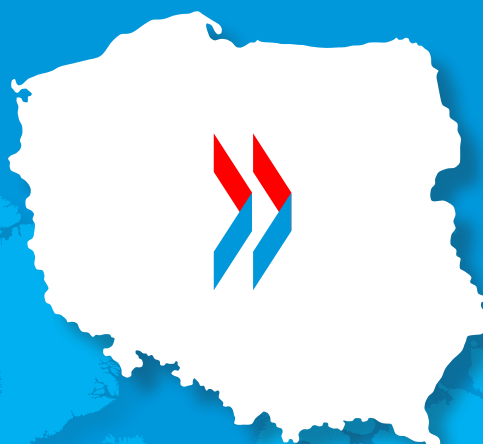


PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT



OECD PISA

PROGRAM MIĘDZYNARODOWEJ OCENY UMIEJĘTNOŚCI UCZNIÓW
WYNIKI BADANIA 2015 W POLSCE



WYNIKI BADANIA 2015 W POLSCE

MINISTERSTWO EDUKACJI NARODOWEJ

Międzynarodowe konsorcjum realizujące badanie OECD PISA 2015:

Educational Testing Servis ETS (USA)
Westat (USA)
Pearson (UK)
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)

Rada Zarządzająca Programem PISA (PGB):

Przedstawiciel Polski – Lidia Olak, Jerzy Wiśniewski

Skład zespołu realizującego badanie PISA w Polsce

Marcin Chrzanowski
Alina Cichomska
Dorota Cyngot
Grażyna Drążyk
Michał Federowicz
Jacek Haman
Zbigniew Marciniak
E. Barbara Ostrowska (kierownik zespołu)
Michał Sitek
Krzysztof Spalik
Agnieszka Sułowska
Grzegorz Zajączkowski
Piotr Zielonka
Piotr Walicki

Raport z badania opracowali: Piotr Bordzoł, Krzysztof Biedrzycki, Krzysztof Bulkowski, Marcin Chrzanowski, Michał Federowicz, Jacek Haman, Zbigniew Marciniak, Dominik Marszał, E. Barbara Ostrowska, Krzysztof Spalik, Agnieszka Sułowska

Redakcja naukowa: Michał Sitek

Redakcja językowa: Natalia Skipietrow

Zespół dziękuje Dyrektorom szkół i Kuratoriom Oświaty za życzliwą pomoc w realizacji badania. Dziękujemy uczniom za udział w badaniach i ich rodzicom za wyrażenie na to zgody.

Za realizację badania w Polsce odpowiadał Instytut Badań Edukacyjnych. Badanie próbne zrealizował Zespół Współpracowników Terenowych IBE, a badanie główne GFK Polonia

Strony internetowe badania PISA w Polsce: www.ibe.edu.pl, www.oecd.org/pisa

ISBN 978-83-65115-87-4

Badanie PISA zostało sfinansowane przez Ministerstwo Edukacji Narodowej.



WSTĘP

Badanie PISA (*Programme for International Student Assessment* – Program Międzynarodowej Oceny Umiejętności Uczniów) realizowane jest przez międzynarodowe konsorcjum nadzorowane przez OECD (Organizację Współpracy Gospodarczej i Rozwoju) i przedstawicieli krajów członkowskich. Jest to największe międzynarodowe badanie umiejętności uczniów na świecie. Jest ono realizowane co 3 lata od 2000 roku we wszystkich krajach OECD, a także w kilkudziesięciu krajach partnerskich. W roku 2015 wprowadzono istotne zmiany w metodologii badania, w szczególności przejście na pomiar komputerowy. Co oznacza, że po raz pierwszy uczniowie rozwiązywali zadania na komputerach, a nie „na papierze”. Dokonano wszelkich starań, by zachować porównywalność wyników w czasie, choć należy mieć świadomość, że tego typu zmiana oznacza, że nie można mieć pewności, czy i w jakim stopniu sposób przeprowadzenia badania wpłynął na wyniki.

Polska od początku uczestniczy w badaniu. Od 2000 roku do 2012 prowadził je Instytut Filozofii i Socjologii PAN, od 2013 – Instytut Badań Edukacyjnych. W realizację badania PISA zaangażowana jest od 2000 roku grupa naukowców, która tworzy interdyscyplinarny zespół badawczy PISA.

W każdej edycji badania jedna z dziedzin: czytanie i interpretacja, umiejętności matematyczne, rozumowanie w naukach przyrodniczych – jest dziedziną wiodącą. W 2015 r. było to rozumowanie w naukach przyrodniczych.

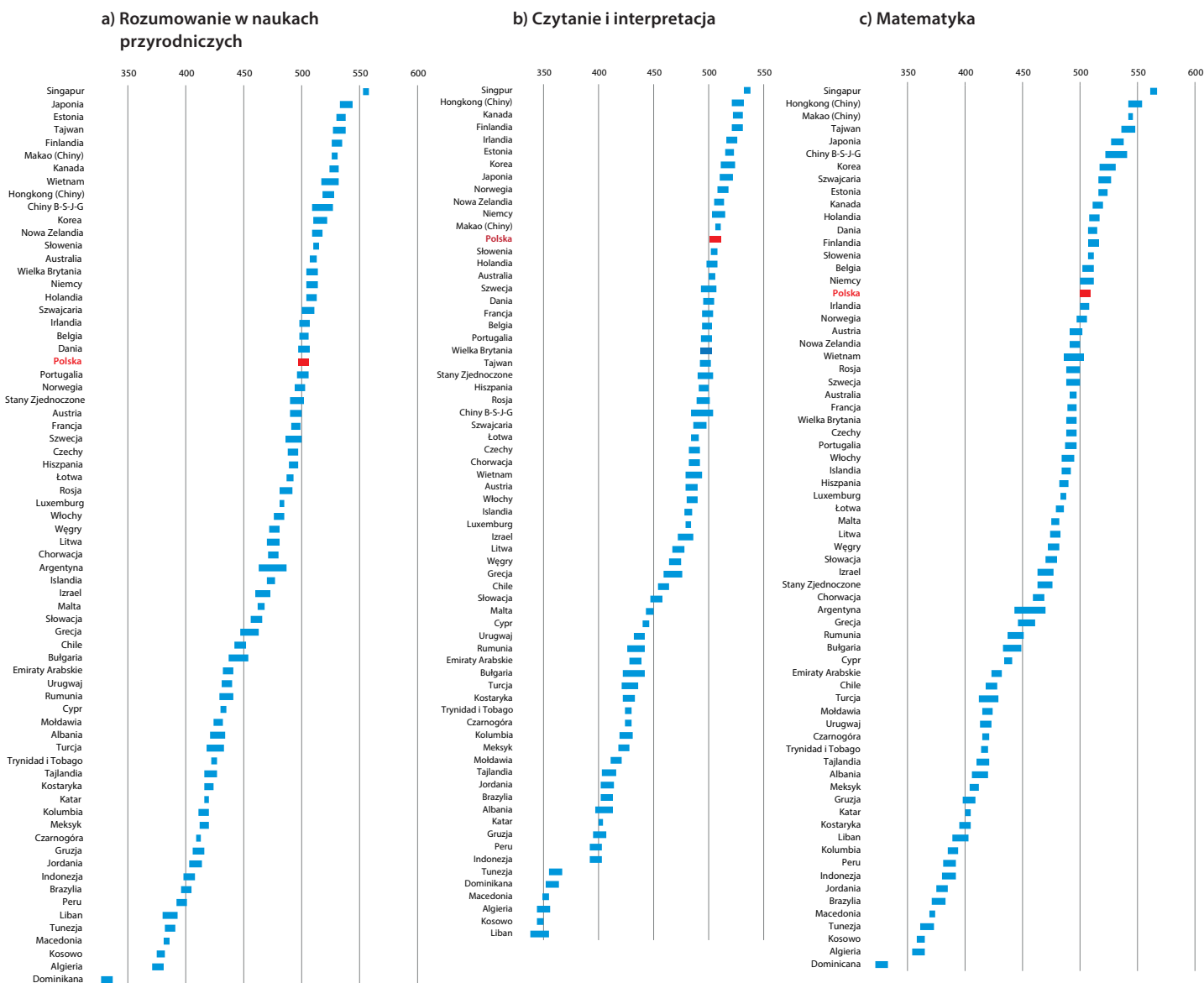
Badanie OECD PISA jest badaniem piętnastolatków – uczniów, którzy w roku poprzedzającym badanie ukończyli lat 15; w 2015 roku byli to uczniowie z rocznika 1999 r.

W PISA 2015 uczestniczyło ponad 500 tys. uczniów z 72 krajów i regionów.

Główne wyniki

- W zakresie rozumowania w naukach przyrodniczych (scientific literacy) wynik polskich uczniów wyniósł 501 pkt, o 24 pkt. mniej niż w badaniu z 2012 r. i był wyższy o 4 pkt. w porównaniu do roku 2006, w którym po raz pierwszy dokonano pełnego pomiaru umiejętności rozumowania w naukach przyrodniczych i opracowano skalę odniesienia dla dalszych pomiarów. Obniżenie wyników pomiędzy latami 2012 i 2015 odnotowano w większości krajów OECD. Wynik polskich uczniów jest zbliżony do wyników piętnastolatków z Irlandii, Belgii, Danii, Portugalii, Norwegii, USA, Austrii i Szwecji – różnice między Polską a tymi krajami były nieistotne statystycznie. Daje nam to dziesiąte miejsce wśród krajów Unii Europejskiej, a statystycznie wyprzedza nas sześć krajów. W czołówce krajów osiągających najwyższe wyniki w badaniu kompetencji przyrodniczych znajdują się kraje Dalekiego Wschodu, a także Finlandia, Estonia i Kanada. Pierwsze miejsce zajmuje Singapur (556 pkt). Z krajów europejskich przoduje Estonia z wynikiem 534 punktów, a w czołówce znajduje się też Finlandia (531 pkt).
- Średni wynik z czytania i interpretacji polskich uczniów wyniósł 506 pkt i jest o 12 pkt niższy od wyniku z 2012 r. W badaniu umiejętności czytania i interpretacji (*reading literacy*) polscy uczniowie znaleźli się na piątym miejscu w Europie. Lepsze średnie wyniki osiągnęli uczniowie z czterech krajów: Finlandii, Irlandii, Estonii i Norwegii. Wśród krajów świata Polska znalazła się na 13 miejscu. Warto zaznaczyć, że jedynie 9 krajów uzyskało wynik statystycznie istotnie lepszy od Polski, zaś w przypadku 3 ich przewaga mieściła się w granicach błędów losowych.

Wykres 1. Wyniki PISA 2015



Średni wynik polskich uczniów w zakresie umiejętności matematycznych (mathematical literacy) to 504 pkt., o 14 pkt. mniej niż w 2012 r. Daje to nam 6 miejsce w UE, za Estonią, Holandią, Danią, Finlandią i Słowenią. Wynik polskich gimnazjalistów jest zbliżony do wyników uczniów z Belgii, Niemiec, Irlandii i Norwegii – różnice między Polską a tymi krajami były nieistotnie statystycznie.

W zakresie rozumowania w naukach przyrodniczych odsetek uczniów najłabszych, czyli poniżej drugiego poziomu umiejętności, w roku 2015 wyniósł 16,3% (w 2012 r. było to 9%, a w 2006 17%). Trudno jednoznacznie określić, na ile zmiana pomiędzy

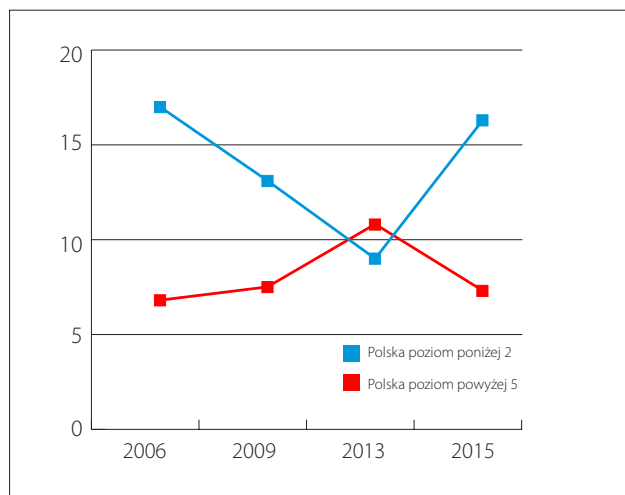
latami 2012 a 2015 wynikała z zastosowania pomiaru komputerowego. Podobne zmiany nastąpiły w pozostałych dziedzinach: w matematyce – 17,2% w 2015 r. (w 2012r. – 14,4%), a w zakresie czytania i interpretacji – 14,4% (w 2012 r. – 10,6%). Analogicznie uczniów z najlepszymi wynikami w rozumowaniu w naukach przyrodniczych było 7,3% w 2015 r. (w 2012 r. – 10,8%, a w 2006 – 6,8%). W matematyce w 2015 r. najlepszych uczniów było 12,2% (w 2012 r. – 16,7%), a w czytaniu – 8,2% (w 2012 r. – 10,1%). Ze względu na zmianę techniki pomiaru wydaje się, że ważniejsze jest odniesienie polskich wyników do wyników innych krajów. Zarówno pod względem odsetka najlepszych, jak i najłabszych uczniów wynik Polski jest lepszy niż średnio w krajach OECD,

z wyjątkiem odsetka uczniów najlepszych w czytaniu i interpretacji, który jest nieco niższy i obszar ten wymaga wsparcia.

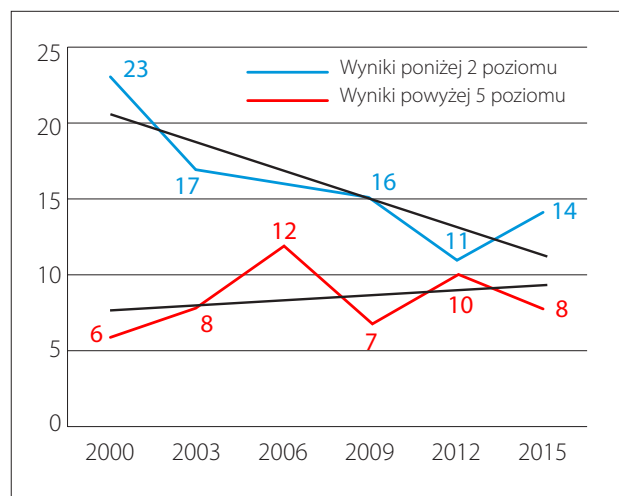
- W rozumowaniu w naukach przyrodniczych w 2015 r. średni wynik chłopców w krajach OECD wyniósł 495 punktów i był o 4 pkt. wyższy od średniego wyniku dziewcząt. W Polsce średni wynik dziewcząt wyniósł 498 pkt., podczas gdy chłopcy osiągnęli 504 pkt. Także w zakresie matematyki odnotowano przewagę chłopców: ich średni wynik to 511 punktów, a dziewcząt – 499 punktów. Po raz pierwszy od lat przewaga chłopców nad dziewczętami w matematyce jest istotna statystycznie (w latach 2009 i 2012 wynosiła tylko 4 punkty). Przewaga chłopców w tej dziedzinie dotyczy także dolnego i górnego krańca skali umiejętności – wśród najsłabszych uczniów (najniższy decyl) wyniki chłopców są o 9 punktów wyższe niż dziewcząt, a wśród najlepszych aż o 19 punktów wyższe. Natomiast w zakresie czytania i interpretacji dziewczęta są zdecydowanie lepsze: w Polsce różnica wyniosła 30 pkt na korzyść dziewcząt, jest ona zbliżona do średniej OECD (27 p.). Dziewczęta uzyskały 521 pkt, a chłopcy 491. We wszystkich trzech dziedzinach chłopcy poprawili swoje wyniki względem dziewcząt pomiędzy latami 2012 a 2015, ale trudno określić czy i w jakim stopniu wpłynęła na to zmiana pomiaru na komputerowy. Wymaga to odrębnych badań.

Wykres 2. Odsetek uczniów, którzy uzyskiwali wyniki poniżej 2 poziomu oraz powyżej 5 poziomu w Polsce w kolejnych edycjach badania (%)

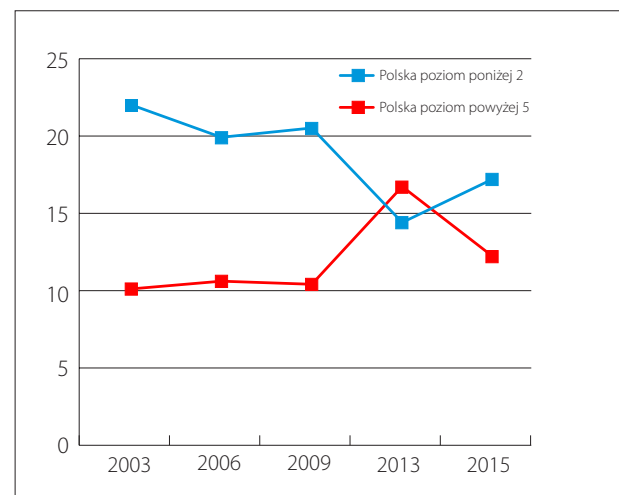
a) Rozumowanie w naukach przyrodniczych



b) Czytanie i interpretacja



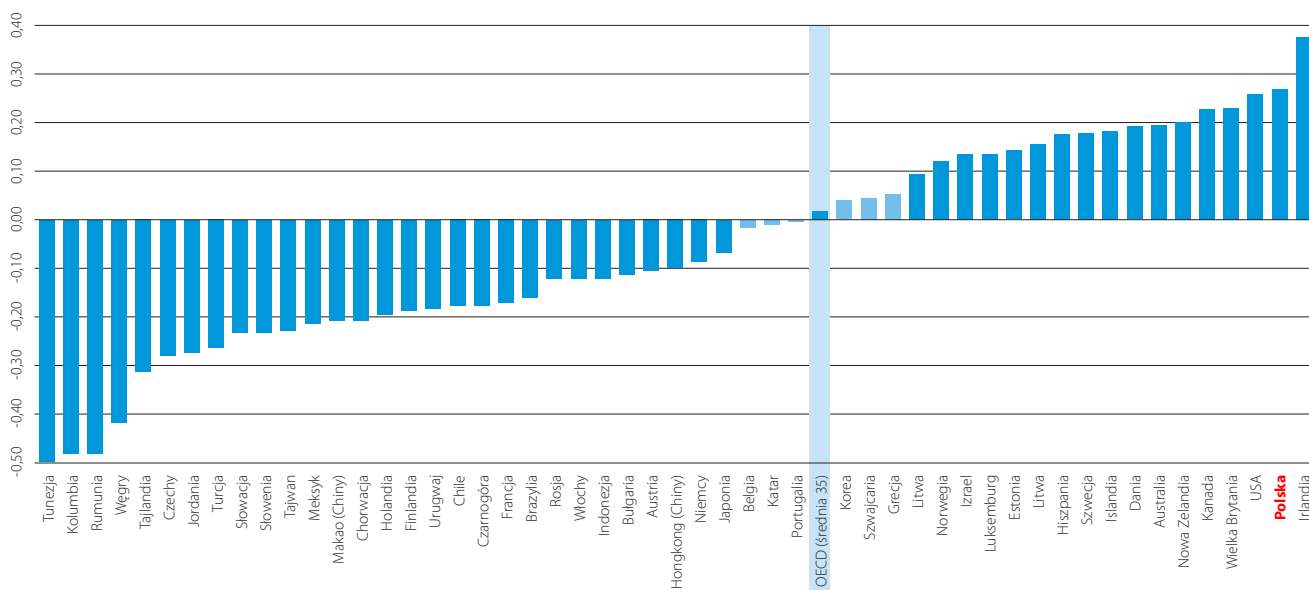
c) Matematyka



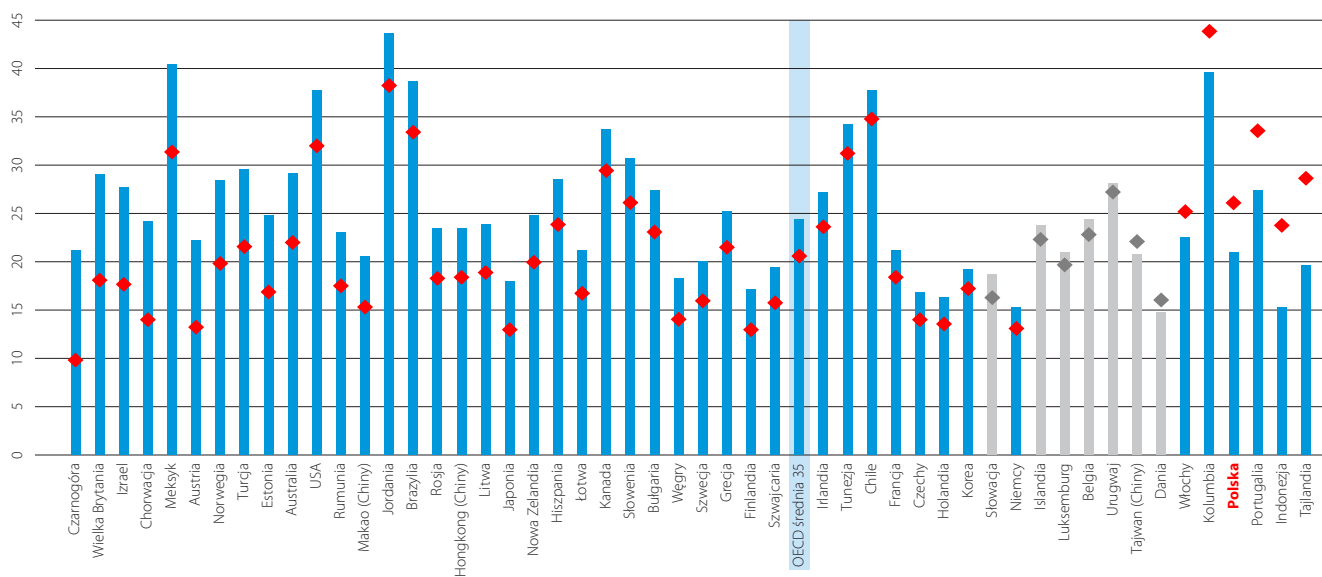
Postawy uczniów wobec nauk przyrodniczych oraz oczekiwania względem kariery zawodowej

W badaniach z 2006 oraz 2015 pytano uczniów o różne aspekty ich postaw wobec nauk przyrodniczych, w tym, czy uczenie się przedmiotów przyrodniczych oraz poznawanie zagadnień naukowych sprawia im przyjemność i daje satysfakcję. Na podstawie tych odpowiedzi opracowano wskaźnik satysfakcji z uczenia się przedmiotów przyrodniczych (*enjoyment of learning science*). Krajami, w których wskaźnik ten wzrósł najbardziej między badaniami 2006 i 2015, były Irlandia i Polska (Wykres 3).

Wykres 3. Zmiana wskaźnika satysfakcji z uczenia się o przyrodzie między badaniami PISA 2006 i 2015. Kraje uszeregowano od najmniejszej do największej zmiany tego wskaźnika. Ciemniejszą barwą zaznaczono różnice istotne statystycznie



Wykres 4. Odsetek uczniów w badaniach PISA 2006 i 2015, którzy pozytywnie odpowiedzieli na pytanie, czy oczekują, że w wieku 30 lat będą pracować w zawodach związanych z naukami przyrodniczymi. Kraje uszeregowano od najwyższej do najniższej wartości różnicy tego odsetka między latami 2015 i 2006. Szarym kolorem oznaczono różnice nieistotne statystycznie.



Uczniowie byli też pytani o to, jak sądzą, jaki zawód będą wykonywać w wieku 30 lat. Odsetek uczniów wskazujących na zawody związane z nauką zmniejszył się, w porównaniu z 2006 r. z ok. 26 do 21 procent. Między 2006 a 2015 nie zmienił się odsetek uczniów wskazujących na zawody inżynierskie i bezpośrednio związane z nauką (te zawody

wskazało w 2015 r. ok. 6% uczniów), zwiększył się odsetek wskazujących na zawody medyczne (wzrost z 8 do 12%), zmniejszył się natomiast odsetek uczniów wskazujących na zawody informatyczne (z 6 do 1%) oraz pozostałe zawody związane z nauką (z 6 do 1%).



INNOWACJE METODOLOGICZNE I CIĄGŁOŚĆ BADANIA

Założeniem badania PISA, od jego pierwszej edycji w 2000 roku, jest dostarczanie szeregu porównywalnych wyników: porównywalnych międzynarodowo, ale także, w kolejnych edycjach, porównywalnych w czasie. Z drugiej strony samo badanie musi iść z duchem czasu: uwzględniać nowe technologie, a także postęp samej teorii pomiaru edukacyjnego. Zmiany w badaniu – niezbędne do utrzymania jego wysokiej jakości – stanowią wyzwanie dla utrzymania porównywalności wyników. Wiedza i doświadczenie zebrane w okresie już prawie dwudziestu lat trwania projektu OECD PISA pozwoliły na zidentyfikowanie szeregu problemów metodologicznych, wprowadzenie szeregu zmian szczegółowych w rozwiązaniach zastosowanych w kolejnych edycjach. Dzięki powszechnemu dostępowi społeczności naukowej do danych PISA i jawności metodologii badania krytyczne analizy – dotyczące w szczególności szczegółowych rozwiązań w skalowaniu danych – prowadzone są nie tylko przez badaczy związanych z konsorcjum międzynarodowym czy krajowymi zespołami realizującymi badanie, ale także przez badaczy z zewnątrz.

Postęp technologii również nie może nie mieć wpływu na badanie PISA. Gdy prowadzona była pierwsza edycja badania PISA, w roku 2000, ponad połowa (52%) polskich piętnastolatków nie miała w domu komputera, blisko jedna trzecia (29%) nigdy nie korzystała z niego w szkole; zaledwie 17% miało dostęp do Internetu w domu, a niespełna 35% kiedykolwiek korzystało z niego w szkole¹. W roku 2015 brak dostępu do komputera i Internetu w domu zdarzał się wyjątkowo, zaś brak dostępu do Internetu w polskim gimnazjum trudno sobie nawet wyobrazić. Komputery (i inne urządzenia spełniające podobne funkcje jak smartfony czy tablety, których w roku 2000 w ogóle jeszcze nie było) odgrywają coraz większą rolę zarówno w życiu uczniów, jak i w nauczaniu, a ekran komputera czy tabletu – chcemy czy nie – staje

się nie mniej naturalnym medium i narzędziem, jak wcześniej papierowy podręcznik, zeszyt, kartka papieru, ołówek. Stopniowe wprowadzanie elementów pomiaru komputerowego – ale tylko w ramach badań opcjonalnych, realizowanych obok głównych badań kompetencji – zaczęto w OECD PISA 2006 (w polskim badaniu od PISA 2009). W badaniu 2015 narzędzia komputerowe całkowicie zastąpiły testy typu „papier i ołówek” i po raz pierwszy w ten sposób realizowany był podstawowy pomiar kompetencji uczniów. Była to – niewątpliwie – największa i najdonioślejsza zmiana i nowość w badaniu PISA 2015, ale nie jedyna.

Przedmiotem ciągłej refleksji – a zatem również ciągłych zmian – są również podstawy koncepcyjne badanych domen wiedzy. W badaniu PISA 2015 główną dziedziną po raz drugi (poprzednio w 2006) było rozumowanie naukowe; w związku z tym szczególnemu przeglądowi i aktualizacji został poddany ramy badania tej części testu.

Czy da się pogodzić zmiany – w narzędziach, w metodologii – z ciągłością, a więc z wymogiem zapewnienia porównywalności wyników w czasie? Na tak postawione pytanie nie da się jednoznacznie odpowiedzieć, a ściślej – nigdy nie będziemy w pełni pewni, czy odpowiedź, której udzieliliśmy, jest poprawna. Najłatwiej ocenić zmiany w sposobie skalowania – można bowiem porównać wyniki wyskalowane na różne sposoby; różnice z reguły nie są wielkie, ale są zauważalne. Konsekwencje zmiany narzędzi – a więc przejścia z pomiaru „papierowego” na „komputerowy” były intensywnie analizowane przede wszystkim w ramach badania próbnego przed badaniem PISA 2015; jego wyniki wskazują, że porównywalność wyników została zachowana, nigdy jednak nie da się usunąć wszystkich wątpliwości.

Na pewno jednak zmiany są niezbędne: zmieniają się sposoby nauczania w szkołach, sposoby korzystania z zasobów

¹ Na podstawie danych z zebranych w OECD PISA 2000.

informacyjnych przez uczniów, zmieniają się wreszcie zasoby naszej wiedzy o pomiarze dydaktycznym i rozwiązania, które były optymalne piętnaście lat temu, dziś mogą być już po części nietrafne.

