

4. Metoda edukacyjnej wartości dodanej w Polsce

*Roman Dolata, Anna Hawrot, Grzegorz Humenny, Aleksandra Jasińska-Maciążek, Anna Rappe,
Ewa Stożek, Tomasz Żółtak*

4.1. Wprowadzenie

Decentralizacja i zwiększanie autonomii szkół zwiększa zapotrzebowanie na informacje o jakości pracy poszczególnych placówek. Tego rodzaju informacje mogą być wykorzystywane zarówno do zewnętrznej oceny pracy szkoły, jak i do doskonalenia jakości nauczania. Egzamin i informacje z testów osiągnięć szkolnych przeprowadzonych w sposób umożliwiający dokonywanie porównań między szkołami są podstawowym źródłem informacji, które można do tego celu wykorzystać. Jednak pełne wykorzystanie informacji płynących z systemu egzaminacyjnego wymaga odpowiednich narzędzi analitycznych. W poprzednim rozdziale opisane zostały porównywalne między latami wyniki egzaminacyjne, innym ważnym instrumentem analitycznym jest metoda edukacyjnej wartości dodanej (EWD).

Wyniki egzaminacyjne uczniów zależą od wielu czynników niezależnych od szkoły, w której uczyli się na danym etapie edukacyjnym, przede wszystkim od uprzednich osiągnięć. Metoda EWD to modele statystyczne pozwalające na wnioskowanie o efektywności szkoły, czyli o wkładzie szkoły w wyniki nauczania. By można stosować metodę EWD, potrzebujemy wyników przynajmniej dwóch pomiarów osiągnięć szkolnych: na początku nauki w danej szkole i na jej zakończenie. Istotą metody EWD jest odejście od patrzenia na szkoły jedynie przez pryzmat wyniku osiąganego przez uczniów na koniec nauki. Ponieważ wyniki nauczania zależą w znaczącej mierze od czynników, które są poza kontrolą szkoły, używanie rezultatów egzaminów końcowych jako miary efektywności musi prowadzić do niesprawiedliwych ocen. EWD jest zdecydowanie bardziej wartościową metodą oceny efektywności nauczania. Uwzględnienie wyników uczniów na początku rozpoczęcia nauki w danej szkole pozwala wskazać szkoły, które mimo relatywnie słabych wyników końcowych mogą pochwalic się wysoką efektywnością nauczania i odwrotnie, szkoły będące wysoko w rankingu wyników testów końcowych, ale niewiele przyczyniające się do postępów uczniów. Należy przy tym zauważyć, że wyniki „na wejściu” informują nie tylko o poziomie osiągnięć szkolnych ucznia na progu szkoły, ale są też „nośnikiem” wiedzy o środowisku rodzinnym dziecka, o jego poziomie zdolności i motywacji szkolnej. Dzieje się tak dlatego, że wyniki egzaminu „na wejściu” są uwarunkowane podobnymi czynnikami, co rezultaty egzaminu końcowego.

Pojęcie edukacyjnej wartości dodanej pojawiło się po raz pierwszy w połowie lat 70. Od początku lat 90. intensywne prace nad metodą edukacyjnej wartości dodanej prowadzone są głównie w Stanach Zjednoczonych i Wielkiej Brytanii, a od początku tego wieku w Australii. Niektórzy badacze uważają pojęcie edukacyjnej wartości dodanej za najważniejsze narzędzie analityczne, jakie w naukach pedagogicznych pojawiło się pod koniec XX wieku (Schagen i Hutchison, 2003). Nie oznacza to oczywiście, że metoda ta jest idealna i nie jest pozbawiona wad.

Na świecie istnieje wiele wariantów metody EWD, co częściowo odzwierciedla zróżnicowane możliwości korzystania z danych, a częściowo różne cele stosowania metody. Na przykład możliwości modelowania zmian umiejętności uczniów są większe w Stanach Zjednoczonych niż w Polsce, ze względu na fakt, że testy przeprowadzane są rokrocznie przez 8 lat nauki. Z drugiej strony, metody szacowania edukacyjnej wartości dodanej powinny być podporządkowane temu, czemu mają służyć i kto będzie odbiorcą informacji. Wskaźniki EWD mogą służyć ocenie efektywności nauczania na potrzeby ewaluacji wewnętrznej, mogą być wykorzystywane przez nadzór pedagogiczny i samorządy w ewaluacji zewnętrznej czy w końcu w ocenie pracy szkoły na potrzeby rodziców

i uczniów wybierających szkołę czy oceniających szkołę, z usług której korzystają (Raudenbush i Willms, 1995). Metoda może być stosowana w analizie efektywności nauczania szkół lub poszczególnych nauczycieli.

W wypadku szacowania EWD na potrzeby profesjonalnych procesów ewaluacji ważne jest uwzględnienie różnic nie tylko w poziomie uprzednich osiągnięć, ale też, w miarę możliwości i udokumentowanej badawczo potrzeby, innych czynników wpływających na postęp edukacyjny. Choć badania pokazują, że wpływ bazy materialnej jest znikomy (por. Hanushek, 2003), to już np. wpływ składu społecznego szkoły może być znaczący (Markman i in., 2003).

Innego sposobu szacowania EWD wymagałyby natomiast wskaźniki adresowane do rodziców i uczniów. W tym wypadku wartość dodana powinna być liczona dla całej placówki, bez wyłączenia wpływu zasobów szkolnych (zarówno materialnych, jak i np. składu społecznego szkoły), bowiem z punktu widzenia rodziców interesujące jest całościowe oddziaływanie danej placówki na postęp w osiągnięciach szkolnych dziecka. W wypadku wskaźników EWD dla rodziców jest też bardzo ważne, czy efektywność nauczania w danej placówce zależy od wstępnego poziomu osiągnięć szkolnych. Szkoła może bardzo efektywnie nauczać uczniów dobrych, a być nieefektywna w pracy z uczniami słabszymi, albo odwrotnie (por. Meyer, 1997).

Z faktu, że EWD jest metodą statystyczną, wynika jej siła analityczna, ale jest to też powód jej ważnego ograniczenia. Metody statystyczne mogą być stosowane tylko do zbiorowości. Można się spierać, czy minimalna liczebność takiej zbiorowości to 10 czy 30 elementów, ale z pewnością dla bardzo małych grup metoda EWD nie może być stosowana. Jeżeli na przykład w danej szkole do egzaminu przystąpiło czworo uczniów, żadna statystyka nie pozwoli na dokonywanie na tej podstawie jakichkolwiek statystycznych uogólnień. Jest to problem w Polsce dość istotny, bo spora grupa placówek to szkoły bardzo małe. Można próbować radzić sobie częściowo z tym ograniczeniem poprzez gromadzenie danych z kolejnych lat i liczenie wskaźników wieloletnich.

Stosowanie modeli EWD do polskich danych egzaminacyjnych napotyka na inne jeszcze ograniczenie. Pomiar osiągnięć szkolnych na kolejnych etapach edukacyjnych i w kolejnych edycjach egzaminów na niezależnie tworzonych skalach – np. wynik na sprawdzianie i egzaminie gimnazjalnym z matematyki nie jest wyrażany na jednej skali; wynik z danego roku nie jest wyrażany na tej samej skali co w innych latach – sprawia, że miary EWD mogą mieć jedynie relatywny sens. Dodatni wskaźnik EWD szkoły mówi, że efektywność nauczania jest w niej w porównaniu ze szkołami w kraju w danym roku ponadprzeciętna, ujemna, lub poniżej przeciętnej. Gdy dysponujemy wynikami testowania porównywalnymi tylko w obrębie danego egzaminu i danego roku, to nie sposób odpowiedzieć na pytanie, czy EWD w skali kraju w kolejnych latach rośnie, czy maleje. Należy jednak wyraźnie powiedzieć, że na potrzeby procesu ewaluacji pracy szkół relatywne miary są bardzo przydatne i nie jest to mankament znacząco zmniejszający użyteczność wskaźników EWD.