Pomiar i skalowanie

Wyniki badania PISA – a więc oszacowania kompetencji w poszczególnych obszarach badania – przypisywane są poszczególnym uczniom na podstawie procedury skalowania, uwzględniającej liczbę poprawnie rozwiązanych zadań oraz ich trudność. Zastosowanie tych procedur daje szereg istotnych korzyści, w porównaniu do oceniania kompetencji przez np. zwykle zliczanie odsetka poprawnie rozwiązanych zadań. W szczególności możliwe jest ocenianie na tej samej skali (a więc zapewnienie porównywalności wyników) uczniów, którzy rozwiązywali częściowo różne zestawy zadań testowych (a w pewnym zakresie – nawet uczniów, którzy rozwiązywali całkowicie różne zestawy zadań), bez przyjmowania trudnego do spełnienia założenia o identycznym poziomie ich trudności. Dzięki temu w badaniu PISA możliwe jest wykorzystanie znacznie większej liczby różnorodnych zadań, niż gdyby wszyscy uczniowie mieli wykonywać te same zestawy testowe; możliwe jest również porównywanie wyników różnych edycji badania, pomimo że jedynie część zadań (tzw. zadania kotwiczące) powtarzanych jest w kolejnych edycjach.

Gdy w 2000 roku przeprowadzaliśmy pierwszą edycję badania OECD PISA, skalowanie wyników testów oparte na modelach teorii odpowiedzi na pytanie testowe (IRT) było w Polsce nowością. Od tego czasu IRT było wykorzystywane w szeregu badań, tak krajowych, jak i międzynarodowych, jak również było przedmiotem szeregu publikacji, zarówno o charakterze wprowadzającym, jak i bardzo zaawansowanym. Z tego względu poniższe omówienie skalowania w badaniu PISA dzieli się na dwie sekcje. Sekcja pierwsza przeznaczona jest przede wszystkim dla czytelników niemających na co dzień do czynienia z pomiarem w paradygmacie IRT – przedstawiamy w niej najważniejsze elementy tej koncepcji, niezbędne dla zrozumienia istoty pomiaru kompetencji w badaniu PISA; zainteresowanych jej głębszym poznaniem odsyłamy do dostępnej w Polsce literatury, w szczególności do książek (Jakubowski i Pokropek, 2009) i (Pokropek (red.), 2015).

W drugiej sekcji podajemy najważniejsze informacje na temat szczegółowych rozwiązań modeli IRT stosowanych w badaniu OECD PISA z podkreśleniem tych elementów, które zostały zmodyfikowane w badaniu OECD PISA 2015 względem poprzednich edycji badania.

Założenia ogólne – podstawy modeli IRT²

W badaniu PISA skalowanie wyników testu opiera się na teorii odpowiedzi na pytanie testowe (IRT – *Item Response Theory*), a ściślej – na uogólnionym modelu Rascha.

Koncepcja ta odwołuje się do następujących założeń:

- To, czy dany uczeń rozwiąże prawidłowo dane zadanie, jest zdarzeniem losowym.
- Prawdopodobieństwo zajścia tego zdarzenia determinowane jest przez dwa czynniki:
 - poziom umiejętności ucznia,
 - poziom trudności zadania³.

Zakłada się przy tym określoną postać funkcji wiążącej prawdopodobieństwo rozwiązania zadania o danej trudności z poziomem umiejętności ucznia. W modelu Rascha jest to, zasadniczo, funkcja logistyczna; poszczególne warianty modeli IRT różnią się od siebie głównie uwzględnieniem pewnych dodatkowych parametrów tej funkcji – obok samej trudności zadania także jego mocy dyskryminacyjnej⁴, a w przypadku zadań zamkniętych – prawdopodobieństwo udzielenia poprawnej odpowiedzi poprzez zgadywanie. Zwyczajowo określa się poziom trudności zadania i poziom umiejętności badanego na tej samej skali, przyjmując, że badany o poziomie kompetencji k rozwiąże zadania o trudności k z prawdopodobieństwem równym $\frac{1}{2}$.

² Tekst zawarty w tej sekcji zawiera fragmenty powtórzone bezpośrednio lub z pewnymi modyfikacjami za raportami z badania PISA 2009 (Federowicz et al., 2010) i PISA 2012 (Federowicz et al., 2013).

³ Określenia „poziom umiejętności” i „poziom trudności” odpowiadają podstawowemu zastosowaniu modelu Rascha – pomiaru kompetencji. W innych jego zastosowaniach mówilibyśmy raczej o „natężeniu cechy” oraz o charakterystyce konkretnego jej wskaźnika. Model Rascha w badaniu PISA używany jest – poza samym badaniem kompetencji – do konstruowania szeregu skal „cech kontekstowych”.

⁴ „Moc dyskryminacyjna” zadania określa, jak silna jest zależność prawdopodobieństwa rozwiązania zadania od poziomu umiejętności ucznia: im jest ona wyższa, tym szybciej wzrasta prawdopodobieństwo rozwiązania wraz ze wzrostem umiejętności, natomiast w przypadku zadań o słabej mocy dyskryminacyjnej różnica szans na rozwiązanie zadania między uczniami o wysokich i niskich kompetencjach może być niewielka.

- Zarówno poziom umiejętności poszczególnych badanych, jak i poziom trudności poszczególnych zadań (i ewentualnie ich moc dyskryminacyjna) traktowane są jako zmienne ukryte (latentne) – ich estymacja jest celem procesu skalowania.

W procesie skalowania jednocześnie szacowane są poziomy trudności zadań oraz kompetencje badanych – polega to na poszukiwaniu (za pomocą przede wszystkim procedur iteracyjnych) takich kombinacji ich wartości, które z największym prawdopodobieństwem prowadzą do uzyskania zaobserwowanych wyników badania (estymacja metodami „największej wiarygodności”). Drugim elementem procesu skalowania jest ocena zgodności założeń modelu z danymi obserwowanymi. Przykładowo, może okazać się, iż bardzo trudno jest utrzymać założenie, że szanse na rozwiązanie danego zadania wynikają z poziomu tej samej umiejętności, która odpowiada za pozostałe badania. W takiej sytuacji może okazać się, że trafniejsze wyniki uzyskamy, pomijając w analizie dane odnoszące się do tego zadania.

Estymacja trudności zadań może być dokonywana na całości danych z badania, możliwe jest jednak także wykorzystanie do oceny umiejętności badanych danych o poziomie trudności zadań oszacowanych uprzednio. Możliwość ta wykorzystywana jest na kilka sposobów; w szczególności:

- Użycie „zadań kotwiczących” o trudności oszacowanej już w poprzednich cyklach badania PISA pozwala zakotwiczyć skale PISA względem wcześniejszych edycji badania, a tym samym osiągnąć porównywalność i współmierność wyników kolejnych cykli PISA. Osiągnięcie tego efektu wymaga jednak wyskalowania „zadań kotwiczących” na odpowiednio bogatym materiale – z tego względu pełna porównywalność wyników kolejnych edycji badania PISA dla danej dziedziny możliwa jest jedynie od momentu, gdy dana dziedzina była głównym przedmiotem edycji (jak np. czytanie ze zrozumieniem w edycji PISA 2000, matematyka – PISA 2003, rozumowanie w naukach przyrodniczych – PISA 2006); porównywanie wyników wcześniejszych edycji wiąże się z większym ryzykiem błędów losowych.
- Skalowanie trudności zadań odbywa się wyłącznie z użyciem wyników pochodzących z podstawowej populacji badanych, a więc – populacji piętnastolatków. W badaniach uzupełniających projekt międzynarodowy – jak w prowadzonych w części poprzednich edycji polskich badaniach uczniów szkół ponadgimnazjalnych – wykorzystywane są trudności zadań oszacowane w międzynarodowej części badania. W ten sposób jednocześnie osiągnane są dwa cele: ocena umiejętności w „dodatkowych populacjach” na tych samych skalach co w przypadku piętnastolatków; a jednocześnie odseparowanie podstawowego badania międzynarodowego od dodatkowych elementów badania specyficznych dla poszczególnych krajów.

Możliwe jest szacowanie parametrów zadań z wykorzystaniem jedynie części badanej próby, a następnie skalowanie umiejętności wszystkich badanych uczniów w oparciu o tak uzyskane parametry zadań. Metoda ta wykorzystywana była w poprzednich edycjach badania PISA ze względu na mniejsze wymagania co do mocy obliczeniowej.

- Istotną korzyścią z zastosowania modelu Rascha jest również możliwość oceny na tej samej skali badanych, którzy wykonywali częściowo różne zestawy zadań. W ten sposób możliwe jest wykorzystanie w badaniu znacznie większej liczby zadań, a więc zbadanie znacznie szerszego spektrum podobszarów poszczególnych umiejętności.

Skale (umiejętności badanych i trudności zadań) w modelu Rascha mają charakter skal przedziałowych. Pozwalają zatem na interpretowanie i porównywanie wielkości różnic między poszczególnymi wynikami (np. między średnimi dla krajów, średnimi dla typów szkół, wynikami poszczególnych badanych). Skale te nie mają jednak obiektywnego punktu zerowego – a zatem nie jest możliwe określanie proporcji między wynikami. Tak więc, przykładowo, bezsensowne byłoby stwierdzenie, że „kraj A uzyskał wyniki o 20% lepszy od kraju B”. Jednocześnie, poziom umiejętności wyrażony w punktach PISA ma charakter relatywny (i nie odnosi się do żadnych obiektywnie zdefiniowanych oczekiwań co do tego, co wiedzieć lub umieć badani powinni): skale skonstruowane są w ten sposób, by wartość 500 punktów odpowiadała średniej wyników krajów OECD w badaniu PISA 2000 oraz by jeden punkt odpowiadał jednej setnej odchylenia standardowego wyników w populacji krajów OECD w badaniu PISA 2000 (choć ze względu na dokładności oszacowania, wsteczną porównywalność wyników PISA należy ograniczyć do edycji, w której dana dziedzina była dziedziną wiodącą).

Probabilistyczny charakter *Item Response Theory* oznacza także, że przy interpretacji wyników badania bierze się pod uwagę, że dwóch uczniów o tym samym rzeczywistym poziomie umiejętności może uzyskać w teście różne wyniki, i vice versa, dwie osoby, które uzyskały taki sam wynik, mogą w rzeczywistości mieć umiejętności o różnych poziomach. Innymi słowy, probabilistyczny charakter odpowiedzi na bodziec testowy jest drugim, obok reprezentatywnego charakteru badania (błędy związane z próbą), źródłem błędów losowych w wynikach badania PISA. Sposobem uwzględniania tych błędów w analizie jest wykorzystanie do szacowania poziomów umiejętności uczniów estymatorów „wartości prawdopodobnych” (*plausible values, PV*). Należy jednak mieć świadomość, że to, czym różnią się modele IRT od innych podejść do testowania (jak np. klasyczna teoria testu), to nie sam fakt, że wyniki pomiaru obarczone są niepewnością – od tej bowiem nie jest wolny żaden model. Ważne jest to, że modele IRT pozwalają – w pewnym zakresie – oszacować skalę tej niepewności i zamknąć ją w modelu probabilistycznym.

Celem badania OECD PISA jest znalezienie najlepszych oszacowań przeciętnych wyników dla określonych populacji i dokonywanie porównań międzygrupowych: między uczniami z różnych krajów, szkół różnych typów, między dziewczętami a chłopcami itd. W przeciwieństwie do egzaminów nie było celem badania dostarczanie danych o wynikach konkretnych uczniów biorących w nim udział. Z tego względu zarówno organizacja testu, jak i procedury skalowania dobrane są taki sposób, aby zminimalizować błędy oszacowań średnich i wariancji poziomów umiejętności wyznaczanych dla podzbiorności uczniów, nawet jeśli dzieje się to kosztem większych błędów oszacowań indywidualnych.

Możliwość uzyskania poprawy jakości oszacowań dla zbiorowości kosztem oszacowań indywidualnych może wydawać się paradoksalna. Przykładem takiego mechanizmu – na poziomie organizacji testu – jest wspomniana już parokrotnie zasada „rozdzielania” pełnej puli zadań między poszczególne zestawy testowe. Łatwo zauważyć, że jest to działanie korzystne z punktu widzenia oceny przeciętnego poziomu umiejętności dla zbiorowości: dzięki wykorzystaniu większej liczby zadań możemy uwzględnić więcej częściowych składowych umiejętności czy wiedzy wchodzącej w skład danej dziedziny, zaś z drugiej strony ewentualne specyficzne czynniki sprzyjające lub utrudniające rozwiązanie danego zadania, niepowiązane z badaną dziedziną, będą miały mniejszy systematyczny wpływ na ogólny wynik. Równie łatwo przy tym zauważyć, że zasada ta utrud-

nia porównywanie wyników poszczególnych uczniów, którzy rozwiązywali przecież odmienne zestawy zadań.⁵

Na poziomie skalowania, przykładanie większej wagi do poprawności oszacowań zbiorowych niż indywidualnych przejawia się, z jednej strony, w stosowaniu estymatorów *plausible values*, pozwalającymi na lepsze oszacowanie wariancji oraz błędu losowego dla średnich, kosztem dodatkowego błędu losowego na poziomie oszacowań indywidualnych, z drugiej zaś strony – w uwzględnianiu w modelach skalowania (a ściślej w modelach wyznaczania rozkładów prawdopodobieństwa, z których losowane są wartości *plausible values*) danych kontekstowych, pochodzących z kwestionariuszy ucznia. Działanie takie byłoby absolutnie nie do przyjęcia, gdyby celem badania była ocena indywidualnego poziomu umiejętności danego ucznia: trudno by zaakceptować procedurę, w której ocena ucznia zależałaby nie tylko od odpowiedzi, jakich udzieli na egzaminie, ale również od np. wykształcenia jego rodziców. W sytuacji, gdy celem badania jest ocena przeciętnych wyników dla zbiorowości, postępowanie takie pozwala jednak zmniejszyć poziom błędów losowych oszacowań. Co więcej, oszacowania poziomów poszczególnych umiejętności podawane są – na podstawie danych kontekstowych oraz wyników testów z zakresu pozostałych umiejętności – nawet w przypadku uczniów, którzy zadań z zakresu danej umiejętności w ogóle nie rozwiązywali. Na poziomie indywidualnym takie oszacowania (które w tym wypadku lepiej byłoby nazywać przewidywaniami) są, oczywiście, zwykle obciążone dużym błędem; jednakże przy szacowaniu parametrów poziomu umiejętności dla zbiorowości lepsze (dokładniejsze) oszacowania uzyskuje się, uwzględniając zarówno oszacowania pochodzące z rzeczywistego testowania danej umiejętności, jak i te wyznaczone jedynie na podstawie danych kontekstowych i z pozostałych testów.

Skalowanie wyników PISA 2015 w porównaniu z poprzednimi edycjami badania

Przy zachowaniu tych samych ogólnych założeń, szczegółowe rozwiązania przyjęte w skalowaniu wyników PISA 2015

⁵ Warto jednak zwrócić uwagę, że właśnie dzięki zastosowaniu modeli IRT porównywanie indywidualnych wyników uczniów, którzy pisali różne zestawy zadań jest możliwe – jednakże byłoby ono obciążone większą niepewnością niż w przypadku porównywania wyników uczniów rozwiązujących dokładnie te same zadania ze względu na konieczność uwzględnienia dodatkowego składnika błędu losowego – tzw. błędu łączenia.

podlegały – w porównaniu z wcześniejszymi edycjami badania – pewnym modyfikacjom:

- Zmieniony został stosowany wariant modelu IRT: o ile w cyklach PISA 2000-2012 stosowany był model jedno-parametryczny (z uwzględnieniem jedynie trudności zadania), przy skalowaniu danych z PISA 2015 wykorzystano elementy modelu dwuparametrycznego (trudność oraz moc dyskryminacji)⁶.
- We wcześniejszych cyklach badania, szacowanie trudności zadań prowadzone było na podpróbie, do której włączano losowo po 500 przypadków z każdego kraju/regionu (do roku 2009 – wyłącznie z krajów OECD), jedynie z bieżącego cyklu badania. Przy skalowaniu danych z PISA 2015 szacowanie trudności zadań przeprowadzono na bazie wyników uzyskanych przez wszystkich uczniów z wszystkich dotychczasowych edycji badania.

Zastosowanie elementów modelu dwuparametrycznego prowadzi do uzyskania lepszego dopasowania modelu do danych, a zatem trafniejszych i obciążonych mniejszymi błędami wyników; korzyści wynikające z oparcia oceny trudności zadań na pełnym zestawie danych są również oczywiste.⁷

Kolejna nowość w skalowaniu wyników PISA 2015 wiązała się ze sposobem uwzględniania różnic w trudności poszczególnych zadań w poszczególnych krajach. Co do zasady, jeśli porównania międzynarodowe mają być w ogóle możliwe, należy przyjąć, że to samo zadanie we wszystkich krajach mierzy te same umiejętności, a więc ma również taką samą trudność – jeśli jakieś zadanie okazuje się silnie zależne od czynników kulturowych, powinno zostać wyeliminowane (najlepiej na etapie badania próbnego; w razie potrzeby – także już po badaniu zasadniczym). W szczególnych przypadkach (np. specyficzne problemy z tłumaczeniem zadania) możliwe jest również pominięcie w analizie rozwiązań danego zadania pochodzących od uczniów jednego z krajów.

W badaniu PISA 2015 dopuszczono również inne rozwiązanie: wyznaczenie dla niektórych zadań poziomu trudności specyficznego dla danego kraju. W pewnym uproszczeniu

⁶ Ponadto, we wszystkich cyklach PISA w przypadku zamkniętych pytań testowych uwzględniane było prawdopodobieństwo przypadkowego odgadnięcia odpowiedzi.

⁷ Warto dodać, że dla danych z polskich edycji badania PISA 2000-2012 podobne analizy, uwzględniające taki sposób skalowania danych, zostały przeprowadzone wcześniej i opublikowane w pracy (Dolata, Jakubowski, i Pokropek, 2013).

– oznacza to, że dane o rozwiązaniach tego zadania nie będą miały wpływu na przeciętny poziom uzyskany przez uczniów danego kraju na skali międzynarodowej, jednakże ich uwzględnienie pozwoli na dokładniejsze określenie poziomu umiejętności poszczególnych uczniów.

Kolejna zmiana w sposobie skalowania danych PISA 2015 dotyczy sposobu traktowania zadań, do których dany uczeń nie dotarł (zadania, na które nie udzielono żadnej odpowiedzi, znajdujące się na końcu zestawu testowego). W poprzednich cyklach zadania takie – przy szacowaniu poziomu umiejętności – były traktowane tak samo jak rozwiązane błędnie; natomiast przy szacowaniu trudności zadań – tak, jakby nie było ich w zestawie testowym. W cyklu PISA 2015 zadania takie były traktowane jako nieistniejące zarówno przy szacowaniu poziomu trudności zadania, jak i przy szacowaniu poziomu umiejętności uczniów. Głównym uzasadnieniem dla takiej zmiany jest uspojnienie sposobu traktowania tych zadań w obu sytuacjach; trzeba jednak stwierdzić, że w przypadku tej zmiany można wskazać argumenty zarówno za jej wprowadzeniem, jak argumenty przeciwnie.

Zmiany w modelu skalowania danych PISA 2015 oznaczają generalnie, że dane z obecnego cyklu są wyskalowanie *lepiej* niż w cyklach poprzednich dzięki lepszemu dopasowaniu zastosowanych modeli. Oznacza to jednak, że w pewnym sensie są wyskalowane *inaczej* – a więc zmiana modelu skalowania jest czynnikiem utrudniającym możliwość analizowania zmian w czasie. Efekt ten na szczęście nie jest duży, a jego wielkość została oszacowana poprzez zastosowanie modelu skalowania danych z PISA 2015 do danych z poprzednich edycji badania. W przypadku Polski zmiana modelu skalowania ze starego na nowy oznaczałaby zmianę średniego wyniku badania PISA 2012 o –2 punkty w przypadku umiejętności matematycznych, –4 punkty w przypadku czytania i rozumienia oraz +4 punkty w przypadku rozumowania w naukach przyrodniczych.

Narzędzia pomiaru: od testów papierowych do testów komputerowych

Narzędziem pomiaru w badaniu PISA są testy. W badaniu PISA 2000-2012 miały one postać papierowych zeszytów, wypełnianych przez uczniów w czasie sesji testowej – tak samo, jak dzieje się to na np. egzaminach zewnętrznych, a często również na wewnętrznych sprawdzianach czy klasówkach. Od edycji OECD PISA 2006 zaczęto uzupełniać ba-

danie główne opcjonalnymi badaniami z użyciem narzędzi komputerowych; w Polsce komponent komputerowy realizowany był po raz pierwszy w PISA 2009 (badanie czytania tekstu w formie elektronicznej). W badaniu PISA 2012 moduł komputerowy był już znacznie rozbudowany i obejmował badanie kompetencji matematycznych, czytanie tekstu w formie elektronicznej, oraz zadania z „rozwiązywania problemów”, wciąż jednak był to komponent realizowany niezależnie od głównego, „tradycyjnego” badania realizowanego za pomocą testów papierowych, niejako torujący drogę do zmiany narzędzia w badaniu głównym, planowanej na edycję OECD PISA 2015. W badaniu PISA 2015 testy komputerowe stały się w większości biorących udział w badaniu krajów jedynym narzędziem badania kompetencji.

Można wskazać kilka powodów, dla których celowa była zmiana narzędzia badawczego. Realizowanie testów bezpośrednio na komputerze upraszcza bardzo złożony proces transferu materiałów – zestawów zadań do uczniów, wypełnionych testów do koderów wprowadzających do komputera odpowiedzi na pytania zamknięte i oceniających rozwiązania zadań otwartych. Komputeryzacja procesu testowania pozwala na większe zróżnicowanie zestawów testowych (a więc użycie większej liczby zadań, a także wykorzystanie większej liczby wariantów ich rotacji, co pozwala na zmniejszenie błędów związanych z czynnikami specyficznymi dla poszczególnych zadań lub dla pozycji zajmowanej w teście); dzięki temu w badaniu PISA 2015 możliwe było użycie podobnej liczby zadań z zakresu czytania i interpretacji oraz umiejętności matematycznych, co z dziedziny wiodącej – rozumowania w naukach przyrodniczych. Korzyści na etapie sprawdzania prac (jak eliminacja ręcznego przepisywania odpowiedzi uczniów, ale również możliwości wynikające ze sprawdzania odpowiedzi na zadania otwarte w formie elektronicznej), a także późniejszego transferu danych są oczywiste.

Nie mniej poważne, ale jednocześnie – znacznie trudniejsze do jednoznacznej oceny – są kwestie o charakterze kulturowym. Komputery, tablety, smartfony w coraz większym stopniu zastępują papier i ołówek w codziennym życiu (zwłaszcza młodzieży) jako narzędzie pracy zawodowej, ale również jako narzędzie dydaktyki – zdobywania wiedzy i jej sprawdzania. Stopniowo to komputer, a nie kartka papieru staje się dla uczniów naturalnym medium. Proces ten zachodzi w różnym tempie w różnych krajach, w różnych środowiskach, również – w różnych kontekstach. Zmiana narzędzia w badaniu PISA, z założenia dotycząca wszystkich uczniów w kraju (i zdecydowaną większość uczniów

uczestniczących w badaniu)⁸, w tym sensie nigdy nie mogła przyjść we właściwym momencie – dla części będzie to zmiana zbyt wczesna, dla części – zbyt późna.

Komputerowe badanie umiejętności matematycznych, czytania oraz „rozwiązywania problemów” realizowane w ramach PISA 2012, było prowadzone jako opcja dodatkowa, realizowana po głównym, „papierowym” badaniu PISA, i wykorzystywało nowe, specjalnie przygotowane zestawy zadań. Wyniki tego badania były zaskakujące i, zwłaszcza z polskiego punktu widzenia, niepokojące. Przypomnijmy – o ile w badaniu PISA 2012 średni wynik polskich uczniów w zakresie umiejętności matematycznych wyniósł 518 punktów, to w badaniu komputerowym był on niższy o 29 punktów (489); jeszcze gorszy był wynik czytania tekstu elektronicznego – 477 punktów, a więc o 41 mniej niż w badaniu głównym. Również wynik testu z „rozwiązywania problemów”, 481 punktów, wypadł znacząco poniżej średniej OECD. Tak silna rozbieżność między wynikiem testu papierowego a komputerowego na niekorzyść wyników badania komputerowego w badaniu OECD PISA 2012 dotknęła obok Polski tylko kilka krajów (Słowenia, Izrael). Badanie dostarczyło przy tym zbyt mało danych, by w sposób jednoznaczny można było wskazać przyczyny powstania takiej rozbieżności. Generalnie mogło być nią albo to, że testy komputerowe mierzyły w istocie inne kompetencje niż testy papierowe (takie wątpliwości zgłaszano zwłaszcza wobec zadań z matematyki), albo to, że ich wyniki były silnie zależne od lokalnych uwarunkowań (jak np. opanowanie przez dzieci w danym kraju narzędzi IT). Źródłem różnic mogą być także czynniki motywacyjne (badanie komputerowe było realizowane po głównej części badania – uczestnicy mogli być zatem zmęczeni lub zdekoncentrowani: istotne jest jednak, czy i na ile efekty z tym związane mogły występować z różnym nasileniem w różnych krajach). Zebrane dane nie pozwoliły jednak ani na pełne potwierdzenie, ani wykluczenie takich hipotez.

Aby ograniczyć ryzyka związane z planowaną zmianą narzędzi w PISA 2015, podjęto działania idące w dwóch kierunkach. Po pierwsze, zadania na testy komputerowe 2015 przygotowywano w taki sposób, by stanowiły możliwie bliskie odpowiedniki zadań w wersjach „papierowych”. Oznacza to

8 W OECD PISA 2015 poszczególne kraje mogły zdecydować się na pozostanie przy pomiarze za pomocą testów papierowych, jeśli ich zdaniem przeprowadzenie badań komputerowych w lokalnych warunkach byłoby zbyt skomplikowane lub zbyt ryzykowne. Zdecydowało się na to 15 z 72 biorących udział w badaniu krajów i regionów (spośród krajów UE były to jedynie Rumunia i Malta). Realizacja badania za pomocą testów papierowych oznaczała również rezygnację z wykorzystania w badaniu nowych zadań – te bowiem zostały przygotowane jedynie w wersji komputerowej.

wprawdzie, że nie wykorzystane zostały w pełni możliwości, jakie daje rozwiązywanie zadań na komputerze (innymi słowy: choć test mógłby bardziej przypominać np. grę komputerową, zdecydowano by w większym stopniu imitował test papierowy), dzięki czemu jednak zmniejszono ewentualne efekty pomiarowe związane ze zmianą narzędzia. Po drugie, w ramach badania próbnego, realizowanego (na znacznie większą skalę, niż w poprzednich edycjach PISA) na wiosnę 2014 roku testowano jednocześnie zadania w wersji papierowej i komputerowej, aby wypracować „model przejścia” umożliwiający interpretację wyników testu komputerowego na tej samej skali co wyników testu papierowego. W badaniu próbnym wzięło udział 2171 polskich uczniów (w skali całego badania międzynarodowego ponad 152 tysiące uczniów), podzielonych losowo⁹ na trzy grupy: rozwiązujących zadania kotwiczące w wersji papierowej, rozwiązujących zadania kotwiczące w wersji komputerowej oraz rozwiązujących nowe zadania w wersji komputerowej; porównanie wyników uczniów z grupy pierwszej i drugiej służyło wypracowaniu „modelu przejścia”.

Tym samym zasady przeliczania wyników uzyskanych w testach komputerowych i papierowych zostały opracowane na mocnej podstawie empirycznej. Trzeba jednak pamiętać, że „model przejścia” jest wspólny dla wszystkich uczestniczących w badaniu krajów, a zatem nie uwzględnia np. wszystkich możliwych różnic kulturowych, w wyniku których w różnych społecznościach efekty zmiany narzędzia mogłyby być różne. Stworzenie modelu w większym stopniu uwzględniającego potencjalne różnice kulturowe wymagałoby przeprowadzenia badań na znacznie większą skalę. Efekty związane ze zmianą narzędzia mogą mieć bowiem bardzo różne, trudne do przewidzenia z góry źródła. Mogą to być z jednej strony czynniki związane z ogólnym opanowaniem przez poszczególnych uczniów narzędzi IT (i w tym zakresie stosunkowo najłatwiej jest ewentualne efekty kontrolować), ale także ze specyfiką stosowania IT w różnych kontekstach, tak w edukacji, jak i poza edukacją. Mogą to być także czynniki związane z konkretnymi, przyjętymi w konstrukcji aplikacji testowej rozwiązaniami. Przykładowo – test papierowy może być przez ucznia rozwiązywany

w dowolnej sekwencji – uczeń może dowolnie wiele razy wracać do podjętego zadania, bądź zadanie, które wpięrow uznał za zbyt trudne, może spróbować rozwiązać po rozwiązaniu łatwiejszych itp. W przypadku testu komputerowego „nawigacja” po poszczególnych zadaniach z natury rzeczy odbywa się inaczej niż w przypadku tradycyjnego testu; dodatkowo testy zastosowane w badaniu PISA 2015 uniemożliwiały „cofanie się” do zadań wcześniej porzuconych lub pominiętych (można było jedynie cofać się do wcześniejszych pytań w ramach tego samego zadania). Jedną z wielkich zalet testów komputerowych jest możliwość sprawdzenia, jaki był rytm rozwiązywania zadań testowych przez poszczególnych uczniów (np. zapisywane są szczegółowe informacje o czasie poświęconym na rozwiązanie poszczególnych zadań czy liczba wykonywanych operacji myszką i na klawiaturze) – analiza takich danych pozwoli w przyszłości na lepsze poznanie mechanizmów określających związek między konstrukcją samej aplikacji stosowanej do testowania a wynikami uczniów; brak jest jednak tak bogatych danych w odniesieniu do testów rozwiązywanych tradycyjnie.

Analiza wyników badania próbnego pozwoliła przyjąć, że wyniki badań za pomocą narzędzi komputerowych, wyskalowane zgodnie z wypracowanym „modelem przejścia”, zasadniczo są porównywalne z wynikami wcześniejszych badań za pomocą tradycyjnych testów papierowych. Skala zmian w narzędziach badawczych, a także relatywnie małe jak dotąd doświadczenia w stosowaniu narzędzi komputerowych do badania kompetencji, nakazują jednak pewną ostrożność: zmiana technologiczna jest jednym z najważniejszych „czynników niepewności” przy interpretacji wyników badania OECD PISA 2015. Zmiana ta jednak musiała w którymś momencie nadejść, a w każdym kolejnym badaniu realizowanym za pomocą nowoczesnych technik komputerowych niepewność ta będzie coraz mniejsza, podobnie jak w coraz większym stopniu będziemy korzystali z możliwości, jakie nowe techniki testowania dają badaczom.

Podsumowanie

Wyniki badania OECD PISA 2015 są porównywalne (a więc wyrażone na tej samej skali) co wyniki poprzednich edycji badania. Taka „porównywalność” nigdy nie ma jednak charakteru bezwzględnego – zawsze obciążona jest pewnym marginesem niepewności, związanym z błędami o charakterze losowym lub systematycznym. Skala zmian metodo-

⁹ W badaniu próbnym, w przeciwieństwie do badania zasadniczego, szkoły dobierane są na zasadzie próby celowej, a nie losowej, jednakże dobór uczniów w szkole przeprowadzany jest wedle tych samych zasad co w badaniu zasadniczym, a więc losowo; to samo dotyczy również przydziału poszczególnych wersji narzędzi testowych. Tym samym na podstawie badania próbnego nie można wnioskować o przeciętnym poziomie umiejętności (nie mamy do czynienia z próbą reprezentatywną szkół, a więc i uczniów), można jednak wnioskować np. o relatywnych poziomach trudności zestawów testowych (próby, na których testowane są różne zadania, nie są wprawdzie w pełni reprezentatywne dla całej populacji, natomiast są w pełni porównywalne ze sobą).

logicznych – przede wszystkim zmiana narzędzia z testów papierowych na aplikację komputerową, ale także zmiany szczegółowych rozwiązań stosowanych przy skalowaniu wyników margines ten z całą pewnością poszerzają. Trzeba jednak jeszcze raz podkreślić, że istnienie tego „marginesu niepewności” powinno skłaniać do ostrożności w formułowaniu wniosków, ale przy jej zachowaniu wnioski dotyczące porównań między krajami, jak również porównań w czasie między poszczególnymi edycjami badania OECD PISA mogą być w pełni uprawnione.

W badaniu PISA 2015 obserwujemy systematyczny spadek przeciętnych wyników w krajach, w których w poprzednich edycjach wyniki te były wysokie, i wzrost w krajach o wynikach niższych, choć spadki te i wzrosty są oczywiście różne w różnych krajach i nie da się wskazać żadnej prostej reguły je wyjaśniającej. W efekcie, pomimo że wyrażone w punktach przeciętne wyniki polskich uczniów są w 2015 roku w poszczególnych dziedzinach nieco niższe niż w roku 2012, Polska generalnie utrzymała swoją wysoką pozycję. Można postawić pytanie, na ile zmiany wyników, które obserwujemy, związane są ze zmianami w konstrukcji badania – a więc właśnie są efektami ruchów w granicach tych „marginesów niepewności”, które poprzez staranne przygotowanie badania staramy się zminimalizować, ale których nigdy nie wyeliminujemy. Znalezienie na nie choć częściowej odpowiedzi nie jest jednak łatwe i wymagać będzie jeszcze wielu analiz i dalszych badań.

Bibliografia

- Dolata, R., Jakubowski, M., Pokropek, A. (2013). *Polska oświata w międzynarodowych badaniach umiejętności uczniów PISA OECD: wyniki, trendy, kontekst i porównywalność*. Wydawnictwa Uniwersytetu Warszawskiego.
- Federowicz, M. i in. (2013). *OECD PISA 2012. Wyniki badania 2012 w Polsce*. Ministerstwo Edukacji Narodowej.
- Federowicz, M., i in. (2010). *Program Międzynarodowej Oceny Umiejętności Uczniów OECD PISA. Wyniki badania 2009 w Polsce*. Ministerstwo Edukacji Narodowej.
- Jakubowski, M., Pokropek, A. (2009). *Badając egzaminy: podejście ilościowe w badaniach edukacyjnych*. Centralna Komisja Egzaminacyjna.
- Pokropek, A. (red.) (2015). *Modele cech ukrytych w badaniach edukacyjnych, psychologii i socjologii: teoria i zastosowania*. Instytut Badań Edukacyjnych.



POPULACJA I PRÓBA

Badaniem OECD PISA objęci są uczniowie piętnastoletni, a dokładniej rzecz biorąc – uczniowie, którzy ukończyli 15 lat w roku poprzedzającym badanie; w przypadku PISA 2015 odpowiadało to uczniom urodzonym w 1999 roku (znajdowali się więc wśród nich również tacy, którzy w chwili badania ukończyli już lat szesnaście; operacyjna definicja „piętnastolatka” przyjmowana w badaniu nie jest zatem w pełni zgodna z naturalną interpretacją słowa „piętnastolatek”). Zdecydowaną większość tej grupy wiekowej stanowią uczniowie gimnazjów (przede wszystkim klasy III, ale również klas II i I). Oprócz tego do badanej populacji należeli piętnastoletni uczniowie szkół artystycznych (przede wszystkim ogólnokształcących szkół muzycznych II stopnia)¹, oraz piętnastoletni uczniowie szkół ponadgimnazjalnych (liceów lub średnich i zasadniczych szkół zawodowych). Z założenia, z populacji badania wyłączeni byli natomiast uczniowie szkół specjalnych i piętnastoletni uczniowie szkół podstawowych (wyłączenia na poziomie szkół), a także uczniowie nie-mogący pisać testu ze względu na niepełnosprawność lub niewystarczającą znajomość języka polskiego (wyłączenia wewnętrzne). Do badanej populacji nie należeli także piętnastolatki, którzy z jakichkolwiek przyczyn znajdowali się poza polskim systemem szkolnym (w szczególności – jeśli realizują obowiązek szkolny poza Polską²). Dane o populacji badanej w PISA 2015 podaje Tabela 1.

¹ Jeśli szkoła artystyczna prowadziła naukę na poziomie gimnazjum (np. ogólnokształcące szkoły muzyczne II stopnia), na użytek doboru próby traktowana była tak, jakby była gimnazjum; także dalej w tym tekście tam, gdzie mowa jest o gimnazjach, włącza się w to również ogólnokształcące szkoły muzyczne II stopnia.

² W takim przypadku mogą jednak należeć do populacji badanej w innym kraju.

Tabela 1.

Piętnastolatki: rok urodzenia	1999
Liczba piętnastoletnich gimnazjalistów ¹ (bez szkół specjalnych)	352,2 tys.
Liczba piętnastoletnich uczniów liceów ogólnokształcących	2,6 tys.
Liczba piętnastoletnich uczniów średnich szkół zawodowych	0,6 tys.
Liczba piętnastoletnich uczniów zasadniczych szkół zawodowych	0,1 tys.
Razem	355,5 tys.
Szacunkowa wielość wyłączeń wewnętrznych	2,4 tys.
Łącznie wielkość badanej populacji	353,1 tys.
Liczba gimnazjów (bez szkół specjalnych)	6639

Badanie PISA realizowane jest na reprezentatywnej próbie losowej. Schemat doboru próby uczniów w badaniu PISA ma charakter dwustopniowego doboru warstwowego z zastosowaniem losowania systematycznego, w którym pierwszym stopniem doboru był wybór szkoły, zaś drugim – losowanie uczniów z uprzednio wylosowanych szkół (również w trybie systematycznego losowania warstwowego). Pierwszy etap losowania – losowanie szkół – realizowane jest przez wchodzącą w skład konsorcjum międzynarodowego amerykańską firmę badawczą WESTAT; rolę polskiego zespołu PISA jest dostarczenie pełnej bazy szkół wraz z niezbędnymi do przeprowadzenia losowania danymi. Drugi etap – losowanie uczniów – realizowany jest przez krajowy zespół PISA przy użyciu narzędzia dostarczonego przez WESTAT (co w szczególności oznacza, że krajowy zespół nie ma wpływu na wynik losowania, natomiast

Tabela 2.

	Gimnazja	Licea ogólnokształcące	Technika	Zasadnicze szkoły zawodowe	Razem
Liczba wylosowanych szkół (próba zasadnicza)	160	12	32	21	225
Liczba wylosowanych szkół, w których uczyli się uczniowie piętnastoletni	160	6	4	0	170
Liczba szkół z próby zasadniczej biorących udział w badaniu	143	6	3	0	152
Liczba szkół rezerwowych biorąca udział w badaniu	17	0	0	0	17
Liczba wylosowanych uczniów	5182	13	5	0	5200
Liczba wylosowanych uczniów wyłączonych z próby	55	1	0	0	56
Próba po wyłączeniach	5127	12	5	0	5144
Liczba uczniów biorąca udział w badaniu	4466	8	4	0	4478
Poziom realizacji próby	87,1%	66,7%	80,0%	-	87,1%

nie ma potrzeby przekazywania do WESTAT-a dodatkowych danych polskich uczniów).

Zachowanie ciągłości zasad doboru próby w badaniu PISA jest jednym z podstawowych środków zapewniania porównywalności wyników w poszczególnych edycjach badania – stąd procedury zastosowane w badaniu PISA 2015 były praktycznie takie same jak w roku 2012. Drobne zmiany w porównaniu z poprzednimi edycjami badania wiązały się z faktem, że w roku 2015 w badaniu polskim nie realizowano żadnych opcji krajowych (a więc nie było ani dodatkowego badania uczniów I klas szkół ponadgimnazjalnych, ani nadreprezentacji szkół prywatnych); a opcjonalne międzynarodowe badanie *Financial Literacy* realizowane było na podpróbie głównej próby badania (w PISA 2012 – na dodatkowej próbie uczniów). W efekcie reguły doboru próby do polskiego badania OECD PISA 2012 były maksymalnie bliskie podstawowemu schematowi przewidzianemu dla badania międzynarodowego.

W losowaniu szkół zastosowano podział na warstwy jawne (*explicite*) wyznaczone ze względu na typ szkoły (gimnazja, licea, szkoły zawodowe); w tych trzech kategoriach losowanie było prowadzone osobno. Dzięki stosowaniu metody losowania systematycznego możliwe było wykorzystanie również warstw *implicite* (określających uporządkowanie operatu losowania) – były nimi (w hierarchii od najważniejszej do najmniej ważnej): w warstwie szkół zawodowych, podział na szkoły zasadnicze i technika; we wszystkich warstwach: publiczność (szkoły publiczne, szkoły prywatne),

wielkość miejscowości (miasto na prawach powiatu, pozostałe miasta powyżej 5 tys. mieszkańców, miasta do 5 tys., wieś); skład szkoły ze względu na płeć (szkoły z dominującym udziałem dziewcząt, mieszane, z dominującym udziałem chłopców); ostatnią cechą (ciągłą) porządkującą operat losowania była wielkość (liczba uczniów) szkoły.

Do polskiego badania OECD PISA 2015 wylosowano próbę 160 gimnazjów (W tym – dwie ogólnokształcące szkoły muzyczne II stopnia), 12 liceów ogólnokształcących oraz 53 szkół zawodowych (32 technika i 21 szkół zasadniczych); dodatkowo, dla każdej wylosowanej szkoły, wylosowano dwie szkoły rezerwowe, na wypadek gdyby wylosowana szkoła odmówiła udziału w badaniu. Z 65 wylosowanych szkół ponadgimnazjalnych do badania zostało zakwalifikowanych jedynie 10: 6 liceów i 4 technika (w pozostałych – w tym wszystkich wylosowanych szkołach zasadniczych – nie było żadnych uczniów piętnastoletnich); z czego ostatecznie uwzględniono w badaniu wyniki uczniów 9 szkół. Spośród 160 wylosowanych gimnazjów z próby zasadniczej, 17 szkół odmówiło udziału w badaniu i zostało zastąpione odpowiednimi szkołami rezerwowymi (tak więc przebadano łącznie – zgodnie z zamierzeniem – 160 gimnazjów). Dane o wylosowanej i zrealizowanej próbie zawarte są w Tabeli 2.

Losowanie uczniów do badania również opierało się na schemacie losowania systematycznego z warstwami *implicite* (uporządkowanie operatu ze względu na płeć i klasę – a więc III, II i I-klasistów); zgodnie z tym schematem przydzielano uczniom również konkretne zestawy testowe

i opcje badania. Do badania wylosowanych zostało łącznie 5200 uczniów, w tym 5182 uczniów gimnazjów, 13 uczniów liceów oraz 5 uczniów techników. Z liczby tej wyłączono 56 uczniów ze względu na niespełnianie innych kryteriów udziału w badaniu (jak nieznanostwo języka, niepełnosprawność uniemożliwiająca udział w teście lub ciężka dysleksja), natomiast 665 uczniów nie brało udziału z badaniem z powodu braku zgody rodziców bądź nieobecności. Ostateczna próba zrealizowana wyniosła 4478 uczniów, w tym 4466 uczniów gimnazjów, 8 uczniów liceów oraz 4 uczniów techników. Oznacza to realizację wylosowanej próby na poziomie 87%, a więc minimalnie lepszym niż w roku 2012 (86%).

Jak widać, badanie polskie OECD PISA 2015, choć teoretycznie obejmuje uczniów zarówno gimnazjów, jak i szkół ponadgimnazjalnych, w praktyce jest prawie wyłącznie badaniem gimnazjalistów (99,7% przebadanej próby). Wynika to, z jednej strony z faktu, że odsetek piętnastolatków w szkołach ponadgimnazjalnych to niespełna 1%; z drugiej jednak jest efektem ścisłego stosowania procedur losowania obowiązujących w badaniu międzynarodowym, akurat w tym przypadku niedostosowanych do specyfiki polskiej sytuacji, w wyniku którego ich udział w próbie wyniósł zaledwie 0,3% (w analizie wyników wielkość ta jest następnie korygowana do prawidłowej dzięki ważeniu danych). Trzeba jednak wyraźnie stwierdzić, że nawet, gdyby próba obejmowała odpowiedni odsetek uczniów liceów i techników – a więc byłoby ich w próbie nie 13, ale około 40 – i tak nie pozwalałoby to na jakiegokolwiek uogólnianie wniosków dotyczących tej kategorii uczniów, a w większości sytuacji postępowanie zgodnie w wymogami formułowanymi przez konsorcjum międzynarodowe przynosiło badaniu zdecydowaną korzyść, wymuszając niejako trzymanie się najwyższych standardów metodologicznych.

Dobór próby a błędy losowe i nielosowe

Próba w badaniu PISA jest **próbą losową**, co oznacza, że dla każdego ucznia z badanej populacji można wyznaczyć znane i niezerowe prawdopodobieństwo, że zostanie on do próby wylosowany (w badaniu polskim wynosiło ono przeciętnie ok. 0,01459; wartość ta mogła się wahać dla uczniów różnych szkół). Próba losowa jest z definicji **próbą reprezentatywną**, jeśli prawdopodobieństwa trafienia do niej są równe dla wszystkich członków populacji; jeśli nie są – staje się reprezentatywna po nadaniu jej elementom odpowiednich **wag**, odwrotnie proporcjonalnych do praw-

dopodobieństw trafienia do próby (tzw. ważenie poststratyfikacyjne). Nierówne prawdopodobieństwa trafienia do próby mogą wynikać np. z celowego nadreprezentowania jakichś podzbiorowości (takich, którym z jakichś względów chcemy poświęcić w badaniu szczególną uwagę), bądź też ich subreprezentacji (np. w celu ograniczenia kosztów związanych z badaniem tam, gdzie wiąże się ono ze szczególnie wysokimi nakładami). Takie nad- i subreprezentowanie poszczególnych części populacji nie narusza reprezentatywności próby pod warunkiem odpowiedniego jej ważenia, natomiast może być korzystne ze względów metodologicznych lub na logistykę badania. W roku 2015 nie stosowaliśmy w badaniu polskim żadnych dodatkowych procedur nad- lub subreprezentacji (w niektórych poprzednich edycjach stosowano nadreprezentację szkół prywatnych); elementem wspólnego dla wszystkich krajów uczestniczących w badaniu schematu doboru próby jest natomiast subreprezentacja szkół bardzo małych, realizowana ze względów logistycznych.

Wyniki uzyskane na próbie reprezentatywnej można uogólniać na populację, z której została ona wylosowana, jednakże są one obarczone **błędami losowymi**. Wielkość tych błędów, oczywiście, nie jest znana, jednakże, opierając się na zasadach rachunku prawdopodobieństwa, można określić wielkość, której błąd losowy nie przekroczy z określonym prawdopodobieństwem (zwykle 95%), wyznaczając w ten sposób **przedziały ufności** dla poszczególnych wyników, których szerokość (a ściślej jej połowę) traktuje się zwykle jako dokładność oszacowania (tzw. „błąd statystyczny”). Dokładność oszacowań (a zatem potencjalna wielkość błędów losowych) zależy od dwóch czynników: wielkości próby (im większa próba, tym większa dokładność oszacowań) oraz schematu jej doboru, który może przyczynić się zarówno do zwiększenia, jak i zmniejszenia dokładności w porównaniu do tzw. „prostej próby losowej”.³

Schemat doboru próby uczniów w badaniu PISA ma charakter dwustopniowego doboru warstwowego z zastosowaniem losowania systematycznego, w którym pierwszym stopniem doboru był wybór szkoły, zaś drugim – losowanie uczniów z uprzednio wylosowanych szkół. Zastosowanie doboru warstwowego umożliwia zmniejszenie skali błędów losowych, natomiast dwustopniowy charakter losowania prowadzi do ich zwiększenia w porównaniu do hipotetycznego badania realizowanego na próbie losowej prostej; wielkość błędów losowych w porównaniu do pró-

³ Zależy ona oczywiście również od rozkładu populacyjnego badanej cechy, ta jednak jest czynnikiem niezależnym od charakterystyki próby losowej.

by prostej jest wypadkową tych dwóch czynników, przy czym znacznie silniejszy jest efekt związany z doбором wielostopniowym: błędy losowe wyników uzyskiwanych na liczącej ok. 5000 elementów próbie mają skalę odpowiadającą kilkusetelementowej próbie prostej. Nie oznacza to oczywiście, że lepszą próbą byłaby właśnie próba prosta: ze względów organizacyjnych znacznie łatwiejsze i tańsze do przeprowadzenia jest badanie na próbie 5000 uczniów zgrupowanych w 180 szkołach, niż, powiedzmy, 800 uczniów, z których każdy jest uczniem innej szkoły; niezależnie od tego badanie wielu uczniów w tej samej szkole jest niezbędne, by móc w badaniu charakteryzować nie tylko zatowarowaną społeczność uczniów, ale także wypowiadać się o systemie edukacyjnym, w którego skład wchodzi cała społeczność szkolna.

Zastosowanie złożonego schematu doboru próby oznacza jednak, że do określania dokładności oszacowań nie można stosować standardowych technik obliczeniowych – proste wykorzystanie metod zaimplementowanych w typowych programach statystycznych prowadziłyby do niedoszacowania błędów. Z tego względu błędy losowe wyników w badaniu PISA szacowane są z wykorzystaniem metod replikacyjnych, a ściślej – techniki *balanced random replicates* w wariancie Fay'a. Metoda ta pozwala na wyznaczanie przedziałów ufności oraz weryfikację hipotez statystycznych z uwzględnieniem specyfiki przyjętego schematu doboru próby.

Mówiąc o reprezentatywności próby oraz dokładności oszacowań, należy pamiętać o kilku zasadach:

- Dokładność oszacowań zależy od liczebności próby, natomiast jest praktycznie niezależna od tego, jaką część populacji stanowi próba⁴. W konsekwencji, przy tym samym schemacie doboru próby, próba o liczebności 5000 osób da taką samą dokładność oszacowania niezależnie od tego, czy wylosowana ona była z populacji liczącej sto tysięcy, milion czy dziesięć milionów ludzi.
- Schemat doboru próby, a także sposób jej warstwowania czy ewentualne sub- i nadreprezentacje, o ile są przeprowadzone prawidłowo i właściwie uwzględnione przy ważeniu próby, nie mają wpływu na jej reprezentatywność (choć mają wpływ na wielkości błędów losowych). W konsekwencji w pełni dopuszczalne jest porównywanie wyników uzyskanych z prób wylosowanych przy użyciu różnych schematów doboru czy

4 Pomijamy tutaj sytuację, gdy próba stanowi rzeczywiście znaczącą część (np. 80%) populacji.

wykorzystujących różne warstwowanie. Pomimo tego w kolejnych cyklach badania PISA dąży się do utrzymania stałych schematów doboru i warstwowania próby, jednakże ewentualne ich zmiany nie stanowią przeszkody w porównywaniu wyników badania PISA z różnymi latami.

Realizacja badania a błędy nielosowe

Oprócz błędów losowych, wyniki każdego badania mogą być obciążone różnorodnymi *błędami nielosowymi* (*systematycznymi*). Ich źródłem może być nietrafność narzędzi pomiarowych, błędy proceduralne w realizacji badania, a przede wszystkim – niepełna realizacja próby. W przeciwieństwie do błędów losowych, których wartości wprowadzić nie znamy, ale ich skalę możemy przewidzieć, błędy nielosowe nie poddają się opisowi probabilistycznemu i nie da się łatwo wskazać, jaką wielkość czy kierunek mogą osiągnąć. Błędów takich nie da się nigdy całkowicie uniknąć, jednak staranne przygotowanie i realizacja badania ma służyć ich minimalizacji.

Poziom realizację próby uczniów w polskim badaniu OECD PISA 2015 wynosił 87% – był więc nieco wyższy niż w badaniu 2012 (ok. 86%)⁵ i spełniał wymagania standardów technicznych badania PISA (min. 80%). Również odsetek szkół z próby zasadniczej uczestniczących w badaniu (89%) znacznie przekracza wymogi międzynarodowe i jest nieznacznie wyższy niż w 2012 roku (88%). Wielkości te warto porównać z poziomami realizacji prób w typowych badaniach społecznych na próbach reprezentatywnych dorosłej ludności Polski, które w przypadku najlepiej realizowanych badań akademickich zwykle nie przekraczają 65%, a w przypadku badań komercyjnych (np. sondaże przedwyborcze) nie osiągają nawet 40% założonej próby.

Można jednak postawić pytanie, czy – a jeśli tak, to na ile – niepełna realizacja próby wpływa na uzyskane wyniki, w szczególności – na średnie liczby punktów uzyskanych w testach kompetencji przez polskich uczniów. Byłoby tak, gdyby przeciętny poziom umiejętności uczniów wylosowanych do próby, ale nie uczestniczących w badaniu, był inny niż uczniów, którzy udział w badaniu wzięli. Tak najprawdopodobniej rzeczywiście jest – wskazują na to np. różnice w poziomie realizacji próby wśród wylosowanych uczniów klasy III gimnazjum (88%) a klas I i II (zaledwie 66%). Pojawia się zatem pytanie, jak duży jest wpływ tego zjawiska na

5 Wg metodologii przyjętej przez konsorcjum międzynarodowe.



wynik badania oraz na możliwość porównywania wyników w czasie lub między krajami. Kwestia ta była przedmiotem szczegółowej analizy polskich wyników badań PISA 2006, 2009 i 2012, gdzie w oparciu o zbierane dane o wynikach egzaminów zewnętrznych uczniów wylosowanych do próby (zarówno biorących, jak i nie biorących udziału w badaniu) szacowano wielkość efektu niepełnej realizacji próby na około 1-3 punkty. Ponieważ parametry realizacji próby (zarówno jej ogólny poziom, jak i struktura) w roku 2015 są podobne jak w poprzednich edycjach badania, można oczekiwać, że również w badaniu PISA 2015, gdyby – hipotetycznie – uzyskano 100% realizacji próby, średnie wyniki polskich uczniów okazałyby się mniejsze o również ok. 1 do 3 punktów w porównaniu z ogłoszonymi wynikami badania (co jest wartością mniejszą, niż niepewność wyniku związana z błędami losowymi).

Należy jednak podkreślić, że wyniki badania PISA mają z zasady charakter relatywny. Porównując wyniki 2015 do poprzednich edycji badania, należy brać pod uwagę przede wszystkim to, czy przewidywana wielkość błędów systematycznych nie uległa zmianie. Ponieważ charakterystyka

realizacji próby PISA 2015 w porównaniu z poprzednimi edycjami nie uległa zauważalnej zmianie, można przyjąć, że również ewentualne zniekształcenia wyników z nią związane są podobne – a zatem zmian średniego poziomu umiejętności między poszczególnymi edycjami badania nie można wiązać z niepełną realizacją próby, choć – zwłaszcza w bieżącej edycji badania – nie można wykluczyć, że ma ona pewien związek np. ze zmianą narzędzia badawczego z testów papierowych na komputerowe.

W kwestii porównań międzynarodowych brak jest danych dotyczących wpływu niepełnej realizacji próby dla średnich wyników uzyskanych w innych krajach niż Polska, można jednak przypuszczać, że podobne mechanizmy jak w Polsce mogą działać także w innych krajach. Ponieważ poziom realizacji próby w Polsce jest podobny jak w innych krajach uczestniczących w badaniu, można oczekiwać, że również wielkość efektów niepełnej realizacji próby będzie podobna. Tak więc niepełna realizacja próby nie stanowi istotnej przeszkody dla dokonywania międzynarodowych porównań wyników – choć, oczywiście, stanowi ona pewien dodatkowy czynnik niepewności.



ROZUMOWANIE W NAUKACH PRZYRODNICZYCH

Istotą dobrej edukacji przyrodniczej jest przede wszystkim wykształcenie u uczniów umiejętności *rozumowania naukowego*. Tę właśnie kluczową umiejętność mierzy badanie PISA w zakresie kompetencji w naukach przyrodniczych (*science literacy*). Badanie to więc służy nie tyle oszacowaniu wiedzy uczniów w zakresie nauk przyrodniczych, ile sprawdzeniu, czy potrafią oni tę wiedzę zastosować, zwłaszcza kiedy muszą twórczo rozwiązać problemy przedstawione w różnych, a nie tylko szkolnych kontekstach. Powszechna i wysokiej jakości edukacja matematyczno-przyrodnicza jest warunkiem postępu naukowego i technicznego, bez którego nie ma rozwoju cywilizacyjnego. Kanon wiedzy przyrodniczej, przede wszystkim biologicznej, zmienia się nieustannie, nowe wyzwania przynoszą także zdobycze techniki, zwłaszcza w zakresie dostępu do informacji oraz jej przetwarzania. Istnieje jednak coś, co wyróżnia dociekanie naukowe i stanowi jego fundament – to sposób weryfikacji twierdzeń nauki, bazujący na próbie ich obalenia za pomocą doświadczeń i obserwacji, czyli przez *falsyfikację*.

Pewność twierdzeń nauki wynika z tego, że przetrwały one nieustanne próby wytrzymałości. Formułowanie hipotez i ich falsyfikowanie za pomocą danych empirycznych jest więc podstawą metody naukowej.

W 2015 r. rozumowanie w naukach przyrodniczych było główną dziedziną pomiaru po raz drugi w historii tego badania. Pomiar kompetencji piętnastolatków w zakresie rozumowania naukowego obecny był w badaniu PISA od jego pierwszej edycji w 2000 r., ale dopiero w 2006 r. był główną dziedziną pomiaru. Oznacza to, że kompetencje przyrodnicze piętnastolatków zostały wtedy wszechstronnie sprawdzone za pomocą rozbudowanego zestawu zadań, obejmujących różne aspekty wiedzy przyrodniczej oraz różne składowe umiejętności rozumowania i wnioskowania naukowego. Pozwoliło to na miarodajne wyskalowanie wyników badania. Dlatego, jak przypominano w kolejnych raportach, wyniki badań z lat 2000 i 2003 nie są w pełni porównywalne z prowadzonymi od 2006 r.

Tabela 1. Aspekty pomiaru rozumowania w naukach przyrodniczych

Kontekst	Zagadnienia osobiste, miejscowe lub narodowe oraz globalne, zarówno współczesne, jak i historyczne, które wymagają rozumienia zagadnień z zakresu nauki i techniki.
Wiedza	Znajomość najważniejszych faktów i pojęć oraz zrozumienie teorii wyjaśniających, które tworzą podstawę wiedzy naukowej. Wiedza ta obejmuje znajomość zjawisk i procesów zachodzących w przyrodzie (<i>content knowledge</i>), znajomość procedur badawczych np. planowania doświadczeń i obserwacji (<i>procedural knowledge</i>) oraz zrozumienie podstaw wnioskowania naukowego (<i>epistemic knowledge</i>).
Umiejętności	Wyjaśnianie zjawisk przyrodniczych w sposób naukowy, planowanie i ocena poprawności procedur badawczych, interpretacja danych i dowodów naukowych.
Postawy	Zespół postaw obejmujących m.in. zainteresowanie nauką i techniką, docenianie naukowego podejścia do zdobywania wiedzy oraz świadomość wpływu człowieka na środowisko.