Metoda EWD może być rozpatrywana w ogólnym kontekście polityki poprawiania jakości oświaty oraz w bardziej swoistym otoczeniu problemów pojawiających się wszędzie tam, gdzie szkoły zaczynają ze sobą konkurować o uczniów. Badania nad funkcjonowaniem systemów oświatowych, w których szkoły rywalizują o uczniów i są finansowane zależnie od ich liczby, wskazują, że poza ewentualnymi korzyściami płynącymi z tych rozwiązań pojawiają się zagrożenia. Najczęściej wskazuje się na silne różnicowanie systemu szkół. Różnice między szkołami zwiększają się. Najślabsze szkoły nie są eliminowane, ale z różnych powodów z coraz mniejszą liczbą uczniów trwają na rynku. Najlepsze placówki, wykorzystując mechanizm zwany „spijaniem śmietanki”, umacniają swoje pozycje. Różnicowanie to zagraża ważnemu celowi polityki oświatowej, jakim jest równość szans edukacyjnych. Z procesem różnicowania szkół mamy też do czynienia w Polsce (por. Dolata, 2008; Dolata, Jasińska i Modzelewski, 2012; Zawistowska, 2012). Ocena szkoły na podstawie wskaźnika EWD może zmniejszyć nacisk na selekcję na wejściu do szkoły, zaś ocena nauczycieli według tej miary może podnieść atrakcyjność pracy w szkołach działających w mniej korzystnych warunkach społecznych. Nie ma dowodów na to, że tak się dzieje, ale jest szansa, która może się zrealizować. Oczywiście EWD z pewnością nie zahamuje segregacji społecznych i ekonomicznych w skali makro, ale w dłuższej

perspektywie, jeżeli będzie elementem szerszej polityki edukacyjnej, w której równość szans będzie brana na serio pod uwagę, może przyczynić się do rewitalizacji szkół pracujących w środowiskach społecznie, kulturowo i ekonomicznie upośledzonych.

Rozwój metody edukacyjnej wartości dodanej w Polsce

W Polsce wykorzystanie modeli EWD w analizie wyników egzaminacyjnych stało się możliwe w 2005 roku. W tym właśnie roku gimnazja ukończył pierwszy rocznik absolwentów, dla których był dostępny zarówno wynik sprawdzianu w klasie VI, jak i egzaminu gimnazjalnego. W 2006 roku ukazały się pierwsze polskie publikacje pokazujące możliwości metody EWD w analizie wyników gimnazjalnych (Dolata, 2006a, 2006b). W tekstach tych wykorzystywano proste modele regresji, w których zmienną wyjaśnianą był wynik z danej części egzaminu gimnazjalnego, a zmienną wyjaśniającą wynik na sprawdzianie w klasie VI szkoły podstawowej. Wskaźniki EWD dla gimnazjów czy oddziałów klasowych szacowane były punktowo, jako średnie reszt z równania regresji. Ze względu na brak scalonych krajowych baz danych z połączonymi wynikami sprawdzianu i egzaminu gimnazjalnego analizy prowadzono na poziomie lokalnym (np. gimnazja w danym mieście). Najważniejszym osiągnięciem tego okresu było uświadomienie decydentom potencjału analitycznego tkwiącego w połączonych, podłużnie analizowanych wynikach egzaminacyjnych. Dzięki temu powołano przy Centralnej Komisji Egzaminacyjnej zespół badawczy ds. rozwoju metody EWD (pierwszy projekt EFS, 2005–2006).

Pierwszy, skromny budżetowo projekt, pozwolił rozpocząć prace nad rozwojem metody EWD na potrzeby polskiego systemu egzaminacyjnego. Pomoc okręgowych komisji egzaminacyjnych umożliwiła scalenie wyników sprawdzianu i egzaminu gimnazjalnego w jedną zintegrowaną bazę danych. Dzięki temu mogła powstać pierwsza edycja Kalkulatora EWD dla gimnazjów. Był to arkusz kalkulacyjny z zaimplementowanym algorytmem wyliczania wskaźników EWD dla szkół lub innych grup uczniów (oddziały klasowe, chłopcy, dziewczęta itp.). Algorytm wykorzystywał oszacowane dla danych krajowych wartości przewidywane dla uczniów o danym wyniku na sprawdzianie. W estymacji wartości przewidywanych wykorzystywano dodatkowo zmienne kontrolne dostępne w systemie egzaminacyjnym – informacje o płci i dysleksji. Wskaźniki EWD szacowane były punktowo i przedziałowo. Szkoły mogły bezpłatnie pobierać ze strony CKE Kalkulator EWD i po wprowadzeniu danych swoich absolwentów przeprowadzać analizy na potrzeby ewaluacji wewnętrznej. W ramach projektu odbyło się pierwsze, kaskadowe szkolenie dyrektorów gimnazjów w wykorzystaniu metody EWD w ewaluacji wewnętrznej. W 2007 roku powstała kolejna edycja Kalkulatora EWD, równolegle testowano bardziej złożone statystycznie modele szacowania EWD. W roku tym uruchomiono też przy CKE kolejny zespół badawczy ds. rozwoju metody EWD (EFS, 2007–2013). Dysponował on już znacznym budżetem i możliwościami działania.

W 2008 roku powstała strona www.ewd.edu.pl. Z czasem strona przerodziła się w portal wiedzy nt. wykorzystania wyników egzaminacyjnych do ewaluacji pracy szkół. W tym samym roku odbyło się kolejne masowe szkolenie potencjalnych użytkowników metody EWD pod nazwą Wiosenna Szkoła EWD. Szkoły EWD stały się w następnych latach główną formą upowszechniania wiedzy o metodzie i możliwościach jej wykorzystania. W 2008 roku ukazała się też przygotowana przez członków zespołu pierwsza polska monografia naukowa poświęcona metodzie *Edukacyjna wartość dodana jako metoda oceny efektywności nauczania* (Dolata, 2007). W 2008 roku miało miejsce jeszcze jedno ważne wydarzenie – opublikowano raport OECD opracowany przez międzynarodową grupę ekspertów poświęcony metodzie EWD *Measuring Improvements in Learning Outcomes* (2008). Jednym z ekspertów był Maciej Jakubowski, członek polskiego zespołu EWD. Następny rok zaowocował opracowaniem całkowicie nowego Kalkulatora EWD. Pod nazwą Kalkulator EWD Plus kryła się samodzielna aplikacja komputerowa do analiz wyników egzaminacyjnych dla gimnazjów. W tym samym roku po raz pierwszy zespół w trybie pilotażowym upublicznił na stronie projektu trzyletnie wskaźniki egzaminacyjne dla gimnazjów. Wskaźniki w nowatorski, graficzny sposób pokazywały pozycję danego gimnazjum w dwuwymiarowej przestrzeni – jedna oś to średni wynik egzaminu, druga

to miara EWD. O ile Kalkulator EWD z założenia był przeznaczony do wykorzystania w ewaluacji wewnątrzszkolnej efektywności nauczania, to publikowane w Internecie, powszechnie dostępne wskaźniki trzyletnie miały służyć szerszej grupie odbiorców. Założonymi użytkownikami tych wskaźników były oczywiście szkoły, ale również organy nadzoru pedagogicznego i organy prowadzące. W końcu 2009 roku rozpoczęła się też realizacja trzech badań podłużnych mających dostarczyć informacji potrzebnych do rozwoju metody EWD, w tym do oceny jej trafności. Dwa z nich były – w szkołach podstawowych i gimnazjach - realizowane przez Zespół EWD, natomiast badanie w szkołach ponadgimnazjalnych na zlecenie CKE prowadził IFIS PAN (informacje o badaniach patrz: www.ewd.edu.pl oraz www.ifispan.waw.pl). Pokłosem tych realizowanych w kolejnych latach studiów empirycznych są dostępne na stronie projektu bazy danych wraz z pełną dokumentacją badawczą oraz trzy ważne monografie: *Ścieżki rozwoju edukacyjnego młodzieży – szkoły pogimnazjalne* (Karwowski 2013), *Trafność metody edukacyjnej wartości dodanej dla gimnazjów* (Dolata i in., 2013) oraz *Kontekstowy model oceny efektywności nauczania po pierwszym etapie edukacyjnym* (Dolata i in., 2014). Wnioski z tych prac zostały wykorzystane w niniejszym rozdziale.

W 2010 roku upubliczniono po raz pierwszy wskaźniki egzaminacyjne dla liceów ogólnokształcących i techników w zakresie matematyki i języka polskiego. Graficzna postać wskaźników była analogiczna jak w wypadku gimnazjów, jednak proces ich wyliczania był bardziej złożony. Na podstawie wykonania zadań maturalnych na poziomie podstawowym i rozszerzonym szacowano za pomocą metody IRT poziom umiejętności matematycznych i polonistycznych maturzystów, następnie, analogicznie jak w wypadku wskaźników gimnazjalnych, szacowano średni wynik i EWD szkoły.

W 2011 roku udoskonalono modele EWD dla gimnazjów, uporządkowano bazę szkół, dzięki czemu można było lepiej przypisywać wyniki egzaminacyjne szkołom, co stało się palącą potrzebą w obliczu publikowania wskaźników EWD dla kilku okresów trzyletnich dla każdej szkoły. W tym samym roku znacząco zmodyfikowano wskaźniki dla LO i techników. Wprowadzono trzy syntetyczne wskaźniki: umiejętności w zakresie przedmiotów humanistycznych, matematyczno-przyrodniczych i oddzielnie matematyki. Wspólne skalowanie zadań z arkuszy egzaminacyjnych z różnych przedmiotów wymagało zastosowania odpowiednich modeli statystycznych, by rozwiązać problem różnic w poziomie umiejętności subpopulacji przystępujących do egzaminu maturalnego z różnych przedmiotów i na różnych poziomach.

Zmiana struktury egzaminu gimnazjalnego wymusiła w 2012 roku opracowanie nowej aplikacji komputerowej – Kalkulatora EWD 100. Narzędzie to, choć funkcjonalnie podobne do Kalkulatora EWD, ma inną architekturę, dzięki której łatwo można implementować nowe funkcjonalności i nowe typy wskaźników egzaminacyjnych. Ważną zmianą jest też przejście w Kalkulatorze EWD 100, analogicznie do wskaźników wieloletnich, na skalę standaryzowane o średniej w skali kraju 100 i odchyleniu standardowym 15. W 2012 roku opublikowano również trzyletnie wskaźniki egzaminacyjne dla LO i techników w zakresie przedmiotów humanistycznych, matematyczno-przyrodniczych i oddzielnie polskiego i matematyki.

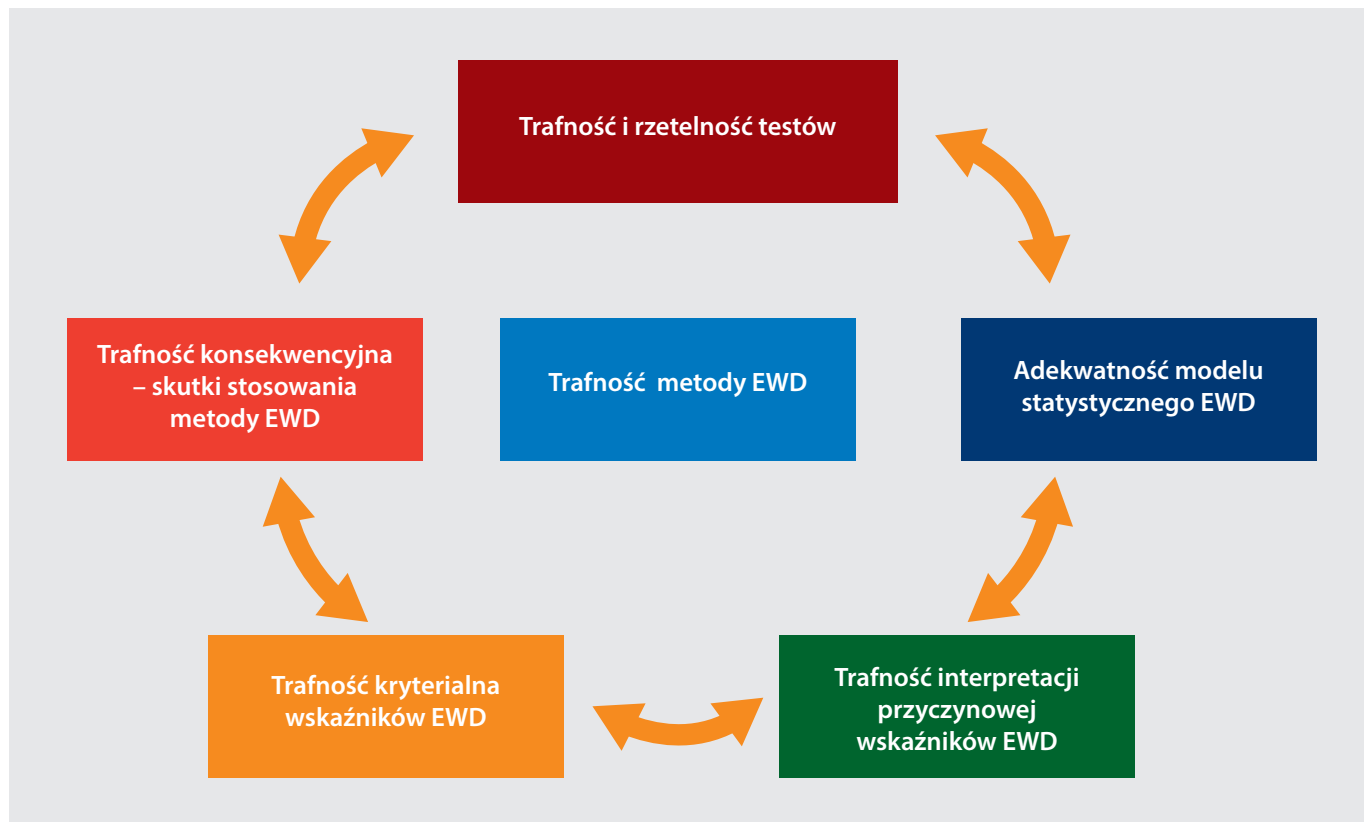
W roku 2013 rozpoczął się zakrojony na blisko trzy lata kolejny projekt finansowany ze środków unijnych, dedykowany rozwojowi metody EWD w Polsce. W latach 2013–2015 trwały prace rozwojowe dotyczące modeli szacowania EWD (między innymi jednoroczny model EWD dla maturalnej matematyki na potrzeby Kalkulatora EWD 100), porządkowano bazy danych, stworzoną nową stronę internetową oraz prowadzono działania upowszechniające i badawcze dotyczące wykorzystania metody EWD w ewaluacji nauczania.