Tabela 2. Konteksty zadań w badaniu rozumowania w naukach przyrodniczych PISA 2015

	Osobisty	Lokalny/narodowy	Globalny
Zdrowie i choroby	Profilaktyka zdrowia, wypadki, odżywianie się	Zapobieganie rozprzestrzenianiu się chorób, wybór pożywienia, zdrowie publiczne	Epidemie chorób zakaźnych
Zasoby naturalne	Osobiste zużycie materiałów i energii	Potrzeby populacji ludzkich, jakość życia, bezpieczeństwo, produkcja i dystrybucja żywności, zaopatrzenie w energię	Zasoby odnawialne i nieodnawialne, wzrost ludności, zrównoważone wykorzystanie gatunków
Jakość środowiska	Działania korzystne dla środowiska, wykorzystanie i usuwanie materiałów i urządzeń	Rozmieszczenie ludności, usuwanie odpadów, wpływ różnych działań na środowisko naturalne	Różnorodność biologiczna, zrównoważone wykorzystywanie środowiska, ograniczenie zanieczyszczeń, produkcja/utrata biomasy, erozja gleby
Zagrożenia	Ocena ryzyka związanego z trybem życia	Nagłe zmiany (np. trzęsienia ziemi, gwałtowne zjawiska pogodowe), powolne, stopniowe zmiany (np. erozja brzegowa, sedymentacja), ocena ryzyka	zmiany klimatyczne, oddziaływanie współczesnej komunikacji
Nowe wyzwania nauki i techniki	Naukowe aspekty hobby, sportu, muzyki i wykorzystywania urządzeń technicznych	Nowe materiały, urządzenia i procesy, modyfikacje genetyczne, technologie medyczne, transport	Wymieranie gatunków, badania kosmosu, pochodzenie i budowa Wszechświata

Założenia teoretyczne badania

Założenia badania 2015¹ są podobne do tych z 2006²., jednak dokonano pewnych zmian, np. określenie badanych obszarów wiedzy zostało na nowo zdefiniowane. Jednocześnie po raz pierwszy całe badanie przeprowadzone było w większości krajów z wykorzystaniem komputera, nie zaś – jak poprzednio – w formie papierowej.

W badaniu 2015 przyjęto, że na pomiar kompetencji piętnastolatków w naukach przyrodniczych składają się następujące aspekty: kontekst zadania oraz wiedza, umiejętności i postawy uczniów (Tabela 1.).

Bardzo ważnym aspektem badania PISA jest kontekst, w jakim sprawdzane są umiejętności uczniów. Choć nawiązuje on do podstaw programowych i programów nauczania krajów biorących udział w badaniu, to zadania rozwiązywane przez uczniów nie mają typowo szkolnego charakteru, ale odnoszą się do sytuacji znanych z codziennego życia, do spraw ważnych lokalnie i globalnie, do różnorodnych zjawisk przyrodniczych. Niekiedy przywoływane są zdarzenia historyczne, aby sprawdzić, czy uczniowie rozumieją procesy czasowe oraz potrafią dostrzec znaczenie rozwoju nauki. W Tabeli 2. przedstawiono różnorodność zadań pod kątem ich kontekstu.

¹ OECD (2016). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy*. PISA, OECD Publishing, Paryż.

² Zespół Badania PISA w Polsce (2007). *Program Międzynarodowej Oceny Umiejętności Uczniów OECD PISA: Wyniki badania 2006 w Polsce*. Ministerstwo Edukacji Narodowej, Warszawa.

Wiedza naukowa

Aby wykonać zadania, uczniowie powinni dysponować odpowiednią wiedzą, w której można wyróżnić trzy zasadnicze aspekty:

- wiedza o **treściach nauki** (*content knowledge*) obejmuje znajomość faktów, pojęć i teorii naukowych wyjaśniających złożoność przyrody i przebieg procesów naturalnych; do tej kategorii należy większość szkolnych treści nauczania przedmiotów przyrodniczych – biologii, chemii, fizyki i geografii;
- wiedza o **procedurach badawczych** (*procedural knowledge*) to znajomość podstawowej metodyki badań naukowych, jak planowanie doświadczeń i obserwacji, świadomość niepewności pomiaru i wynikającej stąd konieczności powtarzania pomiarów i badań, określenie znaczenia próby kontrolnej, znajomość sposobów analizy i prezentacji wyników;
- wiedza o **poznaniu naukowym** (*epistemic knowledge*) to zrozumienie logicznych podstaw dociekania naukowego, na które składają się stawianie pytań badawczych i formułowanie hipotez, konstruowanie modeli zjawisk, weryfikacja hipotez przez obserwacje i doświadczenie, ale także świadomość znaczenia weryfikacji rezultatów jednych badaczy przez drugich poprzez recenzowanie prac lub powtarzanie badań, dzięki czemu twierdzenia nauki są wiarygodne.

W wyborze treści naukowych obecnych w zadaniach kierowano się następującymi przesłankami:

- treści naukowe mają powiązania z rzeczywistymi sytuacjami lub zdarzeniami;
- są to ważne koncepcje albo teorie naukowe o nieprzemijającym znaczeniu;
- treści są dostosowane do poziomu rozwojowego piętnastolatków.

Umiejętności

Na kompetencje przyrodnicze uczniów mierzone w badaniu PISA składają się trzy główne grupy umiejętności (Tabela 3.):

- wyjaśnianie zjawisk przyrodniczych w sposób naukowy,
- planowanie i ocena poprawności procedur badawczych,
- interpretacja danych i dowodów naukowych.

Umiejętności te odwołują się do opisanych powyżej kategorii wiedzy. Wyjaśnianie zjawisk przyrodniczych w sposób naukowy wymaga przede wszystkim (choć nie wyłącznie) opanowania wiedzy z zakresu treści nauki. Wiedza o procedurach badawczych jest niezbędna, aby uczeń mógł zaplanować eksperyment albo ocenić poprawność takiego planu, sporządzonego przez inną osobę. Interpretacja danych i dowodów naukowych wymaga wiedzy o regułach poznania naukowego.

Wymagania poznawcze

Nowością w badaniu 2015 jest określenie poziomu wymagań (*cognitive demand*), niezbędnego dla rozwiązania poszczególnych zadań. Nie należy mylić wymagań poznawczych z trudnością zadania. Empiryczna trudność zadania jest oszacowana na podstawie jego rozwiązywalności w badaniu. Natomiast poziom wymagań poznawczych zależy od tego, jak złożone rozumowanie trzeba przeprowadzić, by rozwiązać zadanie. Zadanie może być trudne, ponieważ wymaga odwołania się do mało znanych treści naukowych, ale może być mało wymagające pod względem rozumowania. Trudność zadania nie musi jednak polegać na znajomości treści. Zadanie może stawiać przed uczniem wysokie wymagania kognitywne, polegające na analizie, ocenie i połączeniu wielu różnorodnych elementów, które same w sobie są każdemu uczniowi znane.

Pomiar kompetencji w naukach przyrodniczych

Charakterystyka zadań wykorzystanych w badaniu

W badaniu użyto 184 zadania; 85 z nich, tzw. zadania kognitywne, było wykorzystywanych w poprzednich edycjach PISA od 2006 r. Nie były one publikowane i pozostały tajne, a tym samym pozwoliły na porównanie wyników między latami. W badaniu 2015 zostały one zaadaptowane do wersji komputerowej. Pozostałe zadania (99) były nowe. Należy podkreślić, że wiele zadań było interaktywnych, a zatem opracowanych specjalnie pod kątem wykorzystania komputera w badaniu. Nie mają one swoich papierowych odpowiedników. Z wszystkich 184 zadań nieco ponad połowa (98) wymagała wiedzy o treściach nauki, 60 odwoływało się do znajomości procedur badawczych, a 26 sprawdzało opanowanie rozumowania naukowego. Pod względem

Tabela 3. Umiejętności z zakresu kompetencji naukowych mierzone w badaniu PISA 2015

Wyjaśnianie zjawisk przyrodniczych w sposób naukowy
<p>Podając, rozpoznając lub oceniając wyjaśnienia różnorodnych zjawisk z zakresu przyrody i techniki, uczeń potrafi:</p> <ul style="list-style-type: none"> ▪ przywołać z pamięci i zastosować odpowiednią wiedzę naukową; ▪ wskazać, wykorzystać lub stworzyć model lub inne przedstawienie wyjaśniające dane zjawisko; ▪ formułować i uzasadniać odpowiednie przypuszczenia; ▪ stawiać hipotezy; ▪ objaśniać potencjalne następstwa wiedzy naukowej dla społeczeństwa.
Planowanie i ocena poprawności procedur badawczych
<p>Opisując i oceniając badania naukowe i proponując sposoby odpowiedzi na pytania badawcze, uczeń potrafi:</p> <ul style="list-style-type: none"> ▪ wskazać problem podejmowany w określonym badaniu naukowym; ▪ odróżnić pytania, na które można odpowiedzieć w sposób naukowy, od niemających takiego charakteru; ▪ podać sposób naukowego poszukiwania odpowiedzi na określone pytanie badawcze; ▪ ocenić różne sposoby naukowego poszukiwania odpowiedzi na określone pytanie badawcze; ▪ opisać i ocenić, w jaki sposób naukowcy starają się zagwarantować rzetelność danych oraz obiektywizm i uniwersalność wniosków.
Interpretacja danych i dowodów naukowych
<p>Analizując i oceniając dane naukowe, tezy i argumenty, podane w różnej formie, a także wyciągając odpowiednie wnioski, uczeń potrafi:</p> <ul style="list-style-type: none"> ▪ przetworzyć dane naukowe podane w jednej formie w inną formę; ▪ analizować i interpretować dane i wyciągać odpowiednie wnioski; ▪ wyodrębnić założenia, wskazać dowody i określić wnioskowanie w tekstach dotyczących nauki; ▪ odróżnić argumenty bazujące na dowodach i teoriach naukowych od tych opartych na innych podstawach; ▪ ocenić wiarygodność naukową tekstów z różnych źródeł (gazet, czasopism, Internetu itp.).

poszczególnych umiejętności naukowych 89 zadań sprawdziło przede wszystkim umiejętność wyjaśniania zjawisk przyrodniczych w sposób naukowy, 39 zadań dotyczyło planowania i oceny poprawności procedur badawczych, a 56 zadań wymagało wykazania się umiejętnością interpretacji danych i dowodów naukowych. 56 zadań stawiało niskie wymagania poznawcze, 15 wymagało wysokiego poziomu umiejętności rozumowania, natomiast większość (113 zadań) zostało zaklasyfikowanych do średniej kategorii.

Określenie poziomów umiejętności uczniów i trudności zadań

Sposób opracowania (skalowania) wyników przyjęty w badaniu PISA pozwala przedstawić na jednej skali poziom trudności zadania oraz wynik ucznia. Oznacza to, że jeśli np. trudność zadania wyniosła 500 pkt., to uczeń, który uzyskał podobny wynik, rozwiązuje to zadanie z prawdopodobieństwem 50%. Ponieważ każde zadanie zostało przeanalizowane pod względem wymagań poznawczych, można na

tej podstawie wyróżnić i scharakteryzować poziomy umiejętności badanych piętnastolatków (Tabela 6).

Wyniki uczniów oraz ich zmiany w latach 2006-2015

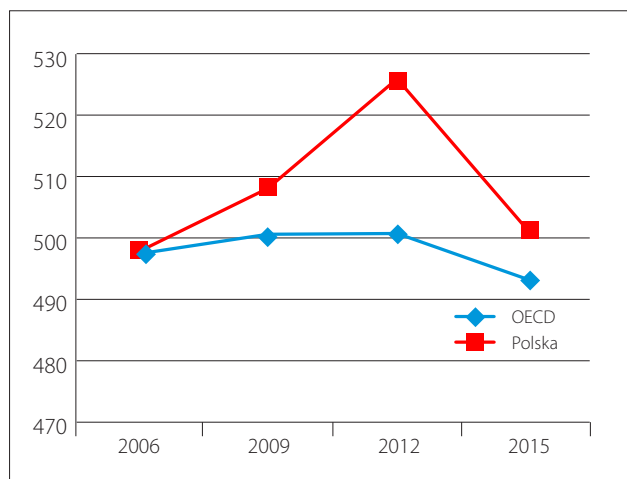
Pomiar kompetencji przyrodniczych był główną domeną badania w 2006 r., dlatego też wyniki z 2006 r. przyjęto za punkt odniesienia do wyników badania 2015 r. Średnie wyniki uzyskane przez uczniów krajów biorących udział w badaniu na ogólnej skali umiejętności przedstawiono w tabeli 5. W obu edycjach badania brała udział inna liczba krajów. W badaniu 2006 uczestniczyło 57 krajów, a średni wynik dla krajów członkowskich OECD był na poziomie 498 punktów, natomiast w roku 2015 do badania przystąpiły 72 kraje lub regiony, a średni wynik krajów OECD wyniósł 493 punkty.

Tabela 4. Średnie wyniki uczniów z pomiaru rozumowania w naukach przyrodniczych w badaniach z lat 2006 i 2015. Białe tło oznacza kraje, których wynik nie był istotnie różny od przeciętnego w krajach OECD, jasnym kolorem wyróżniono kraje o wynikach wyższych, a ciemnym – niższych od przeciętnego. Pogrubioną czcionką wyróżniono kraje europejskie. Dane dla Chin pochodzą tylko z czterech prowincji. Regiony Chin, które niezależnie wzięły udział w badaniu, oznaczono kursywą.

W czołówce krajów osiągających najwyższe wyniki w badaniu kompetencji przyrodniczych znajdują się kraje Dalekiego Wschodu, a także Finlandia, Estonia i Kanada. Pierwsze miejsce zajmuje Singapur (556 pkt), który w 2006 r. nie brał udziału w badaniu PISA. Uczniowie z tego kraju uzyskali średni wynik o 63 pkt wyższy niż średni wynik uczniów krajów OECD. Z krajów europejskich przoduje Estonia z wynikiem 534 punktów, a w czołówce znajduje się też Finlandia (531 pkt), która miała najwyższy wynik w 2006 r. Najniższe wyniki osiągnęli uczniowie z Dominikany (332 pkt), Algierii (376 pkt) i Kosowa (378 pkt). Polscy uczniowie uzyskali wynik 501 pkt i jest on zbliżony do wyników piętnastolatków z Irlandii, Belgii, Danii, Portugalii, Norwegii, USA, Austrii i Szwecji – różnice między Polską a tymi krajami były nieistotne statystycznie. W badaniu 2006 wyniki polskich uczniów były na poziomie średniej OECD, natomiast w 2015 roku nasi uczniowie znajdują się wśród krajów, których wynik jest istotnie wyższy od średniej OECD.

W porównaniu z badaniami z lat 2009 i 2012 zarówno wynik polskich uczniów, jak i średni wynik dla krajów OECD jest niższy (Wykres 1., Tabela 4).

Wykres 1. Zmiany wyników pomiaru rozumowania w naukach przyrodniczych uczniów w Polsce i średnio w krajach OECD w latach 2006-2015.



PISA 2006	
Kraj lub region	Średni wynik
Finlandia	563
Hongkong (Chiny)	542
Kanada	534
Tajwan	532
Estonia	531
Japonia	531
Nowa Zelandia	530
Australia	527
Holandia	525
Liechtenstein	522
Korea	522
Słowenia	519
Niemcy	516
Wielka Brytania	515
Czechy	513
Szwajcaria	512
Makao (Chiny)	511
Austria	511
Belgia	510
Irlandia	508
Węgry	504
Szwecja	503
Polska	498
Dania	496
Francja	495
Chorwacja	493
Islandia	491
Łotwa	490
Stany Zjednoczone	489
Słowacja	488
Hiszpania	488
Litwa	488
Norwegia	487
Luksemburg	486
Rosja	479
Włochy	475
Portugalia	474
Grecja	473
Izrael	454
Chile	438
Serbia	436
Bułgaria	434
Urugwaj	428
Turcja	424
Jordania	422
Tajlandia	421
Rumunia	418
Czarnogóra	412
Meksyk	410
Indonezja	393
Argentyna	391
Brazylia	390
Kolumbia	388
Tunezja	386
Azerbejdżan	382
Katar	349
Kirgistan	322

PISA 2015	
Kraj lub region	Średni wynik
Singapur	556
Japonia	538
Estonia	534
Tajwan	532
Finlandia	531
Makao (Chiny)	529
Kanada	528
Wietnam	525
Hongkong (Chiny)	523
Chiny B-S-J-G	518
Korea	516
Nowa Zelandia	513
Słowenia	513
Australia	510
Wielka Brytania	509
Niemcy	509
Holandia	509
Szwajcaria	506
Irlandia	503
Belgia	502
Dania	502
Polska	501
Portugalia	501
Norwegia	498
Stany Zjednoczone	496
Austria	495
Francja	495
Szwecja	493
Czechy	493
Hiszpania	493
Łotwa	490
Rosja	487
Luksemburg	483
Włochy	481
Węgry	477
Litwa	475
Chorwacja	475
Argentyna	475
Islandia	473
Izrael	467
Malta	465
Słowacja	461
Grecja	455
Chile	447
Bułgaria	446
Emiraty Arabskie	437
Urugwaj	435
Rumunia	435
Cypr	433
Moldawia	428
Albania	427
Turcja	425
Trinidad i Tobago	425
Tajlandia	421
Kostaryka	420
Katar	418
Kolumbia	416
Meksyk	416
Czarnogóra	411
Gruzja	411
Jordania	409
Indonezja	403
Brazylia	401
Peru	397
Liban	386
Tunezja	386
Macedonia	384
Kosowo	378
Algieria	376
Dominikana	332

Tabela 5.	PISA 2015 a 2006		PISA 2015 a 2012	
	Kraj lub region	Wynik 2006	Zmiana od 2006	Wynik 2012
OECD	498	-5	501	-8
Katar	349	68	384	34
Argentyna	391	41	406	27
Kolumbia	388	28	399	17
Portugalia	474	27	489	12
<i>Makao (Chiny)</i>	511	18	521	8
Rumunia	418	16	439	-4
Izrael	454	13	470	-4
Norwegia	487	12	495	4
Bułgaria	434	12	446	-1
Indonezja	393	10	382	21
Brazylia	390	10	402	-1
Chile	438	9	445	2
Japonia	531	7	547	-8
Rosja	479	7	486	0
Urugwaj	428	7	416	20
Dania	496	6	498	3
Meksyk	410	6	415	1
Włochy	475	5	494	-13
Polska	498	4	526	-24
Hiszpania	488	4	496	-4
Estonia	531	3	541	-7
Turcja	424	2	463	-38
Łotwa	490	1	502	-12
Tunezja	386	1	398	-12
Tajwan	532	0	523	9
Francja	495	0	499	-4
Tajlandia	421	0	444	-23
Czarnogóra	412	0	410	1
Luksemburg	486	-4	491	-8
Korea	522	-6	538	-22
Słowenia	519	-6	514	-1
Wielka Brytania	515	-6	514	-5
Szwajcaria	512	-6	515	-10
Irlandia	508	-6	522	-19
Stany Zjednoczone	489	-6	497	-1
Kanada	534	-7	525	2
Niemcy	516	-7	524	-15
Belgia	510	-8	505	-3
Szwecja	503	-10	485	9
Litwa	488	-13	496	-20
Jordania	422	-13	409	-1
Holandia	525	-16	522	-13
Austria	511	-16	506	-11
Nowa Zelandia	530	-17	516	-2
Australia	527	-17	521	-12
Chorwacja	493	-18	491	-16
Islandia	491	-18	478	-5
<i>Hongkong (Chiny)</i>	542	-19	555	-32
Grecja	473	-19	467	-12
Czechy	513	-20	508	-15
Węgry	504	-27	494	-18
Słowacja	488	-28	471	-10
Finlandia	563	-33	545	-15
Liechtenstein	522	*	525	*
Singapur	*		551	4
Wietnam	*		528	-4

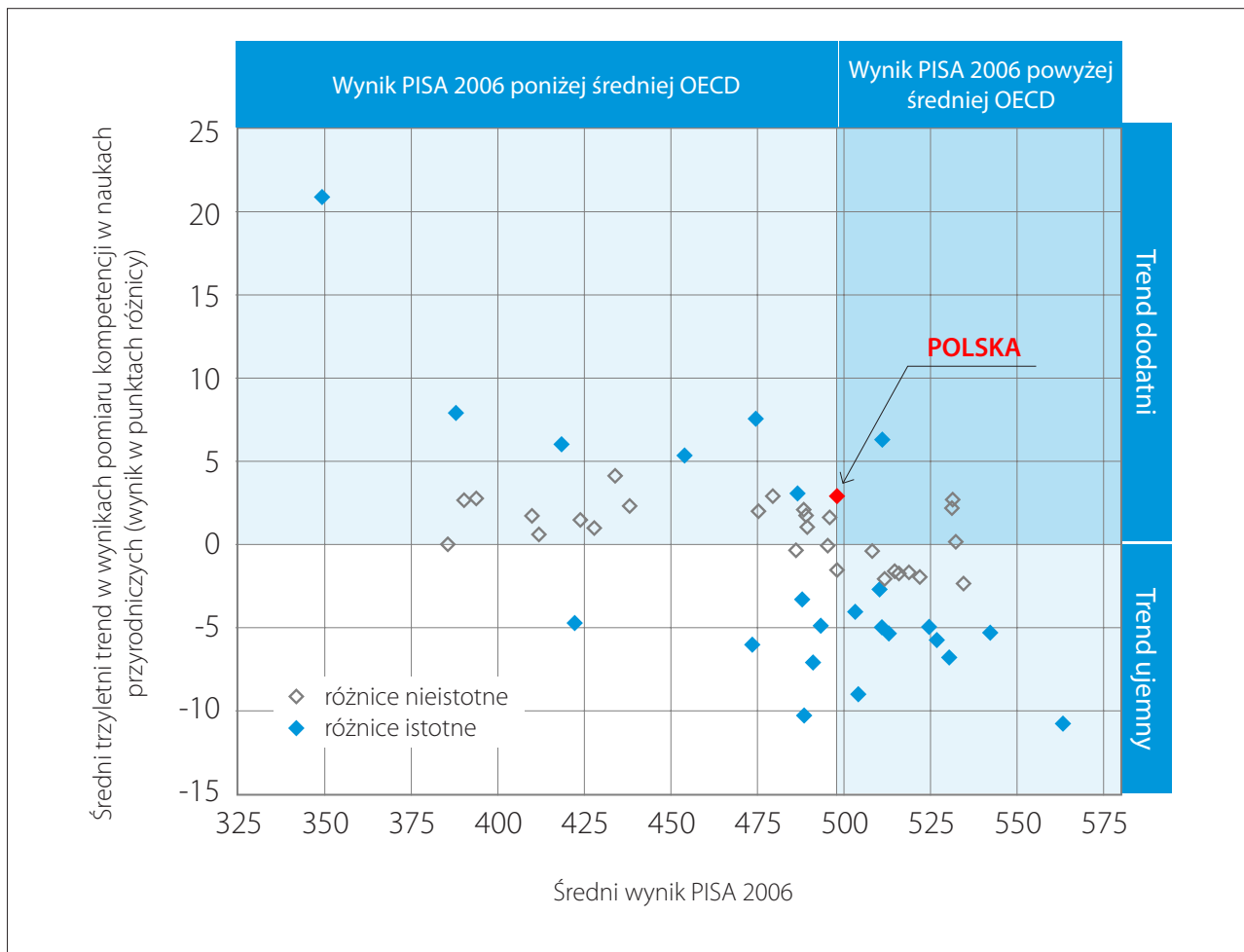
Tabela 5. Różnice wyników uczniów z pomiaru rozumowania w naukach przyrodniczych między badaniem 2015 a 2006 i 2012. Kraje uporządkowano według malejącej różnicy między badaniami 2006 i 2015, w których kompetencje w naukach przyrodniczych były główną domeną pomiaru. Czerwoną czcionką zaznaczono różnice istotne statystycznie. Pogubioną czcionką wyróżniono kraje europejskie, a kursywą – regiony Chin. Gwiazdką oznaczono brak danych.

Porównanie wyników badania z 2015 r. z tymi z lat 2006 r. i 2012 r. dla poszczególnych krajów przedstawiono w Tabeli 5. Można zauważyć, że w pierwszym badaniu komputerowym, jakim było badanie 2015, średni wynik dla krajów OECD był niższy niż w obu badaniach przeprowadzonych z wykorzystaniem zadań drukowanych na papierze. Jednak tylko różnica między latami 2012 i 2015 była istotna statystycznie. Zaledwie 6 krajów odnotowało statystycznie istotny wzrost wyników między latami 2006 a 2012, natomiast aż 14 krajów zanotowało statystycznie istotny spadek. W tej ostatniej grupie znalazły się również kraje-liderzy badania z 2006 r.: Finlandia, Hongkong, Nowa Zelandia i Australia. Wyniki pozostałych krajów, w tym większości krajów europejskich nie różnią się jednak istotnie między badaniami 2006 i 2015 (Estonia: 3 pkt, Słowenia: -6 pkt, Polska: 4 pkt, Irlandia: -6 pkt).

Znaczące różnice można zaobserwować między wynikami badań 2015 i 2012. Dla krajów OECD średni wynik spadł o 8 pkt i była to zmiana statystycznie istotna. Osiem krajów istotnie polepszyło swoje wyniki, ale aż 24 pogorszyły. Liderem spadków była Turcja (-38 pkt), bardzo dużą zmianę odnotowano w Polsce (-24 pkt), w Irlandii (-18 pkt) i Finlandii (-15). Wysoki spadek wyniku pomiaru umiejętności uczniów stwierdzono również w Hongkongu (-32 pkt) i Korei (-22 pkt).

W ocenie systemu edukacji istotne jest nie tylko samo porównanie wyników kolejnych badań, ale także oszacowanie długoterminowych tendencji zmian. Kompetencje polskich 15-latków w naukach przyrodniczych, mimo znacznego obniżenia w latach 2012-2015, na przestrzeni pomiaru 2006-2015 wykazują wyraźną dodatnią tendencję wzrostową (Wykres 2). Średni trzyletni trend w wynikach pomiaru kompetencji przyrodniczych wynosi 2,9.

Wykres 2. Zależność między średnim trzyletnim trendem w wynikach pomiaru kompetencji w naukach przyrodniczych a średnim wynikiem uczniów w badaniu 2006.



Poziomy umiejętności w pomiarze kompetencji uczniów w naukach przyrodniczych

Średni wynik uczniów w PISA służy porównaniom międzynarodowym lub między cyklami badania. W tym celu przydatny jest rozkład wyników, czyli informacja o odsetku uczniów na poszczególnych poziomach umiejętności (poziomy zdefiniowano w Tabeli 6.). Ważne są zwłaszcza zmiany na poziomie 1 i poniżej tego poziomu oraz na poziomie 5 i powyżej. Pierwsza kategoria badanych to uczniowie, nie tylko nierozumiejący otaczającego świata, ale i niezdolni do funkcjonowania w nowoczesnym społeczeństwie, w którym zagadnienia nauki i techniki odgrywają coraz większą rolę. Natomiast uczniowie na poziomach 5 i 6 to potencjalna elita intelektualna, warunkująca postęp naukowo-techniczny. Należy jednak pamiętać, że definicje

poziomów umiejętności w badaniach 2006-2012 oraz 2015 nie do końca się pokrywają, ponieważ ostatnie badanie wykorzystywało już nieco inny zestaw zadań.

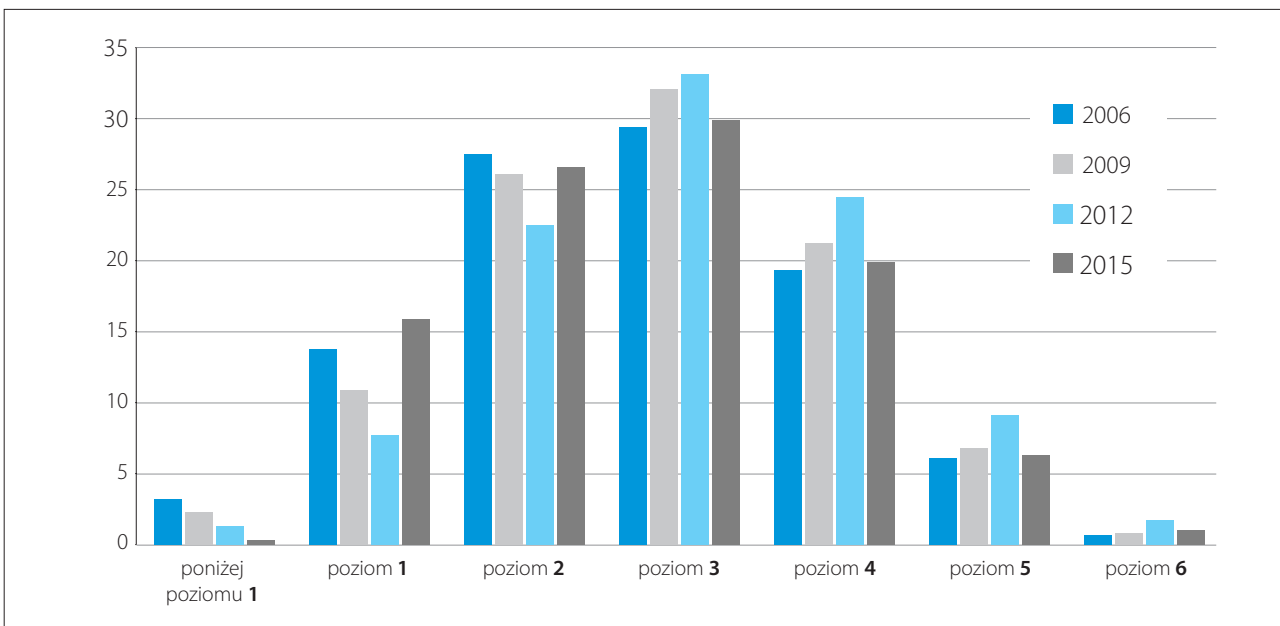
W 2015 roku odsetek uczniów w Polsce na poziomie 1 i 2 wzrósł, a na poziomach od 3-6 spadł w odniesieniu do 2012 roku (Wykres 3.). Ogółem, rozkład wyników z 2015 r. jest zbliżony do wyników z 2006 r. Porównując wyniki dla Polski i OECD (Wykres 4.), można odnotować, że w Polsce jest wyższy odsetek uczniów na poziomach 2-4, a niższy na pozostałych. Choć zatem cieszy niższy odsetek uczniów najsłabszych, to martwi niewysoki odsetek uczniów najwyższych poziomach umiejętności.

Należy jednak zauważyć, że Polska należała do nielicznych krajów o niskich różnicach w odsetku uczniów na najniższych i najwyższych poziomach umiejętności między badaniami 2006 i 2015 (Wykres 5.).

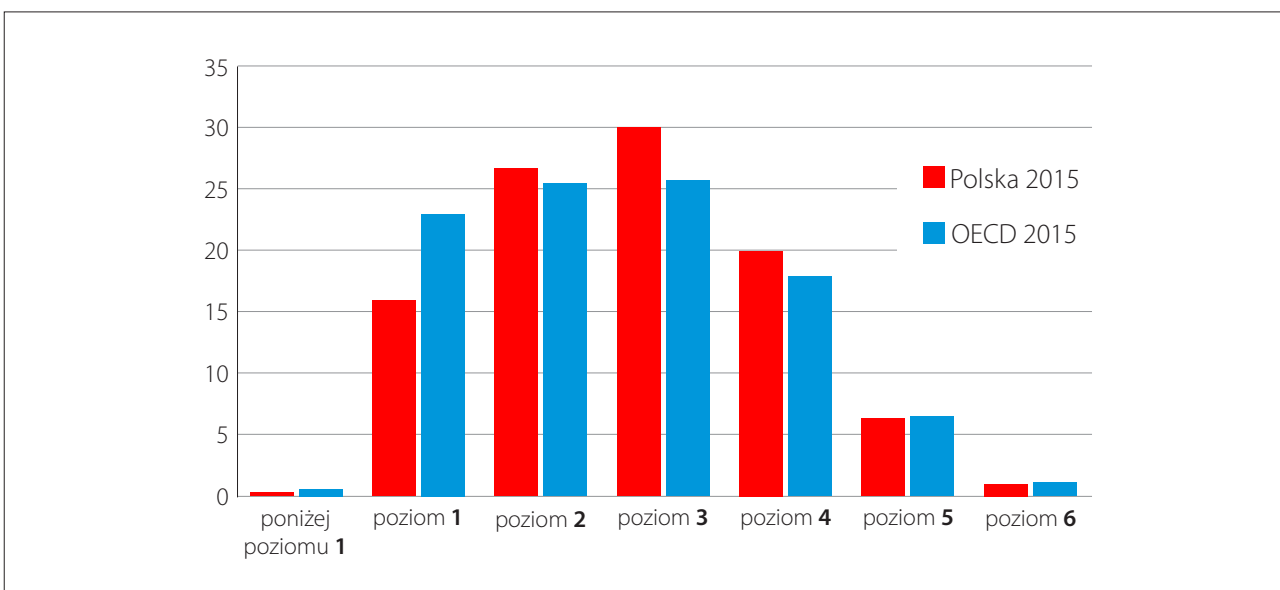
Tabela 6. Opis poziomów umiejętności na skali kompetencji w naukach przyrodniczych

Poziom i dolna granica przedziału	Charakterystyka umiejętności uczniów na podstawie rozwiązanych zadań
<p>6</p> <p>708 pkt.</p>	<p>Uczniowie wykorzystują rozległą wiedzę o treściach naukowych, procedurach badawczych i rozumowaniu naukowym w celu stawiania hipotez wyjaśniających nieznaną im wcześniej zjawiska, zdarzenia i procesy lub formułowania prognoz. Interpretując dane i dowody naukowe, potrafią odróżnić informacje istotne od nieistotnych oraz odwołują się do wiedzy nieobjętej szkolnym programem nauczania. Są w stanie stwierdzić, które argumenty odwołują się do teorii i faktów naukowych, a które bazują na innych rozważaniach. Analizują alternatywne plany badawcze złożonych eksperymentów, obserwacji terenowych i symulacji, potrafią wybrać najlepszy i uzasadnić swój wybór.</p>
<p>5</p> <p>633 pkt.</p>	<p>Uczniowie potrafią wykorzystywać abstrakcyjne pojęcia lub idee naukowe, aby wyjaśnić nieznaną sobie wcześniej zjawiska, zdarzenia lub procesy z wieloma zależnościami przyczynowo-skutkowymi. Wykorzystują bardziej wyrafinowaną wiedzę poznawczą, aby porównać alternatywne schematy eksperymentów i uzasadnić ich wybór. Posługują się wiedzą teoretyczną w celu interpretacji informacji oraz przewidzenia wyniku. Potrafią ocenić różne sposoby naukowego podejścia do tego samego problemu badawczego oraz określić ograniczenia związane z interpretacją danych naukowych, w tym źródła i skutki niepewności.</p>
<p>4</p> <p>559 pkt.</p>	<p>Uczniowie wykorzystują bardziej złożone lub abstrakcyjne treści naukowe (podane w zadaniu lub przywołane z pamięci) w celu wyjaśniania bardziej złożonych lub mniej znanych zdarzeń i procesów. Potrafią przeprowadzić doświadczenie z dwoma lub więcej niezależnymi zmiennymi, ale w ograniczonym zakresie. Są w stanie uzasadnić plan eksperymentu, odwołując się do elementów wiedzy o procedurach i rozumowaniu naukowym. Interpretują dane pozyskane z umiarkowanie złożonego zbioru danych albo dotyczące mniej znanego kontekstu, wyciągają odpowiednie, ogólne wnioski, przedstawiają uzasadnienie swojego wyboru.</p>
<p>3</p> <p>484 pkt.</p>	<p>Uczniowie wskazują lub tworzą wyjaśnienia znanych zjawisk, odwołując się do umiarkowanie złożonej wiedzy faktograficznej. W sytuacjach mniej znanych lub bardziej złożonych potrafią podać wyjaśnienie, jeśli uzyskają odpowiednią podpowiedź lub pomoc. Wykorzystując elementy wiedzy proceduralnej i poznawczej, potrafią przeprowadzić prosty eksperyment. Rozróżniają zagadnienia naukowe od nienaukowych oraz wskazują dane wspierające określoną tezę naukową.</p>
<p>2</p> <p>410 pkt.</p>	<p>Odwołując się do podstawowej, codziennej wiedzy faktograficznej i z zakresu procedur badawczych, uczniowie wskazują odpowiednie wyjaśnienie naukowe, interpretują dane oraz określają pytanie badawcze w prostym doświadczeniu. Wykorzystując podstawową, codzienną wiedzę naukową, potrafią wskazać poprawny wniosek wynikający z prostego zbioru danych. Wykazują się podstawową wiedzą poznawczą, wskazując pytania, na które można odpowiedzieć w sposób naukowy.</p>
<p>1a</p> <p>335 pkt.</p>	<p>Uczniowie potrafią wykorzystać podstawowe wiadomości z codziennego życia oraz znajomość procedur badawczych do rozpoznania wyjaśnienia prostego zjawiska. Z odpowiednią pomocą podejmują problemy badawcze z nie więcej niż dwoma zmiennymi. Potrafią zidentyfikować proste zależności przyczynowo-skutkowe i korelacje oraz interpretować graficzne przedstawienia danych, które nie stawiają wysokich wymagań poznawczych. Potrafią dobrać najlepsze wyjaśnienie naukowe dla określonych danych w znanym już kontekście osobistym, lokalnym lub globalnym.</p>
<p>1b</p> <p>261 pkt.</p>	<p>Uczniowie dysponują podstawową, potoczną wiedzą i wykorzystują ją jedynie do rozpoznania znanych sobie lub prostych zjawisk przyrodniczych. Potrafią dostrzec proste wzorce w danych, znają podstawowe terminy naukowe, przeprowadzają procedury naukowe na podstawie jednoznacznych poleceń.</p>

Wykres 3. Odsetki uczniów na poszczególnych poziomach umiejętności rozumowania w naukach przyrodniczych w Polsce w latach 2006, 2009, 2012 i 2015



Wykres 4. Odsetki uczniów na poszczególnych poziomach umiejętności rozumowania w naukach przyrodniczych w Polsce i OECD w 2015 r.



Wyniki chłopców i dziewcząt

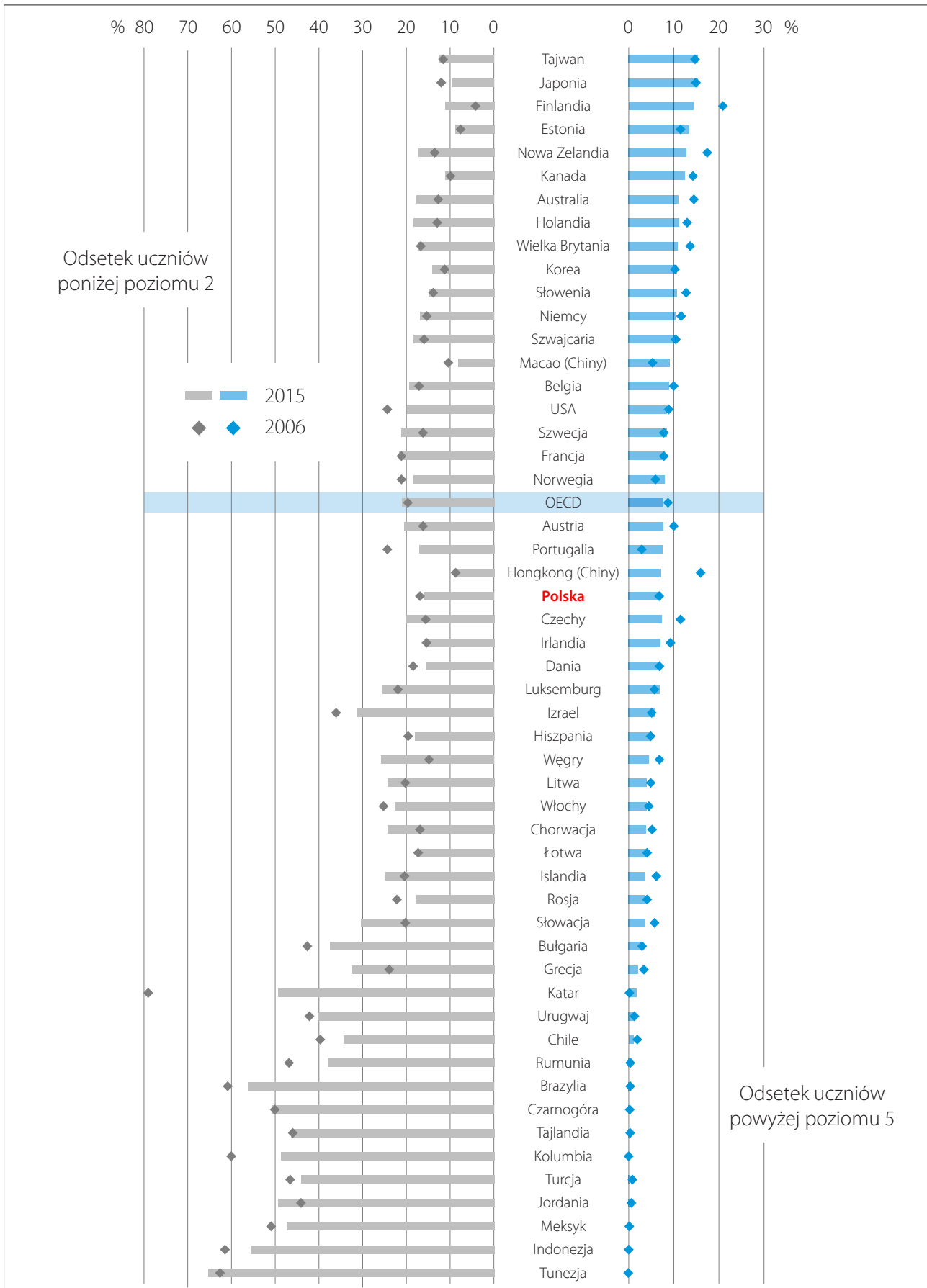
W 2015 r. średni wynik chłopców w krajach OECD wyniósł 495 punktów i był o 4 pkt. wyższy od średniego wyniku dziewcząt (Wykres 5). W Polsce średni wynik dziewcząt wyniósł 498 pkt., podczas gdy chłopcy osiągnęli 504 pkt. Zarówno w wynikach OECD, jak i Polski, różnica między chłopcami a dziewczętami była istotna statystycznie. Odsetek chłopców i dziewcząt na poziomie poniżej 2 był podobny i wynosił około 16%; był przy tym niższy od średniej OECD. Odsetek chłopców na poziomie powyżej 4 był natomiast

istotnie wyższy niż odsetek dziewcząt na tym poziomie, zarówno w Polsce, jak i średnio w krajach OECD.

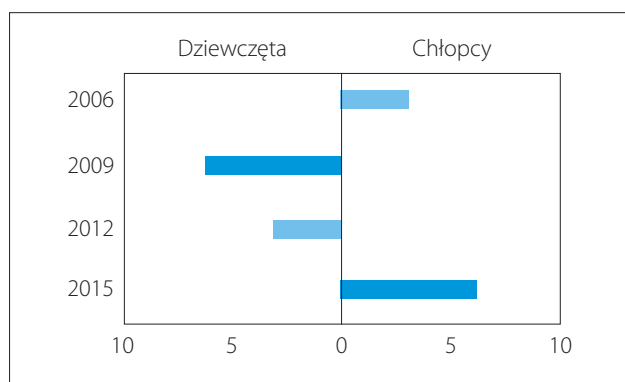
Na Wykresie 6. przedstawiono różnicę między wynikami chłopców i dziewcząt w kolejnych badaniach PISA. W 2006 r. nieco wyższy wynik osiągnęli chłopcy, ale różnica była nieistotna statystycznie. W 2009 r. przewaga dziewcząt była wyraźna (i istotna statystycznie), ale spadła w badaniu 2012 r., a różnica między płciami nie była istotna. W 2015 r. różnica stała się ponownie istotna, tym razem na korzyść chłopców.



Wykres 5. Odsetki uczniów na poziomach 1 i poniżej oraz 5 i powyżej w latach 2006 i 2015 w pomiarze umiejętności rozumowania w naukach przyrodniczych PISA.



Wykres 6. Różnice w wynikach dziewcząt i chłopców w pomiarze rozumowania w naukach przyrodniczych w latach 2006-2015. Statystycznie istotne różnice zaznaczone są ciemniejszą barwą.



Wyniki uczniów w podziale na rodzaj umiejętności i rozwiązywalność poszczególnych zadań

Jak opisano powyżej, kompetencje w zakresie nauk przyrodniczych obejmują trzy umiejętności: (1) wyjaśnianie zjawisk w sposób przyrodniczy, (2) planowanie i ocena poprawności procedur badawczych, (3) interpretacja danych i dowodów naukowych. W kształtowaniu tych umiejętności niezbędne są trzy kategorie wiedzy: o treściach nauki, o procedurach badawczych (o metodyce badawczej) i o podstawach poznania naukowego. Dla wyjaśniania zjawisk przyrodniczych w sposób naukowy niezwykle istotna jest wiedza o treściach naukowych, przy planowaniu i ocenie procedur badawczych ważna jest znajomość tych procedur, natomiast wiedza o poznaniu naukowym jest niezbędna w interpretacji danych i dowodów naukowych. Warto zauważyć, że wyniki polskich piętnastolatków są podobne na wszystkich trzech skalach umiejętności, co wskazuje na zrównoważone kształtowanie tych umiejętności w szkole (Tabela 7). Również na trzech podskalach opisujących treści nauki w 2015 roku uczniowie uzyskali podobne wyniki. W badaniu w 2006 r. natomiast umiejętności odwołujące się do wiedzy faktograficznej przeważały nad umiejętnościami wymagającymi rozumowania naukowego.

Tabela 7. Wyniki uczniów z Polski oraz z OECD na skalach umiejętności przyrodniczych oraz pod względem grup treści naukowych w latach 2006 i 2015.

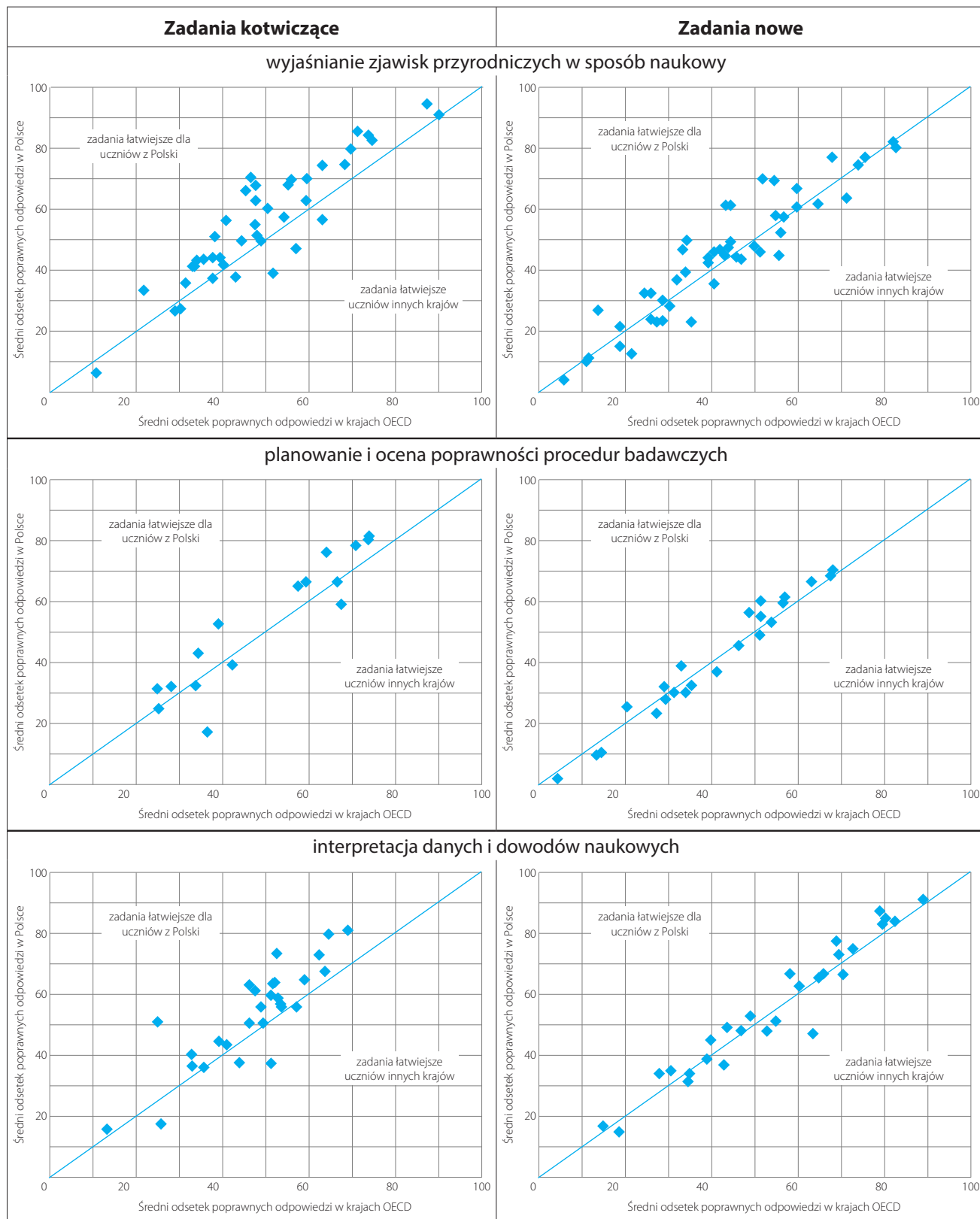
	OECD		Polska	
	2006	2015	2006	2015
Umiejętności				
Wyjaśnianie zjawisk przyrodniczych w sposób naukowy	500	493	506	501
Planowanie i ocena poprawność i procedur badawczych	499	493	483	502
Interpretacja danych i dowodów naukowych	499	493	494	501
Wiadomości – treści nauki				
Świat fizyczny		493	497	503
Świat organizmów żywych		492	501	501
Ziemia i kosmos		494	509	501

Wyniki polskich uczniów w rozwiązywaniu poszczególnych zadań porównano z wynikami, jakie osiągnęli uczniowie z krajów OECD (Wykres 7). Jak widać, polscy uczniowie rozwiązywali większość zadań lepiej od rówieśników z OECD, a różnica nie miała zwykle związku z trudnością zadania ani z kluczową umiejętnością, jaką sprawdzało dane zadanie. Jedynie w wypadku nowych zadań mierzących umiejętności z zakresu planowania i oceny poprawności procedur badawczych (Wykres 8 d) zaobserwowano, że różnica między wynikami OECD a Polski zwiększa się wraz z trudnością zadania (korelacja wyniosła 0,42 i była istotna statystycznie). Znaczy to, że nasi uczniowie mieli większy kłopot z zadaniami o wysokiej trudności niż ich rówieśnicy z OECD.

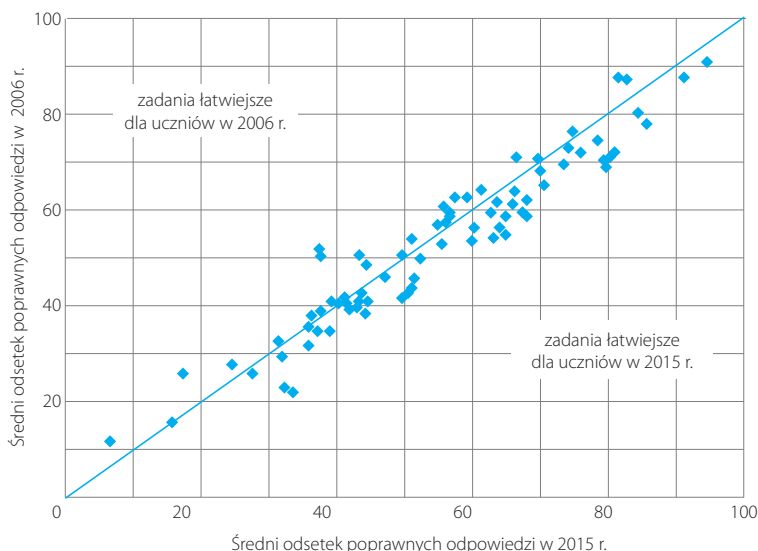
W badaniu PISA 2015 wykorzystano 79 zadań z 2006 roku. W Polsce w 2015 roku w porównaniu z rokiem 2006 procent poprawnych odpowiedzi zwiększył się dla 50 zadań, a zmniejszył dla 29 (Wykres 8).



Wykres 7. Porównanie rozwiązywalności zadań kotwiczących oraz nowych w Polsce oraz w krajach OECD.



Wykres 8. Wyniki polskich uczniów w poszczególnych zadaniach z przedmiotów przyrodniczych w latach 2006 i 2015.



Postawy uczniów wobec nauk przyrodniczych oraz oczekiwania względem kariery zawodowej

W badaniu 2006 oraz 2015 pytano uczniów o różne aspekty ich postaw wobec nauk przyrodniczych, w tym, czy uczenie się przedmiotów przyrodniczych oraz poznawanie zagadnień naukowych sprawia im przyjemność i daje satysfakcję. Na podstawie tych odpowiedzi opracowano wskaźnik satysfakcji z uczenia się przedmiotów przyrodniczych (*enjoyment of learning science*). Krajami, w których wskaźnik ten wzrósł najbardziej między badaniami 2006 i 2015, były Irlandia i Polska (Wykres 9).

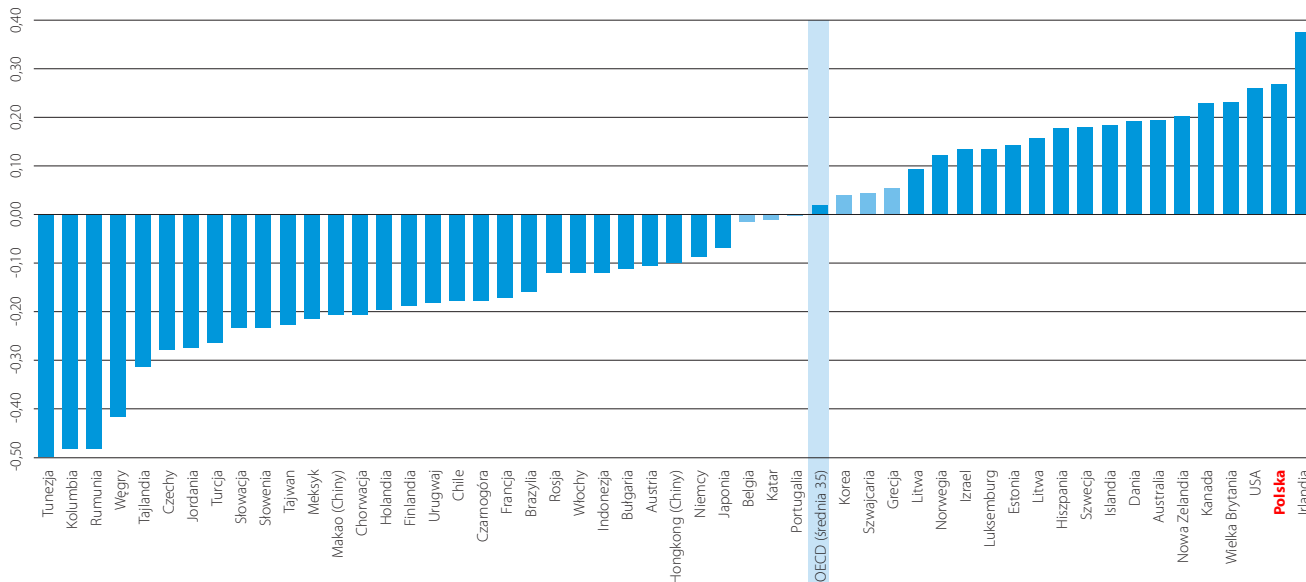
Uczniowie byli też pytani o to, jak sądzą, jaki zawód będą wykonywać w wieku 30 lat. Odsetek uczniów wskazujących na zawody związane z nauką zmniejszył się, w porównaniu z 2006 r. z ok. 26 do 21 procent. Między 2006

a 2015 nie zmienił się odsetek uczniów wskazujących na zawody inżynierskie i bezpośrednio związane z nauką (te zawody wskazało w 2015 r. ok. 6% uczniów), zwiększył się odsetek wskazujących na zawody medyczne (wzrost z 8 do 12%), zmniejszył się natomiast odsetek uczniów wskazujących na zawody informatyczne (z 6 do 1%) oraz pozostałe zawody związane z nauką (z 6 do 1%).