Problem trafności wskaźników EWD

W dotychczasowych pracach nad rozwojem metody EWD w Polsce dużą wagę przykładano do weryfikacji trafności wskaźników EWD. Przed konstruktorami modeli EWD stoi wiele trudnych do pokonania przeszkód. Na podstawie współczesnych ujęć trafności i wymagań stawianych testom używanym w modelowaniu EWD można wskazać na pięć powiązanych z sobą, ale dających się

analitycznie rozdzielić aspektów badania trafności metody EWD. Schematycznie przedstawia je poniższy rysunek.

Rysunek 4.1. Aspekty badania trafności metody EWD



Pierwszy aspekt trafności wiąże się z faktem, że podstawowych danych do modelowania dostarczają testy egzaminacyjne. Jakość tych danych jest warunkiem koniecznym trafności metody EWD. W zakresie jakości danych testowych przede wszystkim stawia się pytanie o to, czy wykorzystane do modelowania EWD testy są dobrą miarą realizacji celów nauczania. Wiąże się to z problemem zakresu dozwolonej generalizacji treściowej wyników testu. Test jest zawsze tylko próbką wymagań programowych, ale jeżeli próbka została dobrze dobrana, to wyniki testu można z dającym się zaakceptować błędem uogólniać na całość wymagań programowych (przynajmniej tę część, która była podstawą budowy testu). Ważne jest zatem, by sprawdzić, czy wyniki testów, które wykorzystujemy w modelowaniu EWD, pozwalają na takie uogólnienie. Jednak sam fakt, że dwa testy dobrze reprezentują wymagania programowe na poziomie kolejnych lat nauczania, jest niewystarczający. Jeżeli wykorzystywane testy nie pozwalają śledzić zmian w poziomie umiejętności na kolejnych etapach edukacji, to należy sprawdzić, czy testy wykorzystane do szacowania danego wskaźnika EWD mierzą ten sam lub przynajmniej bardzo zbliżony konstrukt. Jest to konieczne, by móc interpretować uzyskane wskaźniki jako miary względnego przyrostu osiągnięć szkolnych. To założenie jest w praktyce trudne do spełnienia, bo nawet egzaminy z tego samego przedmiotu, np. egzamin gimnazjalny i matura z matematyki, obejmują różne treści matematyczne, np. składają się z różnych proporcji treści z algebry czy geometrii. Ważna jest również rzetelność pomiaru osiągnięć szkolnych i to, by jego wyniki nie były obciążone działaniem czynników niezwiązanych z poziomem osiągnięć. Ponadto skala, na której wyrażane są wyniki pomiaru, powinna być skalą co najmniej interwałową. Drugi aspekt trafności metody EWD wiąże się ze statystycznym modelowaniem postępu. Jak właściwie uwzględnić w modelowaniu hierarchiczny charakter danych edukacyjnych: szkoła–oddział–uczeń? Jak uwzględnić zjawisko regresji ku średniej? Za pomocą jakiej statystyki opisywać

przeciętny postęp w zakresie osiągnięć szkolnych uczniów w danej szkole? Jak uwzględnić różne przedmioty egzaminacyjne? Jakie zmienne kontrolne uwzględniać w modelowaniu?

Ostatni problem – zmienne kontrolne – prowadzi nas bezpośrednio do trzeciego aspektu trafności metody EWD: przyczynowej interpretacji wskaźników EWD. Jeżeli wskaźniki EWD mają być użyteczne, to musimy mieć argumenty pozwalające oszacowany za pomocą modelu EWD postęp w zakresie osiągnięć szkolnych zasadnie przypisać działaniom szkoły, czyli wskaźniki efektywności nauczania powinny dawać się sensownie interpretować w kategoriach przyczynowo-skutkowych. Ponieważ wykorzystywane w EWD dane nie są zbierane w schemacie eksperymentalnym, w którym uczniowie byłoby losowo przypisywani do szkół, nie możemy konkluzywnie wypowiadać się na podstawie modeli EWD o przyczynowo-skutkowych zależnościach między zmiennymi (Rubin, Stuart i Zanutto, 2004). Innymi słowy nie mamy pewności, że szkoła mająca wysokie wskaźniki EWD miałaby podobne wskaźniki, gdyby uczęszczali do niej inni uczniowie o tych samych wynikach w momencie rozpoczęcia nauki w szkole. W modelowaniu EWD możemy jedynie próbować kontrolować czynniki, które są niezależne od szkoły, a rzutują na postępy uczniów. W sytuacji nielosowego przydziału uczniów do szkół – co ma miejsce właściwie w każdym systemie szkolnym, choć z różnym nasileniem – część wpływu przypisywanego działaniom szkoły może być de facto wynikiem oddziaływania zmiennych zewnętrznych wobec szkoły, nieuwzględnianych w obliczaniu EWD. Możemy wprawdzie włączać tego rodzaju zmienne do modelu – np. zmienną wskazującą na to, czy szkoła działa na wsi, czy w mieście – ale utrudni to rozróżnianie od siebie szkół pracujących efektywnie i nieefektywnie. Działoby się tak dlatego, że włączenie do modelu danej zmiennej kontrolnej oznacza, że zakładamy, że nie jest ona powiązana z efektywnością nauczania. Założenie to ma sens w wypadku zmiennej płci ucznia, ale jest mocno problematyczne w wypadku lokalizacji szkoły. Jeżeli obiektywnie bardziej efektywnie nauczają szkoły miejskie (albo wiejskie), to włączenie tej zmiennej kontrolnej do modelu szacowania EWD szkoły zaniży szacunki dla tych szkół. Niestety, nawet najbardziej złożone modele statystyczne nie są w stanie w pełni rozwiązać problemu nielosowego przydziału uczniów do szkół i oddziałów klasowych. Interpretację wskaźnika jako informacji o efektywności szkoły utrudniać też mogą problemy z oddzieleniem efektu szkoły od wpływu innych czynników, takich np. dynamika procesu grupowego w klasie czy zmiany w środowisku społeczno-ekonomicznym, w którym działa szkoła, z jakich tylko część można przypisać działaniom szkoły (Braun i Weiner, 2007). Choć problem przyczynowej interpretacji wskaźników EWD nie może być w pełni rozwiązany, to badawcza weryfikacja trafności wskaźników EWD może rozwiązać wiele wątpliwości.

Kolejnym aspektem trafności metody EWD jest skonfrontowanie miar postępu z zewnętrznymi w stosunku do wyników nauczania kryteriów. O tym, czy interpretacja przyczynowo-skutkowa jest zasadna, możemy wnioskować nie tylko na podstawie skuteczności kontroli zmiennych powiązanych z selekcją międzyszkolną, ale również patrząc, czy szkoły o wysokim EWD to szkoły, w których zgodnie z naszą normatywną wiedzą o funkcjonowaniu szkoły dobrze się dzieje.

Ostatni wyróżniony na schemacie aspekt trafności metody EWD to konsekwencje stosowania tej metody z punktu widzenia ewaluacyjnej funkcji egzaminów zewnętrznych. Czy metoda służy skutecznej ewaluacji i autoewaluacji szkół, czy jest motorem doskonalenia nauczania? Ten problem jest analizowany w ostatnim rozdziale raportu.