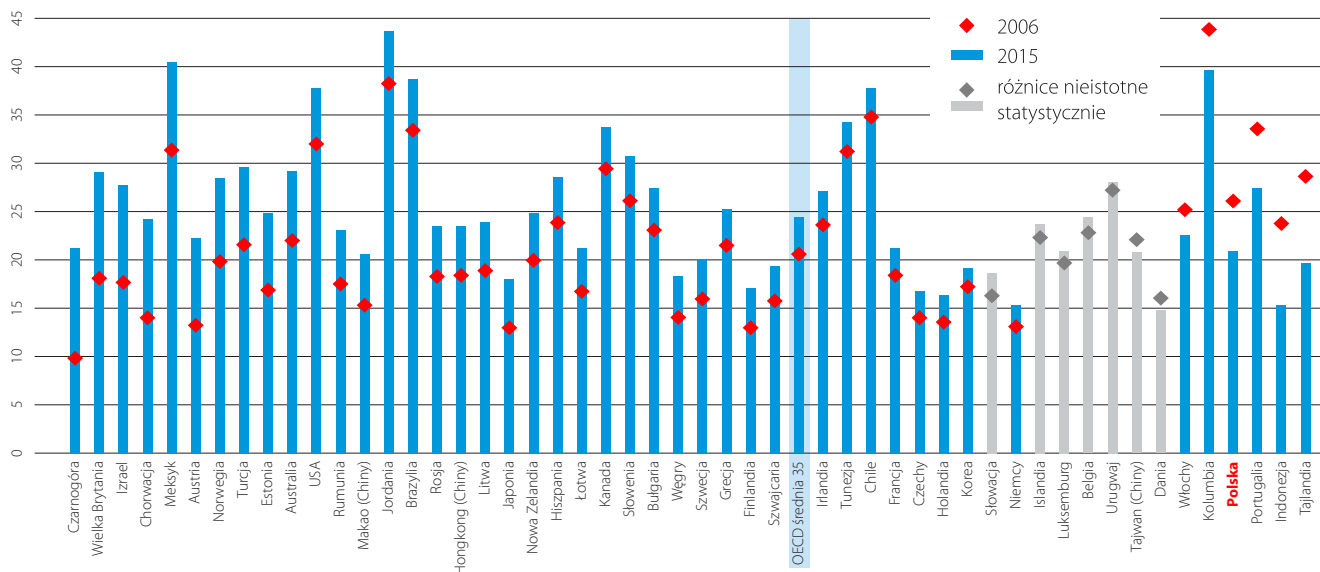
Podsumowanie

W badaniu 2015 polscy uczniowie osiągnęli wynik podobny do uzyskanego w 2006 r., kiedy to kompetencje przyrodnicze były także główną dziedziną pomiaru PISA, ale tym razem był to wynik istotnie powyżej średniej dla OECD. Choć jest to dobra wiadomość, to jednak jest to rezultat znacznie poniżej oczekiwań, do jakich skłaniał bardzo wysoki wynik

Wykres 9. Zmiana wskaźnika satysfakcji z uczenia się o przyrodzie między badaniami PISA 2006 i 2015. Kraje uszeregowano od najmniejszej do największej zmiany tego wskaźnika. Ciemniejszą barwą zaznaczono różnice istotne statystycznie



Wykres 10. Odsetek uczniów w badaniach PISA 2006 i 2015, którzy oczekują, że w wieku 30 lat będą pracować w zawodach związanych z naukami przyrodniczymi. Kraje uszeregowano od najwyższej do najniższej wartości różnicy tego odsetka między latami 2015 i 2006. Szarym kolorem oznaczono różnice nieistotne statystycznie.



polskich piętnastolatków w badaniu 2012. Rozważając tę rozbieżność, należy pamiętać, że w latach 2006-2012 wykorzystywano papierowe zeszyty z zadaniami, natomiast w 2015 r. badanie zrealizowane było za pomocą komputerów. We wszystkich badaniach w latach 2006-2015 wykorzystano te same zadania kotwiczące, ale w 2015 r. zostały one zaadaptowane do rozwiązywania na komputerze, natomiast nowe zadania nie miały już odpowiednika w formie papierowej. Sugeruje to, że głównym problemem polskich uczniów było właśnie rozwiązywanie zadań z wykorzystaniem komputera. Do takiego wyjaśnienia skłaniają też wyniki badania PISA 2012 dotyczącego twórczego rozwiązywania problemów (*Creative Problem Solving*) z wykorzystaniem komputera. Zostało ono przeprowadzone równoległe do badania głównego w 2012 r. W badaniu głównym w zakresie umiejętności rozumowania w naukach przyrodniczych polscy uczniowie uzyskali średni wynik 526 pkt, plasujący ich w czołówce światowej. Natomiast w badaniu rozwiązywania problemów z wykorzystaniem komputera osiągnęli wynik zaledwie 481 pkt (przy średniej OECD 500 pkt), czyli znacznie poniżej oczekiwań.

Niepokój budzi także analiza rozkładu poziomów umiejętności uczniów. W latach 2006-2012 stopniowo malał odsetek uczniów na najniższych poziomach, a rósł na najwyższych. Natomiast w pierwszym badaniu komputerowym rozkład ten jest podobny, jak w 2006 r. W porównaniu z OECD może nas cieszyć niższy odsetek uczniów najsłabszych, jednak niepokoi niewysoki odsetek uczniów na najwyższych poziomach umiejętności, stanowiących przyszłą elitę naukowo-techniczną.

Przejsie z badania papierowego na komputerowe ujawniło także problem różnic między płciami – w 2015 r. chłopcy byli

o 6 pkt istotnie lepsi od dziewcząt, podczas gdy w ostatnim badaniu papierowym (2012 r.) różnica między płciami była nieistotna statystycznie, a średni wynik dziewcząt był o 3 pkt wyższy od wyniku chłopców.

Pozytywnym wynikiem badania 2015 jest to, że Polska jest w grupie krajów o dodatniej średniej długoterminowej tendencji zmian. Pokazuje to stałą poprawę poziomu umiejętności polskich 15-latków. Dalsze badania pokażą, czy jest to tendencja trwała.

Bardzo dobrym rezultatem są także wyniki pomiaru składowych umiejętności w zakresie kompetencji przyrodniczych. W 2006 r. badania pokazały, że polscy uczniowie cechują się ponadprzeciętną wiedzą o treściach naukowych, natomiast słabiej wypadali w zakresie umiejętności wykorzystywania tej wiedzy do wnioskowania, planowania doświadczeń czy w interpretacji faktów naukowych. Natomiast w badaniu 2015 wszystkie komponenty biegłości naukowej były na podobnym poziomie.

Bardzo dobrą wiadomością płynącą z badań jest znaczący wzrost satysfakcji uczniów z nauki przedmiotów przyrodniczych, co jest niezwykle ważne w świetle coraz częstszej i obecnej nawet w mediach publicznych postawy antynaukowej, nieopartej na faktach, ale bazującej na irracjonalnych uprzedzeniach, np. sprzeciw wobec szczepionek i organizmów genetycznie zmodyfikowanych albo zaprzeczanie, że ocieplenie się klimatu stanowi problem globalny. Niepokoi natomiast spadek oczekiwań względem kariery zawodowej w nauce – najwyraźniej uczniowie w coraz mniejszym stopniu uznają taki rozwój zawodowy za atrakcyjny.

Omówienie przykładowych zadań

■ Wiązka zadań „Badanie zbocza góry”

Prezentowana wiązka zadań dotyczy problemu konstrukcji układu badawczego oraz interpretacji otrzymanych danych. We wstępie nakreślono plan eksperymentu oraz postawiono problem badawczy. Wiązka ta ma charakter tradycyjny i równie dobrze mogłaby być wykorzystana w papierowej wersji badania.

PISA 2015

?
◀
▶

Badanie zbocza wzgórza
 Wstęp

Przeczytaj wstęp. Następnie kliknij na strzałkę DALEJ.

BADANIE ZBOCZA WZGÓRZA

Grupa uczniów zauważyła znaczną różnicę w stanie roślinności na dwóch zboczach tworzących dolinę: roślinność jest znacznie bardziej zielona i obfita na zboczu A niż na zboczu B. Tę różnicę pokazano na ilustracji po prawej stronie.

Uczniowie badają, dlaczego roślinność na tych dwóch zboczach tak bardzo się od siebie różni. W ramach tego badania uczniowie przez określony czas mierzą trzy czynniki środowiskowe:

- **Promieniowanie słoneczne:** ile światła słonecznego pada na dane miejsce
- **Wilgotność gleby:** jaki jest poziom wilgotności gleby w danym miejscu
- **Opady deszczu:** ile deszczu spada w danym miejscu



Zadanie 1.

Pierwsze zadanie wymaga sformułowania własnej odpowiedzi i sprawdza umiejętność planowania doświadczenia. Wymaga przede wszystkim wiedzy o poznaniu naukowym, a dotyczy bardzo ważnego zagadnienia źródeł niepewności pomiaru w badaniach.

PISA 2015

?
◀ ▶

Badanie zbocza wzgórza

Pytanie 1 / 2

Zapoznaj się z tekstem "Zbieranie danych" po prawej stronie. Wpisz odpowiedź na pytanie.

Dlaczego, badając różnicę w stanie roślinności między dwoma zboczami tworzącymi dolinę, uczniowie umieszczają na każdym zboczach dwie sztuki każdego urządzenia?

BADANIE ZBOCZA WZGÓRZA

Zbieranie danych

Na każdym ze zboczy uczniowie umieszczają dwie sztuki każdego z trzech urządzeń, jak pokazano poniżej.

Czujnik promieniowania słonecznego: mierzy ilość światła słonecznego, w megadžulach na metr kwadratowy (MJ/m²)

Czujnik wilgotności gleby: mierzy ilość wody wyrażonej jako procent masy gleby

Deszczomierz: mierzy ilość opadów deszczu, w milimetrach (mm)

Aby prawidłowo rozwiązać to zadanie, uczeń powinien rozumieć, w jakim celu w badaniach naukowych dokonuje się co najmniej dwóch niezależnych pomiarów określonej wielkości. Za poprawne odpowiedzi uznawano takie, w których uczniowie wykazali znaczenie takiego podejścia. Poniżej podano przykładowe poprawne odpowiedzi, zachowując oryginalne sformułowania uczniów, opatrzone je także komentarzem. Warto zaznaczyć, że oceniana była jedynie poprawność rozumowania, nie zaś forma czy ewentualne błędy językowe.

- „Żeby można było określić, czy różnica między zboczami jest istotna”. Uczeń rozumie, że pojedynczy pomiar może być – z różnych względów – niewiarygodny.
- „Ponieważ mogą występować różnice na tym samym zboczach”. Uczeń zauważa, że mierzona wielkość fizyczna może cechować się zmiennością.
- „Żeby zwiększyć precyzję pomiaru na każdym ze zboczach”. Uczeń wskazuje, że dla poprawnego wnioskowania należy oszacować zmienność cechy.
- „Na wypadek awarii jednego z dwóch”. Uczeń dostrzega, że urządzenia pomiarowe mogą być zawodne.
- „Żeby porównać różne ilości słońca na zboczach”. Jak powyżej – uczeń odwołuje się do lokalnej zmienności mierzonej wielkości.

Zadaniu przypisano 3. poziom umiejętności na skali kompetencji w naukach przyrodniczych z wartością 517 pkt. Zadanie to prawidłowo rozwiązało 60,7% polskich uczniów, podczas gdy średnia dla OECD wyniosła 51,3%.

Zadanie 2.

Drugie zadanie z prezentowanej wiązki dotyczy problemu zrozumienia wpływu niepewności pomiarowej na możliwość interpretacji otrzymanych danych. Wymagało przede wszystkim umiejętności interpretacji danych oraz odwoływało się do wiedzy o poznaniu naukowym. Rozwiązując zadanie, uczeń musiał wybrać jedną z dwóch odpowiedzi, a następnie wyjaśnić, dlaczego wybrał tę odpowiedź.

PISA 2015
?
◀ ▶

Badanie zbocza wzgórza
Pytanie 2 / 2

Zapoznaj się z tekstem "Analiza danych" po prawej stronie. Kliknij, aby wybrać odpowiedź, a następnie wpisz wyjaśnienie, uzasadniając swój wybór.

Dwoje uczniów nie zgadza się co do przyczyn różnicy w poziomie wilgotności gleby między oboma zboczami.

- Uczeń nr 1 uważa, że różnica w poziomie wilgotności gleby wynika z różnicy w promieniowaniu słonecznym na obu zboczach.
- Uczeń nr 2 uważa, że różnica w poziomie wilgotności gleby wynika z różnicy w opadach deszczu na obu zboczach.

Na podstawie danych powiedz, który z uczniów ma rację?

Uczeń nr 1

Uczeń nr 2

Wyjaśnij swoją odpowiedź.

BADANIE ZBOCZA WZGÓRZA
Analiza danych

Uczniowie obliczają średnie wielkości pomiarów zebranych w danym okresie z każdej pary urządzeń umieszczonych na każdym ze zboczy i obliczają niepewność tych średnich. Uzyskane przez nich wyniki zapisano w poniższej tabeli. Niepewność została zapisana po znaku "±".

	Średni poziom promieniowania słonecznego	Średni poziom wilgotności gleby	Średni poziom opadów deszczu
Zbocze A	3 800 ± 300 MJ/m ²	28 ± 2%	450 ± 40 mm
Zbocze B	7 200 ± 400 MJ/m ²	18 ± 3%	440 ± 50 mm

Aby wynik pomiaru mógł służyć za podstawę do dalszego wnioskowania, niezbędna jest znajomość dokładności, z jaką został on wykonany. Analiza i interpretacja wyników to jeden z kluczowych etapów badania naukowego. W przedstawionym zadaniu uczniowie mają przeanalizować dane dotyczące trzech parametrów charakteryzujących badane zbocza: średniego poziomu promieniowania słonecznego, średniej wilgotności gleby oraz średniego poziomu opadów deszczu. Wartości mierzonych parametrów zostały zestawione w tabeli razem z niepewnościami pomiarowymi średnich. Aby wybrać i uzasadnić prawidłową odpowiedź, niezbędne jest zrozumienie, że pomiar jest nierozdzielnie związany z niepewnością pomiarową.

Rozwiązując zadanie, badany piętnastolatek powinien dostrzec, że rację ma „uczeń nr 1”, ponieważ błąd pomiaru poziomu promieniowania słonecznego jest niski w porównaniu z różnicą między średnimi, co pozwala wnioskować, że różnica jest rzeczywista. Natomiast różnica między średnim poziomem opadów deszczu na obu zboczach jest wielokrotnie mniejsza niż błąd pomiaru – czego prawdopodobnie nie zauważył uczeń nr 2. Za odpowiedzi poprawne uznawano zatem takie, w których badani uczniowie wybrali odpowiedź „Uczeń 1”, a w uzasadnieniu napisali np.:

- „Na zbocze B dociera o wiele więcej promieniowania słonecznego niż na zbocze A, za to ilość deszczu jest ta sama.”
- „Nie ma różnicy w ilości deszczu pomiędzy oboma zboczami.”
- „Jest duża różnica między ilością światła słonecznego docierającego na zbocze A w porównaniu do zbocza B.”

Zadanie to prawidłowo rozwiązało 31,5 % polskich uczniów, nieco mniej niż średnio w OECD (34, 7%). Na skali trudności zadań uzyskało ono 589 pkt. (poziom 4).

Wiązka zadań „Bieganie podczas upału”

Omawiana wiązka zadań jest jednym z narzędzi interaktywnych, które stanowią nową formułę pomiaru kompetencji w naukach przyrodniczych w badaniu PISA. Wiązka *Bieganie podczas upału* dotyczy zagadnienia dociekania naukowego związanego z problemem termoregulacji u biegaczy długodystansowych, biegających w zmiennych warunkach pogodowych. Wyniki symulacji przedstawione w wiązce oparte są o uproszczony model matematyczny funkcjonowania organizmu konkretnej osoby po biegu trwającym jedną godzinę w różnych warunkach. Ta wiązka pytań występowała tylko w badaniu pilotażowym.

PISA 2015

Bieganie podczas upału
Wstęp

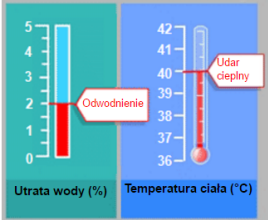
Przeczytaj wstęp. Następnie kliknij na strzałkę DALEJ.

BIEGANIE PODCZAS UPALU

W czasie biegów na długie dystanse temperatura ciała podnosi się i występuje pocenie.

U biegaczy, którzy nie piją dostatecznych ilości wody, aby uzupełnić wodę utraconą przez pocenie się, może wystąpić odwodnienie. Uznaje się, że stan odwodnienia występuje przy utracie wody stanowiącej 2% lub więcej masy ciała. Ta wielkość procentowa została oznaczona na pokazanej niżej podziałce wskazującej utratę wody.

Jeżeli temperatura ciała podniesie się do 40°C lub wyżej, u biegaczy może wystąpić stan powodujący zagrożenie życia, zwany udarem cieplnym. Temperaturę tę oznaczono na pokazanym niżej termometrze temperatury ciała.



Utrata wody (%)	Temperatura ciała (°C)
0	36
1	37
2	38
3	39
4	40
5	41
	42

Po przeczytaniu wstępu uczniowie wykonują wstępną symulację, zapoznając się z działaniem modelu i ze sposobem zbierania danych.

PISA 2015
🔍
⏪
⏩

Bieganie podczas upału

Wstęp

Symulacja opiera się na modelu, w którym oblicza się objętość potu, utratę wody oraz temperaturę ciała biegacza po biegu trwającym jedną godzinę.

Aby zobaczyć, jak działają poszczególne ustawienia w tej symulacji, wykonaj poniższe kroki:

- Przesuń suwak **temperatury powietrza**.
- Przesuń suwak **wilgotności powietrza**.
- Przy **picciu wody** kliknij na odpowiedź Tak lub na odpowiedź Nie.
- Kliknij na przycisk "Start", aby zobaczyć wyniki. Zauważ, że utrata wody stanowiąca 2% lub więcej powoduje odwodnienie, a temperatura ciała wynosząca 40°C lub więcej powoduje udar cieplny. Wyniki zostaną również wyświetlone w tabeli.

Uwaga: Wyniki symulacji są oparte na uproszczonym modelu matematycznym funkcjonowania organizmu konkretnej osoby po biegu trwającym jedną godzinę w różnych warunkach.

Temperatura powietrza (°C)

Wilgotność powietrza (%)

Picie wody Tak Nie

Start

Temperatura powietrza (°C)	Wilgotność powietrza (%)	Picie wody	Objętość potu (litry)	Utrata wody (%)	Temperatura ciała (°C)

W przypadku prezentowanego zadania uczniowie mogą modelować funkcjonowanie organizmu człowieka po biegu trwającym jedną godzinę w różnych warunkach – przy zmieniającej się temperaturze i wilgotności powietrza oraz w zależności od tego, czy biegacz pije w czasie biegu wodę czy też nie. Po wyborze parametrów przez ucznia i naciśnięciu przycisku „Start”, w tabeli pojawiają się wartości wybranych parametrów i odpowiadające im wyniki modelowania zachowania się organizmu biegacza: objętość potu, utrata wody oraz temperatura ciała. Wyniki bieżącej symulacji są dodatkowo prezentowane graficznie na górnym schemacie. Jeśli wybrane warunki wywołają stan odwodnienia lub udaru cieplnego, zagrożenie zdrowia lub życia biegacza jest sygnalizowane czerwonymi flagami. Po wykonaniu wstępnej symulacji uczniowie otrzymują informację, czy dobrze ją przeprowadzili.



PISA 2015
?
◀ ▶

Bieganie podczas upału
Wstęp

Symulacja opiera się na modelu, w którym oblicza się objętość potu, utratę wody oraz temperaturę ciała biegacza po biegu trwającym jedną godzinę.

Aby zobaczyć, jak działają poszczególne ustawienia w tej symulacji, wykonaj poniższe kroki:

- Przesuń suwak **temperatury powietrza**.
- Przesuń suwak **wilgotności powietrza**.
- Przy **picciu wody** kliknij na odpowiedź Tak lub na odpowiedź Nie.
- Kliknij na przycisk "Start", aby zobaczyć wyniki. Zauważ, że utrata wody stanowiąca 2% lub więcej powoduje odwodnienie, a temperatura ciała wynosząca 40°C lub więcej powoduje udar cieplny. Wyniki zostaną również wyświetlone w tabeli.

Uwaga: Wyniki symulacji są oparte na uproszczonym modelu matematycznym funkcjonowania organizmu konkretnej osoby po biegu trwającym jedną godzinę w różnych warunkach.

Objętość potu (litry)

Odwodnienie

Utrata wody (%)

Udar cieplny

Temperatura ciała (°C)

Temperatura powietrza (°C) 20 25 30 35 40

Wilgotność powietrza (%) 20 40 60 Start

Picie wody Tak Nie

Temperatura powietrza (°C)	Wilgotność powietrza (%)	Picie wody	Objętość potu (litry)	Utrata wody (%)	Temperatura ciała (°C)
30	40	Tak	1,2	0,0	39,3

Dojrzeł!

Aby kontynuować, kliknij na strzałkę DALEJ.

PISA 2015
?
◀ ▶

Bieganie podczas upału
Wstęp

Symulacja opiera się na modelu, w którym oblicza się objętość potu, utratę wody oraz temperaturę ciała biegacza po biegu trwającym jedną godzinę.

Aby zobaczyć, jak działają poszczególne ustawienia w tej symulacji, wykonaj poniższe kroki:

- Przesuń suwak **temperatury powietrza**.
- Przesuń suwak **wilgotności powietrza**.
- Przy **picciu wody** kliknij na odpowiedź Tak lub na odpowiedź Nie.
- Kliknij na przycisk "Start", aby zobaczyć wyniki. Zauważ, że utrata wody stanowiąca 2% lub więcej powoduje odwodnienie, a temperatura ciała wynosząca 40°C lub więcej powoduje udar cieplny. Wyniki zostaną również wyświetlone w tabeli.

Uwaga: Wyniki symulacji są oparte na uproszczonym modelu matematycznym funkcjonowania organizmu konkretnej osoby po biegu trwającym jedną godzinę w różnych warunkach.

Objętość potu (litry)

Odwodnienie

Utrata wody (%)

Udar cieplny

Temperatura ciała (°C)

Temperatura powietrza (°C) 20 25 30 35 40

Wilgotność powietrza (%) 20 40 60 Start

Picie wody Tak Nie

Temperatura powietrza (°C)	Wilgotność powietrza (%)	Picie wody	Objętość potu (litry)	Utrata wody (%)	Temperatura ciała (°C)
30	40	Tak	1,2	0,0	39,3

Tak wygląda symulacja, jeśli temperaturę powietrza ustawiono na 30, wilgotność powietrza ustawiono na 40, w przypadku picia wody wybrano ustawienie "Tak", a następnie kliknięto "Start".

Aby kontynuować, kliknij na strzałkę DALEJ.

Zadanie

Zadaniem ucznia jest wskazanie maksymalnej temperatury, w której człowiek może biec przez godzinę przy wilgotności powietrza 40%, nie dostając udaru cieplnego. Uczniowie muszą przeprowadzić symulację dla podanych w pytaniu wartości temperatury powietrza. Udzielają odpowiedzi nie tylko wybierając odpowiednią temperaturę, ale także wskazując dwa wiersze z wynikami, które uzasadniają ten wybór. Na koniec wyjaśniają, w jaki sposób wybrane wiersze z wynikami modelowania uzasadniają wskazaną temperaturę.

Bieganie podczas upału
Pytanie 4 / 5

Jak przeprowadzić symulację

Przeprowadź symulację, aby zebrać dane, opierając się na informacjach podanych poniżej. Kliknij, aby wybrać odpowiedź, zaznacz dane w tabeli, a następnie wpisz wyjaśnienie, uzasadniając swój wybór.

Odpowiedz na podstawie symulacji: Jeżeli wilgotność powietrza wynosi 40%, to jaką jest najwyższa temperatura powietrza, przy której człowiek może biec przez jedną godzinę, nie doznając udaru cieplnego?

20°C
 25°C
 30°C
 35°C
 40°C

Zaznacz w tabeli dwa wiersze z danymi, które uzasadniają wybraną przez Ciebie odpowiedź.

Wyjaśnij, w jaki sposób dane te uzasadniają Twoją odpowiedź.

Temperatura powietrza (°C)	Wilgotność powietrza (%)	Picie wody	Objętość potu (litry)	Utrata wody (%)	Temperatura ciała (°C)

Omawiane zadanie sprawdza umiejętność planowania doświadczenia i wymaga wiedzy o procedurach badawczych. Warto zwrócić uwagę na kontekst zadania, jakim jest zdrowie i choroby człowieka w aspekcie osobistym.

W ocenie wykonania tego zadania uwzględniono zarówno odpowiedzi całkowicie poprawne, jak i częściowo poprawne. Na w pełni poprawną odpowiedź składały się:

wybór temperatury 35 °C, wskazanie wierszy z symulacjami dla wilgotności powietrza 40% i temperatur 35 °C oraz 40 °C oraz wyjaśnienie, które wskazuje lub sugeruje, że najwyższa temperatura powietrza niegrożąca udarem cieplnym wynosi 35 °C, ponieważ temperatura powietrza 40 °C taki udar już wywołuje. Oto przykłady takich wyjaśnień:

- „Kiedy temperatura na zewnątrz rośnie od 35 do 40 °C, temperatura ciała wzrasta powyżej 40 °C, co wywołuje u biegacza udar cieplny.”
- „Przy 40% wilgotności bieganie w temperaturze 40 °C powoduje udar cieplny, ale przy 35 stopniach temperatura ciała biegacza jest poniżej poziomu wystąpienia udaru cieplnego.”
- „Kiedy zwiększa się temperaturę powietrza, przy 40 stopniach biegacz doznaje udaru cieplnego.”
- „Kiedy poziom wilgotności wynosi 40%, u biegacza udar występuje dopiero przy 40 °C. 35 stopni to druga najwyższa temperatura.”
- „40 °C udar cieplny, nie 35.”

Za częściowo poprawne odpowiedzi uznano takie, w niepoprawny był jeden element: wyjaśnienie było niejasne bądź błędne albo nie były wybrane prawidłowe wiersze, albo wskazana była temperatura 40 stopni przy poprawnych pozostałych elementach.

W badaniach pilotażowych zadanie uzyskało 592 punkty na skali trudności, co plasuje je na 4. poziomie.



CZYTANIE I INTERPRETACJA

Czytanie należy do podstawowych umiejętności, w które młodego człowieka wyposaża szkoła. Bez tej umiejętności niemożliwy jest jakikolwiek postęp w pozyskiwaniu wiedzy. Gdy jednak mówimy o czytaniu, możemy mieć na myśli różne sprawności. W sensie podstawowym czytanie oznacza proste dekodowanie tekstu, czyli zrozumienie zawartego w nim przekazu. W szerokim znaczeniu czytanie obejmuje jednak szereg operacji intelektualnych, które pozwalają dotrzeć do głębokich warstw tekstu. Najczęściej operacje te określane są jako interpretacja. Dlatego, żeby uniknąć nieporozumienia, w niniejszym raporcie badana umiejętność określana jest jako **czytanie i interpretacja**.

W latach 2000 oraz 2009 czytanie i interpretacja stanowiły w badaniu PISA główną dziedzinę. W 2000 r. użyto 140 zadań, a w 2009 – 99 zadań. W latach 2003 oraz 2006 do pomiaru tej umiejętności wykorzystano 28 zadań, a w 2012 – 45 zadań. W 2015 – wykorzystano 96 zadań z poprzednich edycji. W latach 2000, 2003, 2006, 2009 oraz 2012 uczniowie rozwiązywali zadania w tradycyjnej formie na papierowych arkuszach – umiejętności czytania tekstów w formie elektronicznej poświęcono osobne, uzupełniające części komputerowe. W 2015 roku w Polsce i w większości innych krajów do pomiaru umiejętności czytania i interpretacji wykorzystane zostały wyłącznie komputery.

Założenia teoretyczne badania

W badaniu PISA umiejętność czytania oznacza zrozumienie pisanych tekstów (w różnej formie i na różnych nośnikach), poddanie refleksji, wykorzystanie do różnych celów (np. pogłębienia wiedzy czy działania w życiu społecznym) oraz zaangażowanie czytelnika w niesione przez nie treści.

Zadania sprawdzają umiejętność czytania w trzech obszarach: **wyszukiwania informacji, interpretacji tekstu oraz refleksji i oceny**. W badaniu z roku 2015 wyszukiwaniu informacji poświęcone było 21 zadań, interpretacji 53 zadania, refleksji i ocenie 22 zadania. W obszarze wyszukiwania informacji uczeń miał odpowiedzieć na pytania dotyczące tego, co w tekście wyrażone jest wprost. W obszarze interpretacji zadaniem ucznia było wydobycie sensu, który wynika z całości wypowiedzi, dostrzeżenie nieoczywistych związków między częściami tekstu, wyciągnięcie wniosków dotyczących znaczeń wynikających ze sposobu wyrażenia przekazu. W obszarze refleksji i oceny uczeń miał zestawiać informacje zawarte w tekście z wiedzą czerpaną z innych źródeł, w tym celu powinien posłużyć się związłą i trafną argumentacją.

W 2015 r. uczniowie czytali teksty i udzielali odpowiedzi na komputerze. Trudno jednoznacznie stwierdzić, jaki mógł być wpływ tej zmiany na wyniki, najprawdopodobniej nie była ona jednak obojętna, gdyż medium oddziałuje na sposób odbioru przekazu i wykonywane operacje, poza tym wykorzystanie do badania komputera wymagało od uczniów przynajmniej elementarnej biegłości w wykorzystaniu technologii cyfrowej.