Struktura rozdziału⁴⁴

Pierwszy podrozdział poświęcony jest kontekstowym modelom oceny efektywności nauczania dla I etapu edukacyjnego. Dla tego etapu nie może być stosowana standardowa metoda EWD, ponieważ nie dysponujemy i nie będziemy dysponować pomiarem osiągnięć szkolnych na progu

⁴⁴ W rozdziale wykorzystano materiał faktograficzny i fragmenty dwóch prac, które powstały w zespole EWD: Dolata R., Hawrot A., Hummeny G., Jasińska A., Koniewski M., Majkut P. i Żółtak T., (2013) *Trafność metody edukacyjnej wartości dodanej dla gimnazjów*. Warszawa, Instytut Badań Edukacyjnych. Dolata, R., Hawrot, A., Humenny, G., Jasińska-Maciążek, A., Koniewski, M. i Majkut, P., (2014). *Kontekstowy model oceny efektywności nauczania po pierwszym etapie edukacyjnym*. Warszawa: Instytut Badań Edukacyjnych.

klasy I. By używać do oceny efektywności nauczania wyników pomiarów osiągnięć przeprowadzanych na zakończenie I etapu edukacji (np. takich jak *Ogólnopolskie Badanie Umiejętności Trzecioklasistów OBUT*), należy w inny sposób kontrolować wpływ pozaszkolnych czynników rzutujących na wyniki uczniów. Wykorzystując dane z przeprowadzonego badania, pokazujemy, że w takiej sytuacji możemy model EWD zastąpić statystyczną kontrolą statusu społecznego i innych cech środowiska rodzinnego ucznia. Drugi podrozdział dotyczy modelu EWD dla II etapu edukacyjnego zbudowanego na podstawie wyników testu diagnostycznego OBUT (miara osiągnięć na „wejściu”) i rezultatów sprawdzianu w klasie VI (miara osiągnięć na „wyjściu”). Przedstawiony został model statystyczny oraz możliwości wykorzystania go przez szkoły. W trzecim podrozdziale opisane jest zastosowanie metody EWD dla gimnazjów. Po charakterystyce modeli statystycznych używanych do szacowania jednorocznych i trzyletnich wskaźników EWD, przedstawiono wyniki badań trafności wskaźników i na kilku przykładach wskazano na możliwości wykorzystania przez szkoły wyników egzaminacyjnych w procesie ewaluacji wewnętrznej. Czwarty, ostatni podrozdział traktuje o zastosowaniu metody EWD w liceach ogólnokształcących i technikach. Analogicznie jak w wypadku gimnazjów, opisano modele statystyczne wykorzystywane do obliczania jednorocznych i trzyletnich wskaźników EWD, wskazano na kluczowe wyniki badania trafności oraz opisano możliwości zastosowania metody przez szkoły. W wypadku wskaźników maturalnych bardzo ważnym i trudnym do rozwiązania problemem jest skalowanie wyników egzaminacyjnych, obszerny zatem fragment tego podrozdziału został poświęcony temu zagadnieniu. Każdy podrozdział wieńczy krótki opis ograniczeń metody i perspektyw jej rozwoju w Polsce.

4.2. Kontekstowy model oceny efektywności nauczania po pierwszym etapie edukacyjnym

Pierwsze lata życia uważane są za najważniejsze dla sukcesu edukacyjnego uczniów. W porównaniu do późniejszych okresów nauczania szkoły podstawowe w Polsce na pierwszym etapie edukacyjnym mają do dyspozycji niewiele narzędzi do diagnozy i monitorowania efektów kształcenia. Prowadzone w ramach projektu *Badania dotyczące rozwoju metodologii szacowania wskaźnika edukacyjnej wartości dodanej* (EWD) analizy podłużne w szkołach podstawowych (więcej o badaniu w ramce 4.1) pozwoliły opracować metodę oceny efektywności nauczania w klasach I–III, która jest przykładem tzw. kontekstowych modeli oceny nauczania. W tej części raportu przedstawione zostaną najważniejsze wyniki tych analiz. Kontekstowe modele oceny efektywności nauczania stosowane są od wielu lat w skali lokalnej w Polsce, a w skali krajowej np. w Australii.

Ramka 4.1. Opis podłużnego badania uwarunkowań wyników nauczania w szkołach podstawowych

Badanie podłużne w szkołach podstawowych rozpoczęto w roku szkolnym 2009/2010. Objęło ono reprezentatywną, ogólnopolską próbę losową 312 oddziałów klas pierwszych ze 180 szkół podstawowych (ok. 5 tys. uczniów). Do roku 2013 badanie było realizowane w ramach projektu *Badania dotyczące rozwoju metodologii szacowania wskaźnika edukacyjnej wartości dodanej* (EWD), a od 2014 roku w ramach projektu *Rozwój metody edukacyjnej wartości dodanej na potrzeby wzmocnienia ewaluacyjnej funkcji egzaminów zewnętrznych*. Badanie obejmuje pełen cykl szkoły podstawowej (śledzone są losy tych samych uczniów) i skończy się w 2015 roku. Uczestniczą w nim uczniowie, ich rodzice, nauczyciele oraz dyrektorzy szkół. Głównym celem badania jest opracowanie metod oceny efektywności nauczania na I i II etapie edukacyjnym. Ponadto zebrane dane mają pozwolić na lepsze poznanie indywidualnych, rodzinnych i szkolnych czynników odpowiedzialnych za osiągnięcia szkolne uczniów.

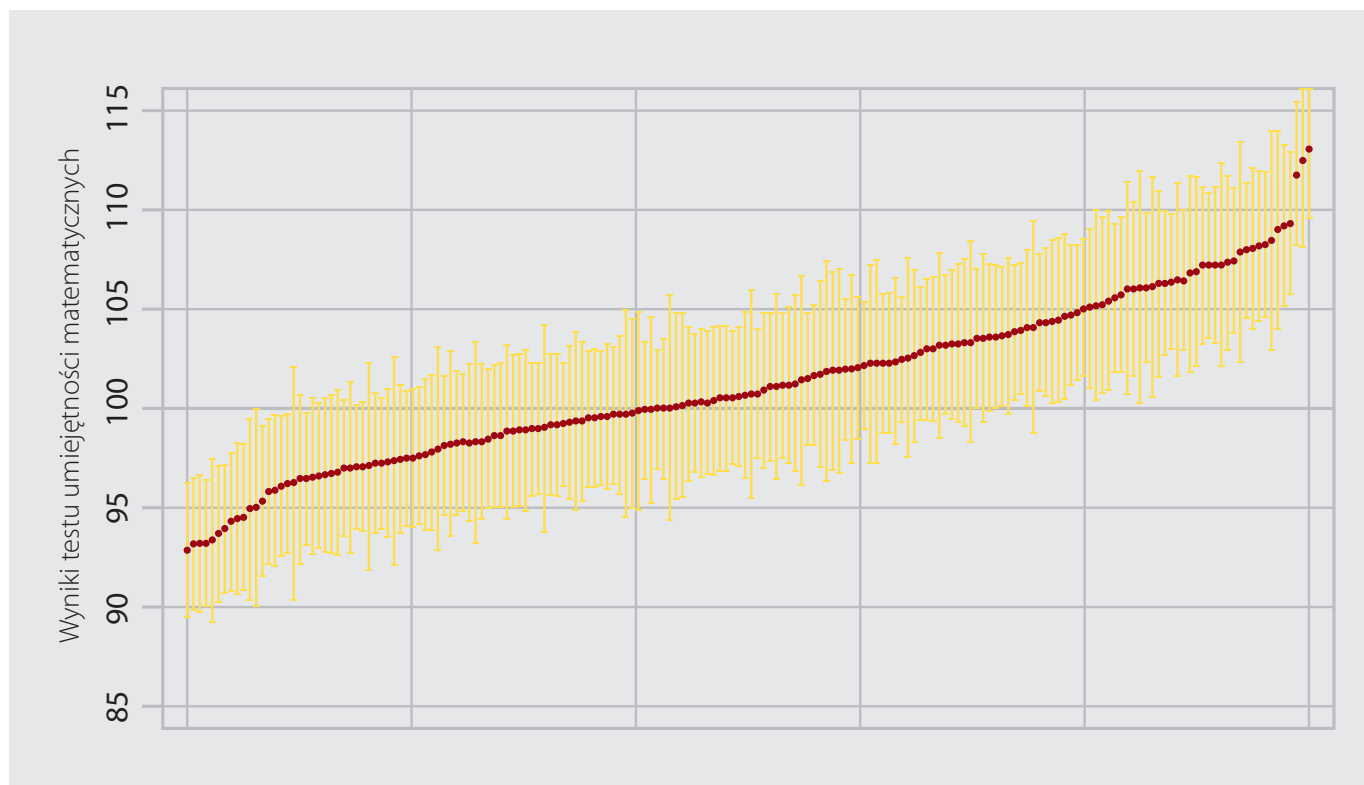
4. Metoda edukacyjnej wartości dodanej w Polsce

Podczas pierwszego etapu badania, który odbył się, kiedy badani uczniowie uczęszczali do pierwszych klas, zebrano m.in. dane opisujące uczniów (płeć, data urodzenia, poziom inteligencji) i ich rodziny (różne miary statusu społeczno-ekonomicznego, aspiracje edukacyjne względem dzieci, struktura rodziny). W kolejnym etapie badania, który odbył się na przełomie III i IV klasy, przeprowadzono m.in. pomiar osiągnięć szkolnych z wykorzystaniem trzech standaryzowanych testów, o udokumentowanej wysokiej jakości (Jasińska i Modzelewski, 2014), mierzących osiągnięcia w zakresie umiejętności czytania, świadomości językowej¹ i umiejętności matematycznych. Więcej informacji na temat badania można znaleźć na stronie: www.ewd.edu.pl.

¹ Test świadomości językowej mierzył zasób słownikowy, elementy wiedzy o języku i umiejętności związane z pisanem tekstów.

Szkoły podstawowe różnią się wynikami uzyskiwanymi przez uczniów w testach osiągnięć szkolnych. Zróżnicowanie to obserwujemy już po pierwszym etapie edukacyjnym, choć, jak pokazały przeprowadzone badania, nie jest ono bardzo duże. Podział na szkoły wyjaśnia 8,6% zróżnicowania wyników testu umiejętności czytania, 11,8% zróżnicowania wyników testu świadomości językowej oraz 10,5% zróżnicowania wyników testu matematycznego. Ilustrację, jak te odsetki przekładają się na realne różnice między szkołami, pokazano na rysunku 4.2. Zaprezentowano na nim średnie wyniki uzyskane przez uczniów w 180 badanych szkołach w teście matematycznym (czarne punkty) wraz z przedziałami ufności (pionowe kreski). Oś pionowa to wynik testu przedstawiony na standardowej skali o średniej krajowej 100 i odchyleniu standardowym 15. Wyniki szkół zostały uporządkowane rosnąco według wartości średniej.

Rysunek 4.2. Zróżnicowanie międzyszkolne wyników nauczania po III klasie na przykładzie testu matematycznego. Średnie wyniki szkół wraz z 90-procentowymi przedziałami ufności



Przedział ufności wskazuje na zakres wyników, w którym z 90-procentowym prawdopodobieństwem znajduje się prawdziwy wynik szkoły. Dwie szkoły istotnie różnią się wynikami, jeśli ich przedziały ufności są rozłączne. Średni wynik w populacji uczniów wynosi 100 punktów. Wobec tego, jeśli cały przedział ufności znajduje się powyżej wartości 100, z dużą pewnością można powiedzieć, że uczniowie w tej szkole osiągnęli przeciętnie wyższy wynik niż w kraju; jeśli poniżej – uzyskali średnio wynik niższy. Jeśli przedział ufności zawiera wartość 100, średnia szkoły na podstawie uzyskanych w badaniu wyników jest statystycznie nierozróżnialna z przeciętną dla kraju.

Rysunek ten pokazuje skalę zróżnicowania międzyszkolnego wyników uczniów w zakresie osiągnięć matematycznych. Mimo że jest duża grupa szkół, które nie różnią się między sobą pod względem średniego poziomu osiągnięć szkolnych swoich uczniów, to widzimy też sporą liczbę szkół o wynikach istotnie powyżej i poniżej przeciętnych. Natomiast różnica między średnimi wynikami szkoły o najniższej i najwyższej średniej wynosi prawie 20 punktów, czyli 1 i 1/3 odchylenia standardowego wyników uczniów. Międzyszkolne zróżnicowanie osiągnięć szkolnych uczniów ze świadomości językowej jest porównywalne, a z zakresu umiejętności czytania trochę mniejsze.

Czy oznacza to, że szkoły, które osiągnęły średnio wyższe wyniki, uczyły lepiej, efektywniej niż te, które uzyskały słabsze rezultaty? Niekoniecznie. Osiągnięcia szkolne uczniów są bowiem częściowo kształtowane przez czynniki, na które szkoła nie ma wpływu. Spośród pozaszkolnych czynników najsilniej determinuje poziom osiągnięć szkolnych uczniów po III klasie szkoły podstawowej inteligencja (Dolata i in., 2014). Ma ona nieco większe znaczenie dla osiągnięć z matematyki niż ze świadomości językowej i umiejętności czytania. Spośród czynników statusowych największe znaczenie ma zasobność gospodarstwa domowego (mierzona zestawem pytań o posiadane dobra mogące wspierać proces uczenia się, jak np. spokojne miejsce do nauki dla dziecka, dostęp do Internetu, aparat cyfrowy, globus, a także liczba książek w domu dziecka) oraz wykształcenie rodziców. Im wyższy poziom zasobności lub wyższy poziom wykształcenia rodziców, tym więcej punktów otrzymują uczniowie na testach osiągnięć. Pozytywny związek stwierdzono także dla aspiracji edukacyjnych rodziców rozumianych jako pożądaną przez rodziców poziom formalnego wykształcenia dla ich dzieci.

Uczniowie, którzy poszli do szkoły rok później niż ich rówieśnicy (mieli więc odroczony start szkolny), uzyskali znacząco niższe wyniki niż średnio ich klasowi koledzy i koleżanki. Z kolei dzieci posłane do szkoły wcześniej – jako sześciolatki – uzyskały wyniki porównywalne z uczniami, którzy rozpoczęli naukę w wieku siedmiu lat. Należy jednak pamiętać, że w roku szkolnym 2009/2010 była to grupa nieliczna i silnie pozytywnie wyselekcjonowana. Podobnie opóźniony start szkolny jest bezpośrednio powiązany z czynnikami mającymi znaczenie dla późniejszych sukcesów szkolnych. Omawianych związków nie można zatem interpretować jako wpływu wieku. Analiza głównej grupy wiekowej wskazuje jednak, że wiek biologiczny uczniów ma znaczenie. Uczniowie starsi (urodzeni bliżej początku roku) uzyskali trochę wyższe wyniki w testach osiągnięć (Dolata i in., 2014).