Wyniki polskich uczniów na tle innych krajów

W roku 2015 polscy uczniowie w badaniu PISA w dziedzinie czytania i interpretacji uzyskali średnio 506 punktów. Ich wynik był wyższy od średniego wyniku krajów OECD o 13 punktów (w 2012 roku był wyższy o 22 punkty). W porównaniu do innych krajów świata Polska znalazła się na wysokim, trzynastym miejscu; w Europie, jeśli uwzględnić błąd statystyczny, lepszy średni wyniki osiągnęli tylko uczniowie

Kraj/region	Średni wynik
Singapur	535
Hongkong (Chiny)	527
Kanada	527
Finlandia	526
Irlandia	521
Estonia	519
Korea	517
Japonia	516
Norwegia	513
Nowa Zelandia	509
Niemcy	509
Makao	509
Polska	506
Słowenia	505
Holandia	503
Australia	503
Szwecja	500
Dania	500
Francja	499
Belgia	499
Portugalia	498
Wielka Brytania	498
Tajpej (Republika Chińska)	497
Stany Zjednoczone	497
Hiszpania	496
Rosja	495
B-S-J-G (Chiny)	494
Szwajcaria	492
Łotwa	488
Czechy	487
Chorwacja	487
Wietnam	487
Austria	485
Włochy	485
Islandia	482
Luksemburg	481
Izrael	479
Autonomiczne Miasto Buenos Aires	475
Litwa	472
Węgry	470
Grecja	467
Chile	459
Słowacja	453
Malta	447
Cypr	443
Urugwaj	437
Rumunia	434
Zjednoczone Emiraty Arabskie	434
Bułgaria	432
Turcja	428
Kostaryka	427
Trinidad i Tobago	427
Czarnogóra	427
Kolumbia	425
Meksyk	423
Mołdawia	416
Tajlandia	409
Jordania	408
Brazylia	407
Albania	405
Katar	402
Gruzja	401
Peru	398
Indonezja	397
Tunezja	361
Dominikana	358
Macedonia	352
Algieria	350
Kosowo	347
Liban	347

z pięciu krajów. Wyniki polskich uczniów porównywalne są wynikami uczniów z następujących krajów: Nowa Zelandia, Niemcy, Makao, Słowenia, Holandia, Australia, Szwecja, Dania, Francja, Belgia. W 2015 roku Polska była więc nadal w gronie krajów o najwyższym wyniku, mimo że jej średni wynik, w porównaniu do roku 2012, był niższy (Tabela 1).

Warto spojrzeć na wyniki osiągnięte przez polskich uczniów w dłuższej perspektywie. Od roku 2000 do 2006 polscy uczniowie poprawiali swoją pozycję. W 2000 ich wynik wyniósł 479 punktów, w 2003 – 497, w 2006 – 508. Od tamtej pory wyniki ulegają wahaniom, ale nie wskazują na jednoznacznie tendencję (2009 – 500, 2012 – 518, 2015 – 506 pkt). Między 2000 a 2015 rokiem w Polsce średni wynik z czytania i interpretacji wzrósł o 27 pkt.

Dla porównania w Tabeli 2. zamieszczono wyniki z badań z poprzednich edycji w krajach Unii Europejskiej. Warto zauważyć, że średni wynik krajów OECD między rokiem 2000 a 2015 zmniejszył się, o ile w pierwszym z tych lat wyniósł on 500 punktów, o tyle w ostatnim badaniu spadł do 493 punktów.

Tabela 1. Średnie wyniki w zakresie czytania i interpretacji w krajach objętych badaniem. Polska na tle świata w 2015 r.; pogrubioną czcionką zaznaczono kraje europejskie. Białym tłem oznaczone są kraje, których wynik nie jest istotnie różny od średniego wyniku w krajach OECD, jasnym kolorem wyróżniono kraje o wynikach lepszych, a ciemnym – gorszych od średniego wyniku OECD.

Tabela 3. Opis poziomów umiejętności na skali osiągnięć czytania i interpretacji

Poziom	Umiejętności typowe dla danego poziomu
Poniżej poziomu 1b (poniżej 262,04 punktów)	Uczeń nie radzi sobie z tekstem, nawet na podstawowym poziomie.
Poziom 1b (262,04-334,75 punkty)	Uczeń wydobywa z tekstu informacje wyrażone wprost, radzi sobie z tekstem łatwym, o znanej mu formie, dostrzega proste związki między tekstem a osobistym doświadczeniem
Poziom 1a (334,75-407,47 punktów)	Uczeń wyszukuje jasno wskazane informacje w tekście o znanej mu formie, określa temat tekstu, dostrzega związki między informacjami zawartymi w tekście a wiedzą potoczną.
Poziom 2 (407,47-480,18 punktów)	Uczeń wydobywa z tekstu jedną lub kilka informacji określonych w poleceniu, dostrzega główną ideę tekstu, dostrzega związki tekstu z wiedzą posiadaną z innych źródeł.
Poziom 3 (480,18-552,89 punkty)	Uczeń wyszukuje informacje, wydobywa z tekstu główną ideę, potrafi dokonywać połączeń napotkanych w tekście elementów.
Poziom 4 (552,89-625,61 punktów)	Uczeń wydobywa z tekstu kilka zawartych w nim informacji, potrafi ogarnąć sens całości tekstu, wykorzystuje posiadaną wiedzę do dokonywania ocen, które potrafi właściwie uzasadnić.
Poziom 5 (625,61-698,32 punktów)	Uczeń potrafi wydobyć szereg informacji, które są zawarte głęboko w tekście, interpretuje szczegóły tekstu i tekst w całości, dokonuje krytycznej oceny tekstu.
Poziom 6 (powyżej 698,32 punktów)	Uczeń bezbłędnie dociera do wszystkich informacji, dokonuje głębokiej interpretacji tekstu, stawia hipotezę dotyczącą całościowej oceny tekstu.

Poziom 2 można przyjąć za minimalny poziom kompetencji, poniżej którego rośnie ryzyko nieradzenia sobie w dalszej edukacji i w dorosłym życiu. Takie kryterium uzgodniły kraje Unii Europejskiej jako ważny wskaźnik osiągania celów polityki edukacyjnej – wykorzystywane jest jako wskaźnik zagrożenia wykluczeniem społecznym.

Ostatnia edycja badania potwierdza, że wyniki polskich uczniów należą do mało zróżnicowanych. Stosunkowo nieliczne są grupy skrajne: uczniów poniżej poziomu 2, a także na poziomach 5 i 6 (Wykresy 1, 2 i 3).

Trzeba podkreślić, że w Polsce między rokiem 2000 a 2015 zdecydowanie spadła liczba uczniów uzyskujących wyniki poniżej 2 poziomu. Obecnie mniejszy niż w Polsce odsetek tych uczniów odnotowuje się niewielu krajach i regionach: Hongkongu (9%), Irlandii (10%), Estonii, Kanadzie, Finlandii, Singapurze (po 11%), Makao (12%) oraz Japonii (13%). Warto zwrócić uwagę, że w Europie są tylko trzy kraje, w których odsetek uczniów uzyskujących najniższe wyniki jest mniejszy niż w Polsce.

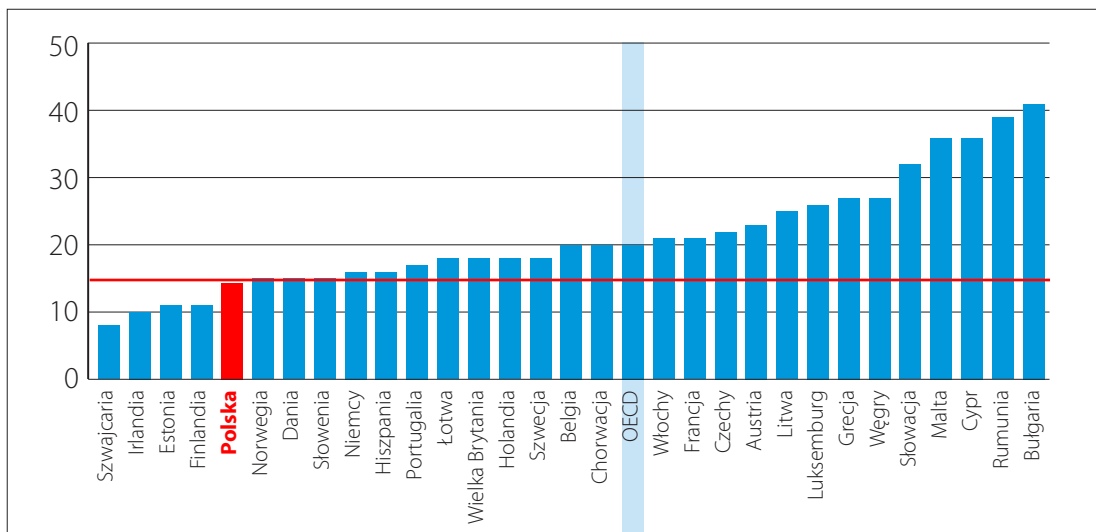
Mniej zadowolający jest efekt pracy z uczniami najzdolniejszymi. Obecnie odsetek polskich uczniów, którzy w dziedzinie czytania i interpretacji uzyskali wyniki na 5 i 6 poziomie (8,2%), jest zbliżony do średniej OECD (8,4%). W Europie w jedenastu krajach pod tym względem jest lepiej.

O ile od 2000 roku można było odnotować systematyczne zmniejszanie się odsetka polskich uczniów uzyskujących wyniki na poziomie niższym niż 2, o tyle w 2015 ich liczba wróciła do poziomu odnotowanego w 2009 r.

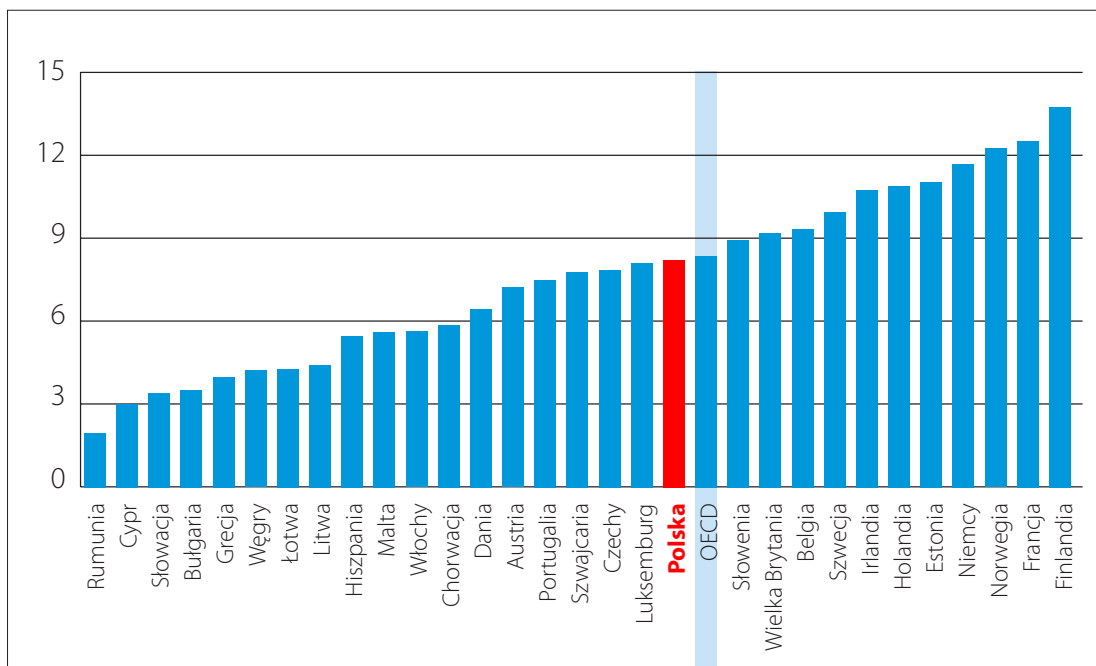
Warto również zauważyć, że na początku stulecia w Polsce nastąpił szybki wzrost odsetka uczniów uzyskujących najwyższe wyniki (podwojenie). Potem ów odsetek podlegał wahaniom między 7 a 12%. Można zatem mówić o względnej stabilizacji. (Wykres 4).



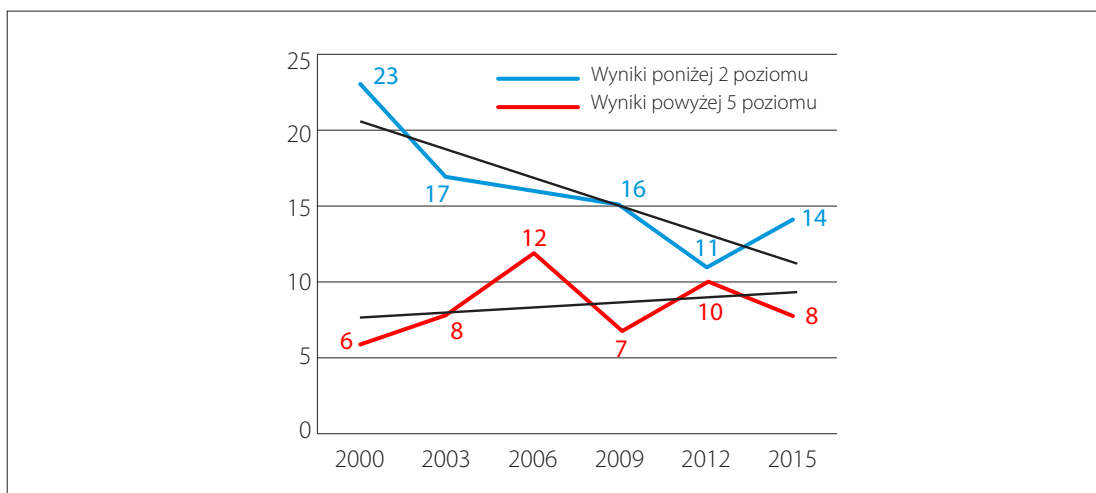
Wykres 2. Odsetek uczniów, którzy uzyskali wyniki poniżej 2 poziomu w krajach Unii Europejskiej i związanych z Unią Europejską w 2015 r. (%). Kraje uporządkowano według wzrastającego odsetka uczniów.



Wykres 3. Odsetek uczniów, którzy uzyskali wyniki powyżej 5 poziomu w krajach Unii Europejskiej i związanych z Unią Europejską w 2015 r. (%). Kraje uporządkowano według wzrastającego odsetka.



Wykres 4. Odsetek uczniów, którzy uzyskiwali wyniki poniżej 2 poziomu oraz powyżej 5 poziomu w Polsce w kolejnych edycjach badania (%)



Wyniki chłopców i dziewcząt

W badaniu umiejętności czytania i interpretacji zauważalna jest różnica między wynikami chłopców i dziewcząt, na korzyść dziewcząt. W 2015 w OECD średni wynik chłopców wyniósł 479, a dziewcząt 506. W Polsce wyniki wyglądały następująco: chłopcy 491 pkt., dziewczęta 521 pkt. Zatem różnica wyniosła 30 pkt. na korzyść dziewcząt, jest ona zbliżona do średniej OECD (27 pkt., Wykres 5).

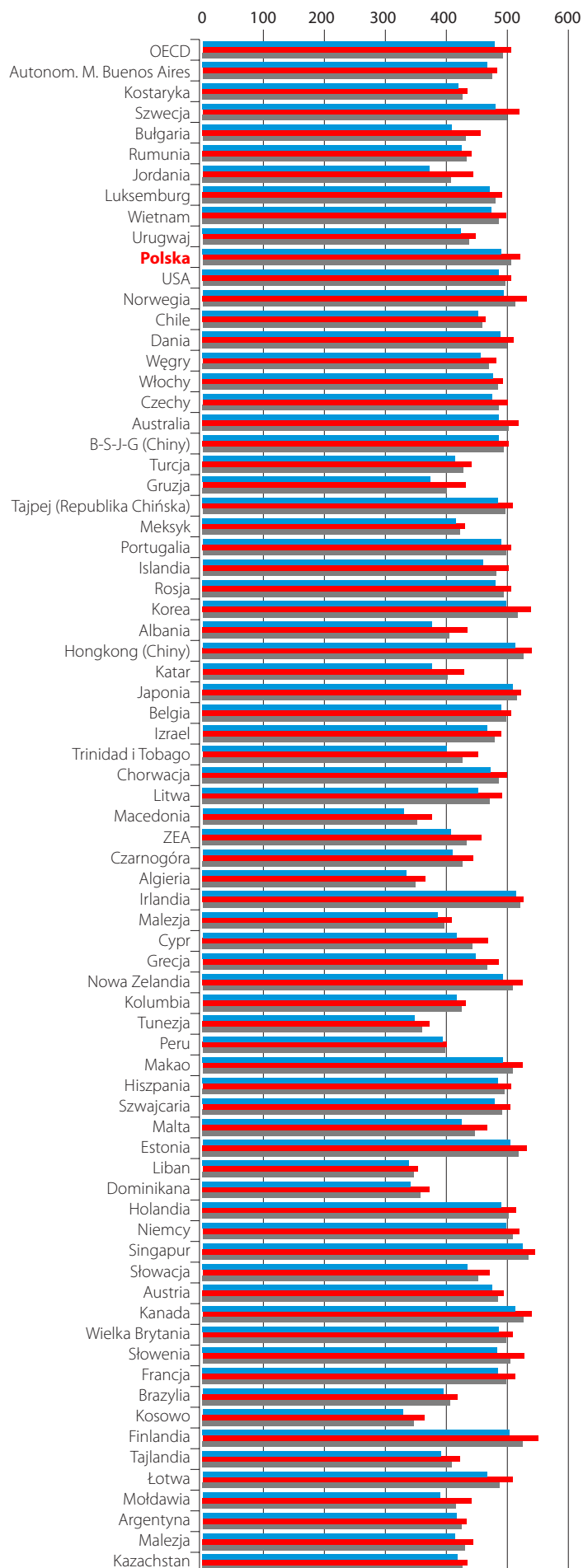
Między rokiem 2000 a 2015 nieznacznie zmniejszyła się ta różnica: w 2000 roku chłopcy w Polsce osiągnęli średni wynik 461 pkt., a dziewczęta 497 pkt., toteż różnica wtedy wynosiła 36 pkt. W 2012 roku różnica między chłopcami a dziewczętami wynosiła 42 pkt.

Interesujące jest porównanie różnic w rozbiciu na poziomy umiejętności. Od 2009 roku odsetek chłopców uzyskujących najniższe wyniki utrzymuje się na względnie stabilnym poziomie. Warto w tym miejscu zauważyć, że tylko w pięciu krajach UE odsetek chłopców uzyskujących najniższe wyniki był w 2015 roku mniejszy niż w Polsce. Podobnie stabilny jest wynik dziewcząt. Badanie w 2012 r. pokazało zmniejszenie się odsetka uczennic uzyskujących wyniki poniżej drugiego poziomu. W 2015 r. odsetek dziewcząt z takimi wynikami nieco wzrósł, mimo to wyniki te należą nadal do najkorzystniejszych w UE – podobnie jak w przypadku chłopców tylko w pięciu krajach (w tym w pozostającej formalnie poza UE Norwegii) jest mniejszy odsetek dziewcząt uzyskujących najłabsze wyniki.

Jeśli chodzi o chłopców uzyskujących najwyższe wyniki, to ich odsetek w Polsce jest zbliżony do przeciętnego w UE. Jest on dosyć stabilny w ujęciu czasowym, jakkolwiek daje się zauważyć wyraźny wzrost liczby uczniów uzyskujących najwyższe wyniki w stosunku do badania z 2009 roku (niemal podwojenie). Odsetek dziewcząt uzyskujących najwyższe wyniki również jest zbliżony do średniej dla UE. W ujęciu czasowym można zauważyć względną stabilizację od 2009 roku.

Wykres 5. Średnie wyniki dziewcząt i chłopców w krajach i regionach uczestniczących w badaniu PISA 2015

■ Chłopcy
■ Dziewczęta
■ Wszyscy uczniowie





Podsumowanie

Wyniki badania PISA z 2015 roku pokazują, że polscy uczniowie dobrze opanowali umiejętności z dziedziny czytania i interpretacji. Spadek średniej liczby punktów (dający się również zaobserwować dla średniej liczby punktów w OECD) mieści się w zakresie niedającym powodów do niepokoju, zwłaszcza że porównanie wyników z kolejnych edycji badania wskazuje na względną stabilizację, która została osiągnięta po reformie przeprowadzonej na przełomie stuleci (między wynikami z 2000 i 2015 roku odnotowuje się zauważalny wzrost średniej liczby punktów). W okresie

2000–2015 zauważalny jest także spadek odsetka uczniów, którzy uzyskują wyniki poniżej 2 poziomu oraz nieznaczny wzrost odsetka tych najlepszych. Trudno jednoznacznie stwierdzić, co jest przyczyną słabszego wyniku w roku 2015 w porównaniu z rokiem 2012.

Badanie PISA z 2015 roku potwierdza obserwacje z poprzednich edycji, że polski system kształcenia umiejętności czytania i interpretacji jest skuteczny, jakkolwiek konieczne jest zwrócenie większej uwagi na rozszerzanie grupy uczniów biegłych w wykonywaniu operacji najtrudniejszych, wymagających znacznej samodzielności myślenia.



MATEMATYKA

Pomiar umiejętności matematycznych w badaniu PISA koncentruje się na tym, w jakim stopniu piętnastoletni uczniowie są w stanie wykorzystać swoją wiedzę i umiejętności matematyczne, gdy stają przed koniecznością rozwiązania problemów praktycznych, jakich dostarcza im otaczający świat. Nawet bardzo dobra, ale tylko teoretyczna znajomość narzędzi matematyki nie wystarczy. Dobry wynik uzyskują uczniowie, którzy umieją rozumować matematycznie, a także potrafią skutecznie wykorzystywać pojęcia i narzędzia matematyczne do opisu, analizy i prognozowania różnych zjawisk. Są to wymagania znacznie wykraczające poza umiejętność mechanicznego odtworzenia wyuczonych pojęć i procedur szkolnych. Aby im sprostać, potrzebne są solidne umiejętności i dobre rozumienie matematyki na poziomie gimnazjalnym.

W 2015 roku w badaniu PISA wzięli udział uczniowie z aż 72 krajów lub regionów świata. Daje nam to, po raz kolejny, możliwość przyjrzenia się, jak z określonym powyżej niełatwym wyzwaniem radzą sobie polscy uczniowie w porównaniu z rówieśnikami z innych krajów.

Część matematyczna badania zawierała 70 zadań o różnym stopniu trudności. W niektórych można było wybrać odpowiedź z zaproponowanej listy; jednak w większości należało opracować i przedstawić odpowiedź samodzielnie.

Umiejętności polskich uczniów

Matematyka była główną dziedziną badania PISA w latach 2003 oraz 2012. Po badaniu w roku 2003 została zdefiniowana skala matematyczna, do której odnoszone są wyniki uczniów uzyskane w kolejnych cyklach. Skala ta została ska-

librowana tak, by średni wynik krajów OECD (w roku 2003) był równy 500, a odchylenie standardowe 100.

Zastosowanie uogólnionego modelu Rascha daje możliwość umieszczenia na tej skali zarówno wszystkich badanych uczniów, jak i wszystkich zadań. Zadania, które na tej skali znajdują się wysoko, mierzą umiejętności złożone. Uczniowie uzyskujący wysoki wynik radzą sobie z takimi zadaniami lepiej od innych.

Dla lepszego objaśnienia tych różnic oraz uzyskanych wyników na skali wyróżniono sześć poziomów oraz podano typowe umiejętności uczniów, plasujących się na każdym z nich (Tabela 3).

W badaniu PISA 2015 średni wynik krajów OECD z matematyki wyniósł 490 punktów.

O wyniku kraju decyduje średnia wyników wszystkich uczniów z tego kraju uczestniczących w badaniu. Wyniki te obarczone są jednak błędem statystycznym, który można oszacować wykorzystując odpowiednie techniki statystyczne. Dzięki temu wiemy, że dwa kraje, które mają różne wyniki, mogą mimo to być statystycznie nieodróżnialne. Dlatego należy zachować dużą ostrożność przy interpretowaniu listy krajów uszeregowanych według średniego wyniku punktowego (Tabela 1).

Dla każdego kraju można wskazać grupę krajów, których wyniki nie różnią się od jego wyniku w sposób statystycznie istotny. Dla Polski tymi krajami są Belgia, Niemcy, Irlandia i Norwegia. Wynik istotnie wyższy od Polski uzyskało 14 krajów i regionów, w tym 6 krajów europejskich: Szwajcaria, Estonia, Holandia, Dania, Finlandia i Słowenia.

Tabela 1. Średnie wyniki uczniów z matematyki w badaniu PISA 2015.

Pogrubioną czcionką wyróżniono kraje europejskie. Kolor tła wyróżnia kraje, które uzyskały wyniki (odpowiednio, od góry) statystycznie istotnie wyższe, statystycznie nieodróżnialne oraz statystycznie istotnie niższe niż średnia krajów OECD.

Kraj/region	Średni wynik 2015	Kraj/region	Średni wynik 2015
Singapur	564	Węgry	477
Hongkong	548	Słowacja	475
Makao	544	Izrael	470
Tajwan	542	USA	470
Japonia	532	Chorwacja	464
B-S-J-G (Chiny)	531	Kazachstan	460
Korea	524	Grecja	454
Szwajcaria	521	Malezja	446
Estonia	520	Rumunia	444
Kanada	516	Bułgaria	441
Holandia	512	Cypr	437
Dania	511	Emiraty	427
Finlandia	511	Chile	423
Słowenia	510	Turcja	420
Belgia	507	Mołdawia	420
Niemcy	506	Urugwaj	418
Polska	504	Czarnogóra	418
Irlandia	504	Trynidad	417
Norwegia	502	Tajlandia	415
Austria	497	Albania	413
Nowa Zelandia	495	Argentyna	409
Wietnam	495	Meksyk	408
Rosja	494	Gruzja	404
Szwecja	494	Katar	402
Australia	494	Kostaryka	400
Francja	493	Liban	396
Portugalia	492	Kolumbia	390
Wielka Brytania	492	Peru	387
Czechy	492	Indonezja	386
Włochy	490	Jordania	380
Islandia	488	Brazylia	377
Hiszpania	486	Macedonia	371
Luksemburg	486	Tunezja	367
Łotwa	482	Kosowo	362
Malta	479	Algieria	360
Litwa	478	Dominikana	328

Źródło: Raport międzynarodowy: PISA 2015 Results Excellence and equity in education, Volume I, OECD 2016, tabela I.502

Tabela 2. Średnie wyniki uczniów z matematyki w krajach o najwyższych wynikach w badaniu PISA 2012 oraz zmiana wyniku między rokiem 2012 a 2015 (w punktach).

Kolorem czerwonym zaznaczono obniżenie wyniku punktowego, a czarnym poprawę wyniku. Zmiany istotne statystycznie zaznaczono pogrubioną czcionką. Gwiazdka oznacza, że kraj lub region nie brał udziału w badaniu PISA 2015.

Kraj/region	Średni wynik w 2012	Zmiana od 2012
Szanghaj (Chiny)	613	*
Singapur	573	-9
Hongkong	561	-13
Tajwan	560	-18
Korea	554	-30
Macao (Chiny)	538	+6
Japonia	536	-4
Liechtenstein	535	*
Szwajcaria	531	-10
Holandia	523	-11
Estonia	521	-1
Finlandia	519	-8
Kanada	518	-2
Polska	518	-14
Belgia	515	-8
Niemcy	514	-8
Wietnam	511	-16
Austria	506	-9
Australia	504	-10
Irlandia	501	+3
Słowenia	501	+9
Dania	500	+11
Nowa Zelandia	500	-5
Czechy	499	-7
Francja	495	-2
Wielka Brytania	494	-2
Islandia	493	-5
Łotwa	491	-9
Luksemburg	490	-4
Norwegia	489	+13
Portugalia	487	+5
Hiszpania	487	-1
Włochy	485	-5
Rosja	482	+12
Słowacja	482	-7
USA	481	-11

Źródło: Raport międzynarodowy: PISA 2015 Results Excellence and equity in education, Volume I, OECD 2016, tabela I.504

Na wyniki badania PISA 2015 oczekiwano w wielu krajach z dużym zainteresowaniem także ze względu na to, że towarzyszyła mu zmiana, która mogła istotnie te wyniki zaburzyć: w 59 krajach (w tym w Polsce) przeprowadzono badanie nie jak dotychczas w wersji papierowej, lecz za pomocą komputerów. Uczniowie rozwiązywali dokładnie te same zadania co w 2012 roku, ale nie otrzymali ich w formie książeczki, lecz czytali je odpowiednio nawigując na ekranach, a swoje rozwiązania wpisywali z klawiatury komputera lub zaznaczali myszką. Nie było jasne, jak ta zmiana wpłynie na wyniki badania.

Okazało się, że w zdecydowanej większości krajów, które w 2012 roku miały wyniki zbliżone do średniej OECD lub wyższe, w roku 2015 nastąpił spadek wyniku o kilka, kilkanaście lub nawet kilkadziesiąt punktów. Tylko trzy kraje (Słowenia, Dania i Norwegia) na 29 krajów z tej grupy zanotowały istotny statystycznie wzrost wyniku (Tabela 2).