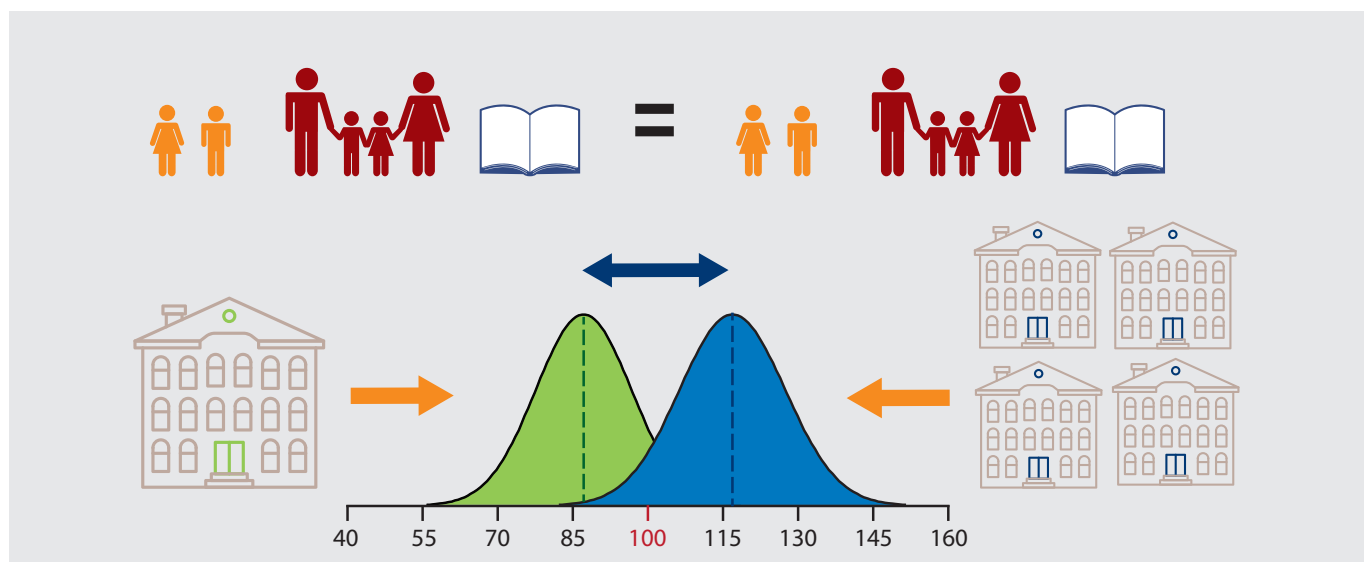
Gdyby szkoły nie różniły się poziomem wymienionych cech uczniów, każda z nich miałaby takie same szanse na uzyskanie podobnych średnich wyników. Tak jednak nie jest. Badania pokazały, że szkoły znacząco różnią się pod względem cech uczniów powiązanych z osiągnięciami szkolnymi. Największe zróżnicowanie stwierdzono dla inteligencji uczniów oraz rodzinnych czynników statusowych, takich jak wykształcenie rodziców, wskaźnik statusu społeczno-ekonomicznego, zasobność rodziny czy prestiż wykonywanego przez rodziców zawodu. Szkoły różniły się także pod względem aspiracji rodziców wobec wykształcenia ich dzieci. Najmniejsze, choć także istotne statystycznie, zróżnicowanie międzyszkolne zaobserwowano dla wieku uczniów.

Kontekstowy model oceny efektywności nauczania

Rezultaty przytaczanych badań potwierdziły, że to, jakie wyniki nauczania uzyskuje szkoła, zależy nie tylko od wysiłku nauczycieli czy oferty edukacyjnej, ale także od tego, jacy uczniowie do niej uczęszczają. Żeby więc móc ocenić wkład danej szkoły w wyniki nauczania, należy porównać je z wynikami szkół, które pracowały z podobnymi uczniami i w podobnym środowisku. Porównań

takich dokonuje się z wykorzystaniem złożonych metod statystycznych. W pewnym uproszczeniu ideę stosowanych modeli analizy ilustruje poniższy rysunek.

Rysunek 4.3. Idea kontekstowych wskaźników efektywności nauczania



Wyniki nauczania mierzone testem uzyskane w danej szkole A (przedstawione na rysunku za pomocą zielonego rozkładu) porównujemy z wynikami szkół pracujących z uczniami o podobnych cechach (niebieski rozkład). Średni wynik w tych szkołach (niebieska przerywana linia) jest punktem odniesienia, to najbardziej prawdopodobny wynik, jaki może uzyskać szkoła, dla której są one grupą odniesienia (takie samo nasilenie cech pozaszkolnych decydujących o osiągnięciach szkolnych). Nazywamy go wynikiem statystycznie przewidywanym, ponieważ – na podstawie danych uzyskanych w badaniu – taki właśnie średni w skali kraju wynik w teście uzyskano, pracując z uczniami o założonych cechach (założonych, czyli takich jak w szkole A).

Następnie wynik przez szkołę faktycznie uzyskany (średni wynik jej uczniów – oznaczony zieloną przerywaną linią) porównujemy z wynikiem dla niej przewidywanym (najbardziej prawdopodobnym). Różnica to wartość kontekstowego wskaźnika efektywności nauczania. Nazywamy go kontekstowym, bo pozwala na porównanie szkół podobnych do siebie ze względu na wiele ważnych cech opisujących kontekst pracy szkoły.

Jeśli w danej szkole uzyskano średnio wyniki niższe niż wartość statystycznie oczekiwana (tak jak na rysunku 4.3), wartość wskaźnika będzie ujemna. Będzie to świadczyć o poniżej przeciętnej efektywności nauczania. Jeśli szkoła uzyskała wynik wyższy niż przewidywany, to wartość wskaźnika efektywności nauczania będzie dodatnia. Będzie to oznaczało ponadprzeciętną efektywność. Jeśli uzyskany przez szkołę średni rezultat i wynik przewidywany będą podobne, wskaźnik efektywności będzie przyjmował wartość w okolicy zera. Będzie to świadczyć o tym, że szkoła naucza z efektywnością taką jak większość szkół w kraju.

Wskaźnik efektywności nauczania jest więc liczony w taki sposób, że szansa na jego wysoką wartość nie zależy od tego, z uczniami o jakich cechach (uwzględnionych w modelu) pracowała szkoła. Szkoła, która świetnie pracowała z uczniami z niekorzystnego środowiska społecznego, osiągnie dodatnią wartość wskaźnika, mimo że jej średnie wyniki w testach osiągnięć będą prawdopodobnie niższe niż szkoły, która słabo pracowała z uczniami pochodzącymi z „dobrych domów”.

Prace nad kontekstowymi wskaźnikami efektywności nauczania były możliwe dzięki badaniu przeprowadzonemu z udziałem 180 szkół podstawowych (patrz ramka 4.1). Dane zebrane w jego toku wykorzystano do zbudowania oraz przetestowania kilku alternatywnych modeli i wyboru optymalnego. Obliczono także wskaźniki dla 180 szkół biorących udział w projekcie.