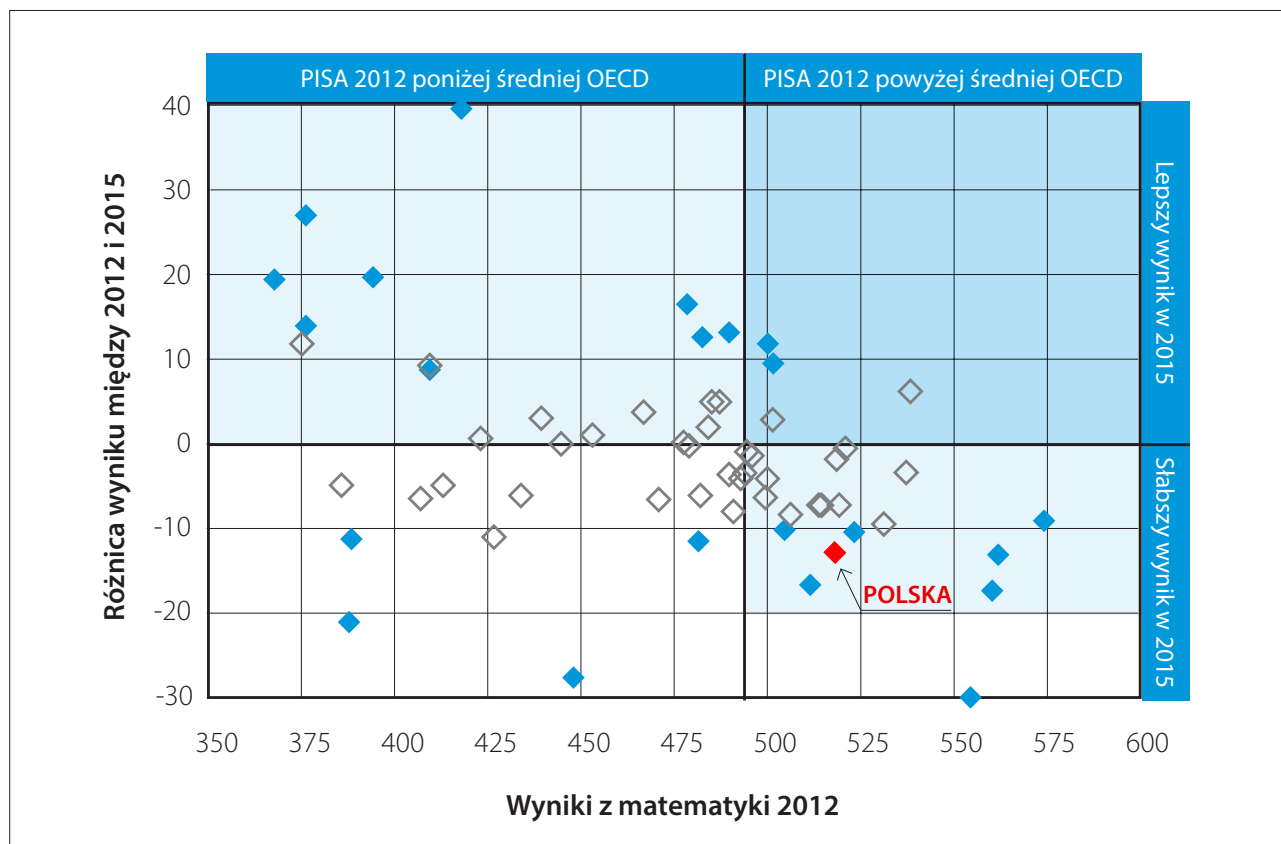
Jednak mimo tych zmian, lista krajów, które w badaniu PISA 2015 osiągnęły najlepsze wyniki z matematyki uległa tylko niewielkim zmianom w stosunku do badania PISA 2012. Można zatem stwierdzić, że w zakresie umiejętności matematycznych polscy gimnazjaliści utrzymali swoją wysoką pozycję.

To samo zjawisko ilustruje Wykres 1. Pokazuje on także, że znaczna większość krajów, których wyniki z matematyki są istotnie statystycznie lepsze w 2015 roku niż w 2012, to kraje o bardzo niskich wynikach – poniżej 425 punktów.



Wykres 1. Różnica wyniku z matematyki między badaniami PISA 2012 i PISA 2015 w poszczególnych krajach, w zależności od wyniku z matematyki w badaniu PISA 2012 w tych krajach.

Niebieskim rombem zaznaczono zmiany istotne statystycznie, przezroczystym rombem zmiany nieistotne statystycznie.



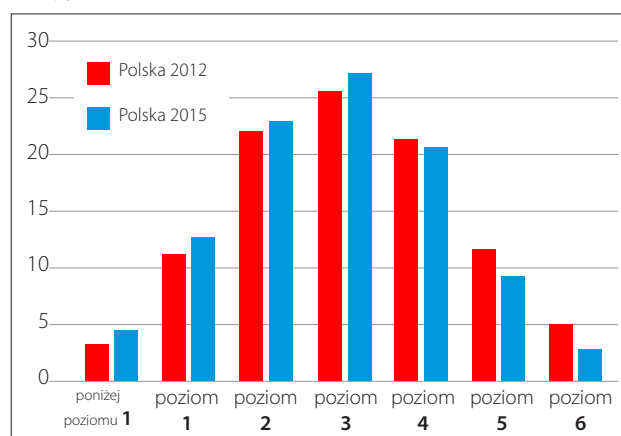
Źródło: Raport międzynarodowy: PISA 2015 Results Excellence and equity in education, Volume I, OECD 2016, diagram I.505

Zmiany w poziomach osiągnięć

Słabszemu wynikowi kraju towarzyszy z reguły zwiększenie odsetka uczniów na najniższych poziomach umiejętności, a także obniżenie odsetka uczniów na najwyższych poziomach. Podobnie jak w wielu najlepszych krajach świata, także w Polsce nastąpiło takie przesunięcie (Wykres 2.).

W badaniu PISA 2015 odsetek uczniów o umiejętnościach poniżej poziomu 2 wzrósł w Polsce z 14,4% w 2012 roku do 17,2% w roku 2015. Także odsetek uczniów na dwóch najwyższych poziomach umiejętności zmniejszył się z 16,7% do 12,2%.

Wykres 2. Odsetki uczniów w Polsce na poszczególnych poziomach umiejętności w badaniu PISA 2015.



Dla lepszego objaśnienia tych różnic oraz uzyskanych wyników na skali wyróżniono sześć poziomów oraz podano typowe umiejętności uczniów, plasujących się na każdym z nich (Tabela 3).

Tabela 3. Opis poziomów umiejętności na skali osiągnięć matematycznych

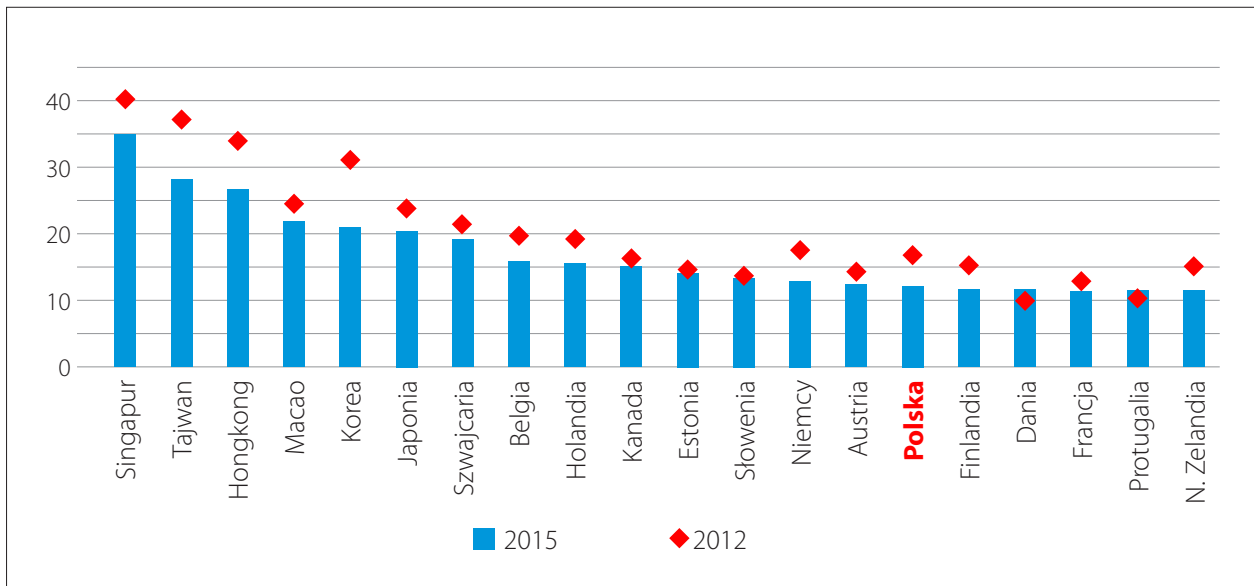
Poziom (dolna granica)	Umiejętności typowe dla danego poziomu umiejętności matematycznych
Poziom 6 (669 pkt.)	Uczeń potrafi analizować i uogólniać informacje zgromadzone w wyniku zbadania samodzielnie zbudowanego modelu złożonej sytuacji problemowej. Umie połączyć informacje pochodzące z różnych źródeł i swobodnie przemieszczać się między nimi. Potrafi wykonywać zaawansowane rozumowania i umie wnioskować matematycznie. Umie połączyć rozumowanie z biegłością w wykonywaniu operacji symbolicznych i formalnych podczas twórczej pracy nad nowym dla siebie kontekstem. Potrafi precyzyjnie formułować komunikat o swoim rozumowaniu, uzasadniając podjęte działania.
Poziom 5 (607 pkt.)	Uczeń umie modelować złożone sytuacje, identyfikując ograniczenia i precyzując zastrzeżenia. Potrafi porównywać, oceniać i wybierać odpowiednie strategie rozwiązywania problemów związanych ze zbudowanym modelem. Wykorzystuje dobrze rozwinięte umiejętności matematyczne, z użyciem odpowiednich reprezentacji, w tym symbolicznych i formalnych. Potrafi krytycznie ocenić swoje działania, zakomunikować swoją interpretację oraz sposób rozumowania.
Poziom 4 (545 pkt.)	Uczeń umie efektywnie pracować z podanymi wprost modelami złożonych sytuacji realnych, identyfikując ograniczenia i czyniąc niezbędne założenia. Potrafi wybierać oraz łączyć informacje pochodzące z różnych źródeł, wiążąc je bezpośrednio z kontekstem realnym. Umie w takich kontekstach stosować ze zrozumieniem dobrze wyuczone techniki. Potrafi konstruować komunikaty opisujące swoje interpretacje, argumenty i działania.
Poziom 3 (482 pkt.)	Uczeń umie wykonać jasno opisany algorytm, także wymagający sekwencyjnego podejmowania decyzji. Potrafi wybierać i stosować proste strategie rozwiązywania problemów. Potrafi interpretować i wyciągać bezpośrednie wnioski z danych pochodzących z kilku źródeł. Umie przedstawić wyniki nieskomplikowanych interpretacji i rozważań.
Poziom 2 (420 pkt.)	Uczeń umie rozpoznać i zinterpretować sytuację wymagającą tylko prostego kojarzenia. Potrafi wydobyć istotną informację z pojedynczego źródła i użyć na raz jednej formy reprezentacji danych. Umie zastosować prosty wzór lub przepis postępowania. Potrafi wyciągnąć bezpośrednie wnioski i dosłownie zinterpretować wyniki.
Poziom 1 (358 pkt.)	Uczeń umie rozwiązywać typowe zadania, w których wszystkie dane są bezpośrednio podane, a zadane pytania są proste. Potrafi wykonywać czynności rutynowe, postępując zgodnie z podanym prostym przepisem. Podejmuje działania oczywiste, wynikające wprost z treści zadania.
Poniżej poziomu 1	Uczeń wykazuje brak umiejętności nawet na poziomie 1.

Źródło: Raport międzynarodowy: PISA 2015 Results Excellence and equity in education, Volume I, OECD 2016, tabela I.502

Należy podkreślić, że zarówno pod względem odsetka najlepszych, jak i najgłupszych uczniów wynik Polski jest wciąż bardzo dobry na tle innych krajów świata. Ilustrują to Wykresy 3. i 4.

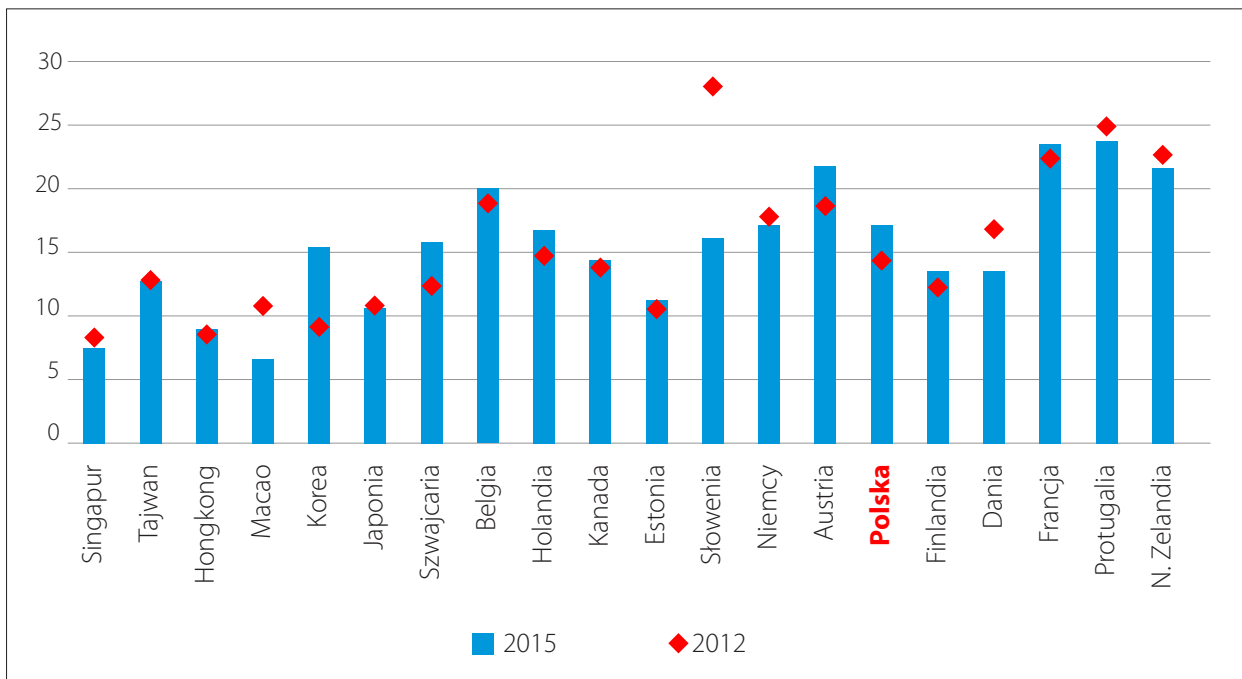
Wykres 3. Odsetki uczniów na poziomach 5 lub 6 w poszczególnych krajach w latach 2012 i 2015.

Na wykresie uwzględniono kraje, które w badaniu PISA 2015 mają najwyższe odsetki uczniów na tych poziomach umiejętności. Kraje uporządkowano według malejącego odsetka uczniów na tych poziomach w 2015 roku.



Wykres 4. Odsetki uczniów na 1. poziomie lub niżej w poszczególnych krajach w latach 2012 i 2015.

Na wykresie uwzględniono te same kraje, co na wykresie 3 i uporządkowano je w takiej samej kolejności jak na wykresie 3.

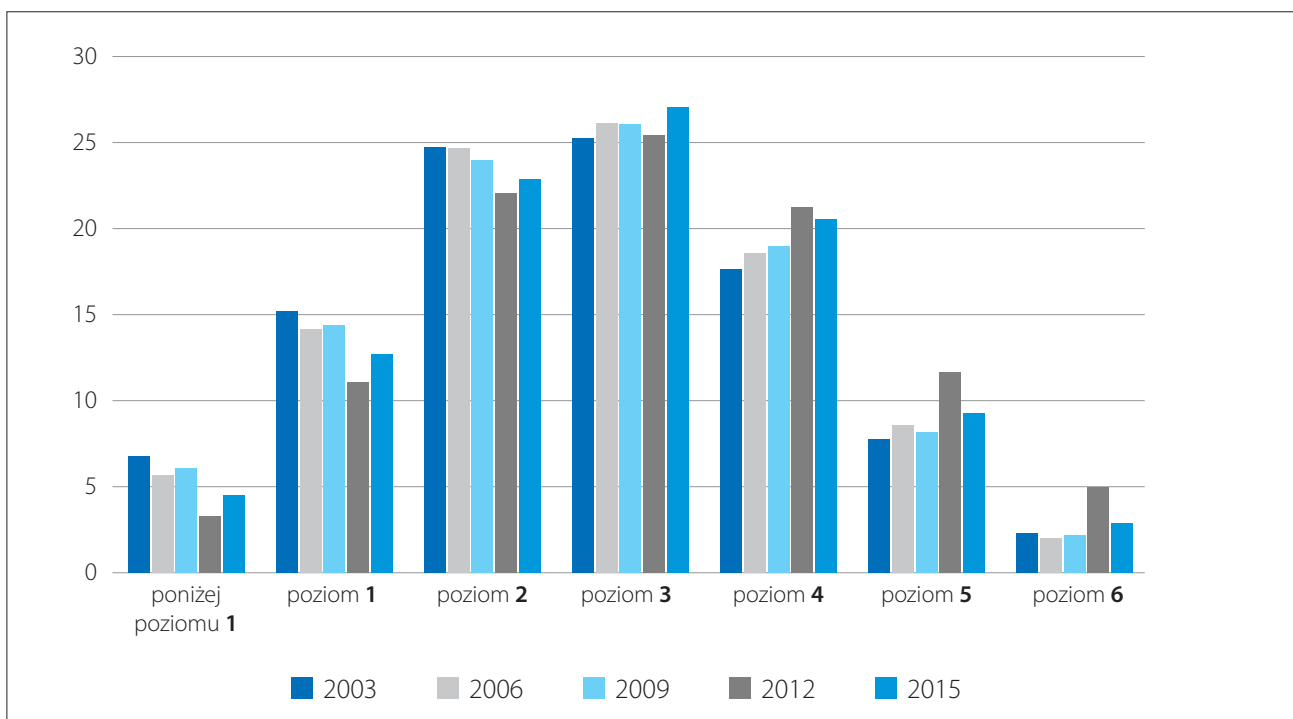


Opracowano na podstawie raportu międzynarodowego: PISA 2015 Results Excellence and equity in education, Volume I, OECD 2016, diagram I.509

Warto spojrzeć na tę kwestię z perspektywy historycznej. Na Wykresie 5. przedstawiona jest zmiana rozkładu uczniów pomiędzy poszczególne poziomy umiejętności matematycznych na przestrzeni pięciu ostatnich edycji badania PISA. Pokazuje on, że w 2015 roku uczniów o umiejętnościach poniżej

poziomu 2 było procentowo mniej niż w latach 2003-2009. Podobnie odsetki uczniów na poziomach 5 i 6 są wyższe w roku 2015 niż w latach 2003-2009, w których były one zbliżone do średniego odsetka w krajach OECD.

Wykres 5. Odsetki uczniów w Polsce na poszczególnych poziomach umiejętności w latach 2003-2015.



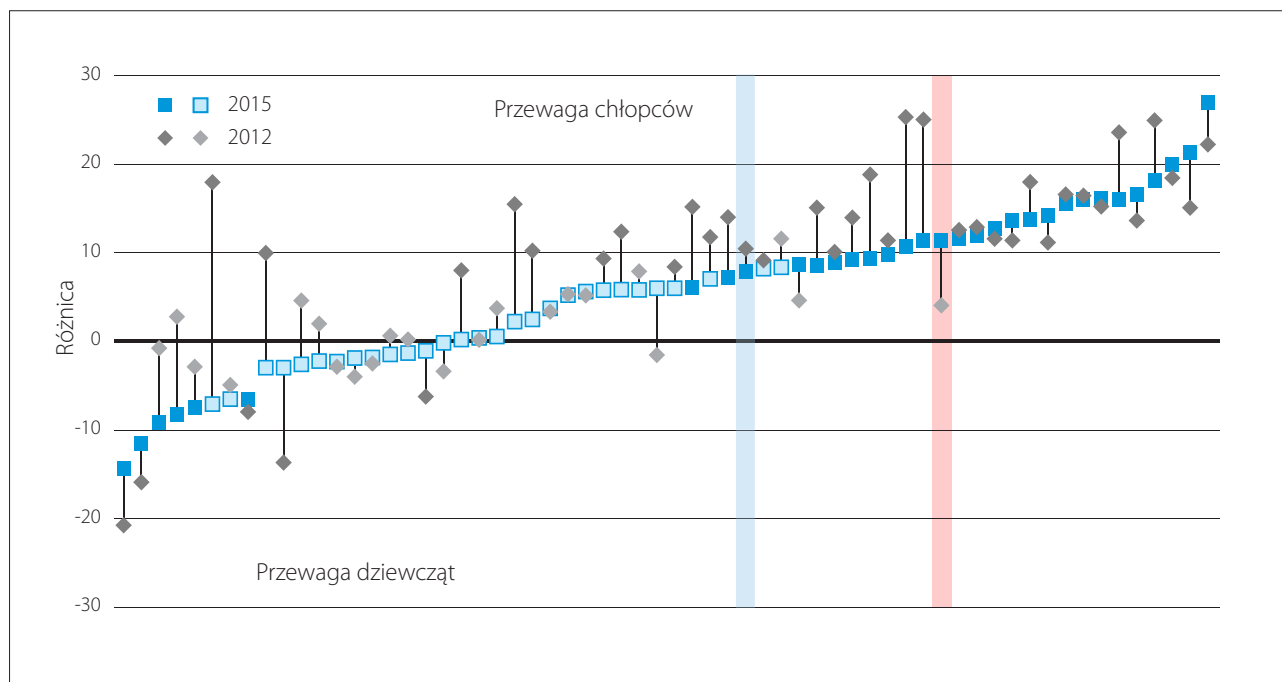
Porównanie wyników chłopców i dziewcząt

W badaniu PISA 2015 w wielu krajach nastąpiła znaczna zmiana zróżnicowania wyników dziewcząt i chłopców w stosunku do notowanych wcześniej. Efekt ten przedstawiony jest na Wykresie 6.

Dane uporządkowano począwszy od kraju o największej różnicy na korzyść dziewcząt w 2015 roku, do największej różnicy na korzyść chłopców w 2015 roku. Niebieskim kwadratem i ciemnym rombem zaznaczono różnice wyniku istotne statystycznie, natomiast błękitnym kwadratem i jasnym rombem różnice nieistotne statystycznie. Czerwona ramka wskazuje różnice dla Polski, a błękitna - średnie różnice dla krajów OECD.

Po raz pierwszy od wielu lat chłopcy w Polsce uzyskali istotną statystycznie przewagę nad dziewczętami w zakresie umiejętności matematycznych – nieistotna statystycznie czteropunktowa różnica z roku 2012 zwiększyła się w 2015 roku do 12 punktów. Średni wynik chłopców wyniósł 511 punktów, zaś dziewcząt 499 punktów. Przewaga chłopców dotyczy także dolnego i górnego krańca skali umiejętności. W dolnym (najniższym) decyly wyniki chłopców i dziewcząt to odpowiednio 396 i 387 punktów (9 punktów różnicy), zaś w ostatnim (najwyższym) decyly te wyniki to odpowiednio 627 oraz 608 punktów (aż 19 punktów różnicy na korzyść chłopców).

Wykres 6. Różnica między wynikami z matematyki osiąganymi przez dziewczęta i chłopców w latach 2012 i 2015 (wynik chłopców minus wynik dziewcząt w punktach).



Źródło: Raport międzynarodowy: PISA 2015 Results Excellence and equity in education, Volume I, OECD 2016, diagram I.511

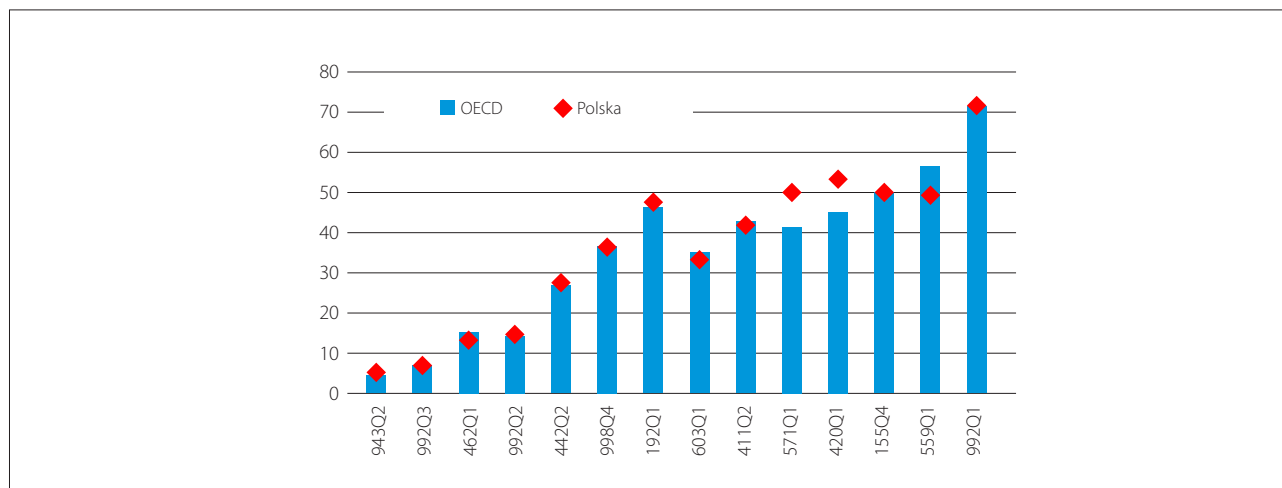
Rozumowanie matematyczne

Jednym z wyznaczników sukcesu polskich gimnazjalistów w 2012 roku była istotna poprawa wyników w zadaniach wymagających rozumowania matematycznego (mierzących V. wymaganie ogólne podstawy programowej). Wtedy, po raz pierwszy od początku badania PISA, polskim uczniom uda-

ło się uzyskać dla prawie każdego z tych trudnych zadań wynik taki sam lub lepszy od średniej OECD.

Badanie PISA 2015 potwierdza trwałość tego zjawiska. Wykres 7. pokazuje porównanie wyników polskich uczniów ze średnim wynikiem OECD dla wszystkich zadań, które wymagały rozumowania matematycznego.

Wykres 7. Wyniki uczniów w Polsce i w krajach OECD w badaniu PISA 2015 w zadaniach wymagających rozumowania matematycznego (w procentach).



Wieloletni trend zmiany wyników

Raport międzynarodowy komentuje także trend poziomu umiejętności matematycznych, obserwowany w kolejnych cyklach badania od 2003 roku. Stwierdzono w nim, że średni wynik pomiaru umiejętności matematycznych w krajach OECD systematycznie obniża się, średnio o 1,7 punktu na każdy cykl badania PISA. W tym kontekście odnotowano, że polscy gimnazjaliści zachowują trend wznoszący, poprawiając swój wynik średnio o 5 punktów z cyklu na cykl. Wraz z Polską w tym kontekście wymieniona jest jeszcze tylko Brazylia oraz Tunezja, poprawiające swoje wyniki średnio odpowiednio o 6,2 oraz 3,8 punktu na cykl. Oba te kraje, w przeciwieństwie do Polski, pozostają jednak ciągle w strefie wyników słabszych od średniej OECD (odpowiednio 377 oraz 367 punktów w 2015 roku).

Podsumowanie

W zakresie matematyki badanie PISA mierzy, w jakim stopniu piętnastoletni uczniowie są w stanie wykorzystać swoją wiedzę i umiejętności matematyczne do rozwiązania problemów. Mierzy zatem, czy uczniowie umieją rozumować matematycznie, a także potrafią skutecznie wykorzystywać

pojęcia i narzędzia matematyczne do opisu, analizy i prognozowania różnych zjawisk.

W 2015 roku uczniowie rozwiązywali z matematyki te same zadania co w roku 2012, ale po raz pierwszy pomiar umiejętności odbywał się przy pomocy komputerów. Być może w związku z tą zmianą w zdecydowanej większości krajów, które w 2012 roku miały wyniki zbliżone do średniej OECD lub wyższe, w roku 2015 nastąpił spadek wyniku o kilka, kilkanaście lub nawet kilkadziesiąt punktów. Jednak mimo tych zmian, lista krajów, które w badaniu PISA 2015 osiągnęły najlepsze wyniki z matematyki uległa tylko niewielkim zmianom w stosunku do badania PISA 2012. W zakresie umiejętności matematycznych polscy gimnazjaliści utrzymali swoją wysoką pozycję. Także pod względem odsetka najlepszych, jak i najsłabszych uczniów wynik Polski jest wciąż bardzo dobry na tle innych krajów.

W roku 2012 polskim uczniom udało się po raz pierwszy uzyskać dla prawie każdego z najtrudniejszych zadań, wymagających rozumowania matematycznego, wynik taki sam lub lepszy od średniej OECD. Badanie PISA 2015 potwierdziło trwałość tego zjawiska.

Raport międzynarodowy zauważa, że polscy gimnazjaliści w zakresie matematyki zachowują trend wznoszący, poprawiając swój wynik średnio o 5 punktów z cyklu na cykl.