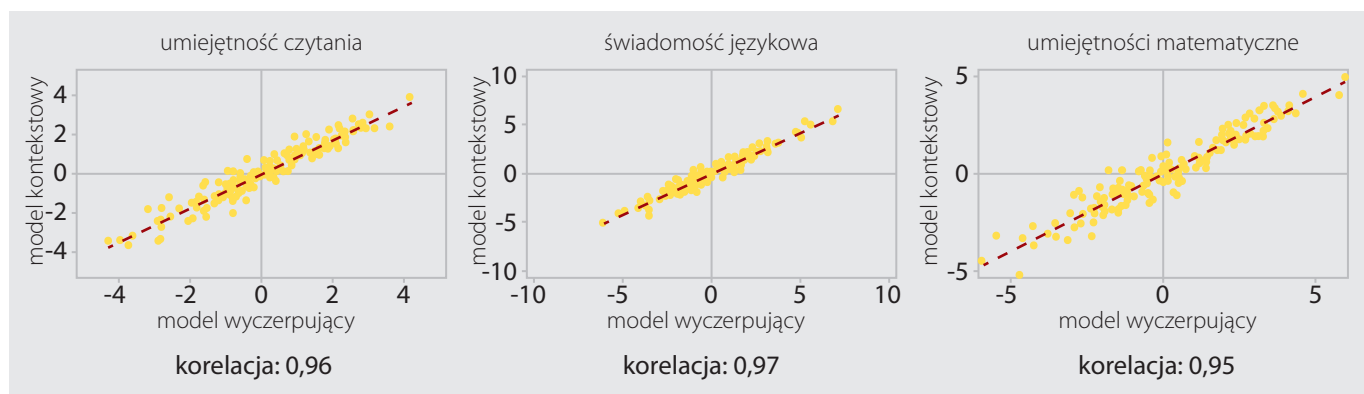
Jakie cechy uczniów uwzględniono w kontekstowych modelach oceny efektywności nauczania? Wybierając zmienne opisujące środowisko pracy szkoły, kierowano się dwoma kryteriami. Po pierwsze, wybrane charakterystyki miały wyjaśniać znaczącą część indywidualnego i międzyszkolnego zróżnicowania osiągnięć szkolnych. Po drugie, wykorzystane informacje musiały być możliwe do pozyskania przez szkoły. W sytuacji badawczej można było zebrać znacznie więcej informacji (np. o poziomie inteligencji uczniów), jednak gdyby wskaźniki kontekstowe miały być wyliczane na większą skalę, trzeba by opierać się przede wszystkim na takich danych, które są możliwe do zebrania w sposób wiarygodny przez szkoły. Przeprowadzone badanie pozwoliło jednak sprawdzić, czy taki zdawałoby się ograniczony zbiór informacji można uznać za wystarczający.

W kontekstowym modelu oceny efektywności nauczania po pierwszym etapie edukacyjnym spośród zmiennych opisujących rodzinę ucznia wykorzystano wskaźniki statusu społeczno-ekonomicznego rodziny ucznia (informację o wykształceniu rodziców i liczbie posiadanych książek, która jest wyrazem zarówno kapitału kulturowego rodziny, jak i jej zasobności) oraz informację o aspiracjach edukacyjnych rodziców względem ich dzieci. W modelu uwzględniono także informację o płci i wieku uczniów. Dodatkowo środowisko, w którym pracuje szkoła, opisano za pomocą średniej wskaźnika wykształcenia rodziców uczniów w szkole. Oznacza to, że kontekstowe wskaźniki efektywności nauczania pokazują wyniki szkół na tle innych, które pracują w takich samych warunkach, ze względu na wymienione charakterystyki. Wskaźniki takie wyliczono dla trzech obszarów osiągnięć szkolnych odpowiadającym trzem wykorzystanym w badaniu testom: umiejętności czytania, świadomości językowej i umiejętności matematycznych.

Trafność kontekstowych wskaźników efektywności nauczania

Skąd jednak wiemy, że uwzględnione w modelu zmienne w wystarczającym stopniu pozwalają uwzględnić to, że szkoły pracują z uczniami o różnym potencjale? Dzięki przeprowadzonemu badaniu, w którym zebrano informację o wielu innych cechach uczniów, ważnych dla ich przyszłych osiągnięć, możliwe było wyliczenie wskaźników uwzględniających szeroki wachlarz charakterystyk opisujących kontekst pracy szkoły i porównanie ich z zaproponowanymi wskaźnikami efektywności nauczania. W rozbudowanych modelach (nazwanych wyczerpującymi), poza zmiennymi uwzględnionymi podczas wyliczania omówionych kontekstowych wskaźników efektywności nauczania, wzięto dodatkowo pod uwagę informację o poziomie inteligencji uczniów (jako charakterystykę indywidualną uczniów oraz średnią w szkole), zasobności rodziny, prestiżu wykonywanych przez rodziców zawodów oraz strukturze rodziny. Wskaźniki obliczone z modeli wyczerpujących porównano z proponowanymi wskaźnikami kontekstowymi. Zestawienie to dla każdego z trzech obszarów osiągnięć pokazano na poniższym rysunku.

Rysunek 4.4. Porównanie kontekstowych wskaźników efektywności nauczania ze wskaźnikami obliczonymi z modeli wyczerpujących



4. Metoda edukacyjnej wartości dodanej w Polsce

Na osiach poziomych opisano wartości wskaźników obliczonych z modeli wyczerpujących, na osiach pionowych – wartości wskaźników z proponowanych modeli kontekstowych. Kropki pokazują pozycje badanych szkół na tych dwóch wymiarach. Pod wykresami podano współczynniki korelacji między dwoma wskaźnikami. Jeśli korelacja byłaby równa 1, oznaczałoby to, że wskaźniki wyliczone z obu modeli przyjmują takie same wartości, a punkty na wykresach ułożyłyby się na prostej linii. Jeśli współczynnik korelacji wyniósłby 0, oznaczałoby to, że nie ma żadnego związku między wartościami wskaźników obliczonych z obu modeli, więc punkty na wykresach przyjęłyby kształt rozproszonej chmury.

Prezentowane wyniki pokazały, że wskaźniki wyliczone z proponowanych modeli kontekstowych przyjmują wartości bardzo podobne do tych z modeli zawierających znacznie więcej charakterystyk opisujących uczniów i środowisko pracy szkoły. Oznacza to, że mimo iż w proponowanych modelach nie uwzględniamy np. informacji o inteligencji uczniów czy prestiżu zawodu rodziców, to pozostałe zmienne w wystarczającym stopniu pozwalają opisać międzyszkolne zróżnicowanie niezależnych od szkoły cech uczniów mających znaczenie dla osiągnięć uczniów. Jest to jeden z argumentów na rzecz trafności proponowanych wskaźników.

Innym spojrzeniem na problem trafności proponowanych wskaźników jest sprawdzenie, czy w stosunku do nieprzetworzonych wyników testów są one znacząco słabiej zależne od czynników pozaszkolnych wyznaczających osiągnięcia szkolne. Przeprowadzone badania potwierdziły, że średnie wyniki testów osiągnięć uczniów w szkołach są silnie determinowane przez kontekst pracy szkoły. Zależą one od charakterystyk takich jak: średni poziom inteligencji uczniów w szkole czy średni status społeczno-ekonomiczny (mierzony wykształceniem rodziców, wskaźnikiem prestiżu wykonywanego zawodu lub wskaźnikiem zasobności gospodarstwa domowego). Wartości kontekstowych wskaźników efektywności nauczania niemal trzykrotnie słabiej zależą od średniego poziomu inteligencji uczniów, a co więcej, nie są powiązane ze statusem społeczno-ekonomicznym rodzin uczniów w szkole. W tabeli 4.1 zaprezentowano współczynniki korelacji między wskaźnikami opisującymi kontekst pracy szkoły a średnimi wynikami w testach osiągnięć oraz kontekstowymi wskaźnikami efektywności nauczania. Pogrubione wartości wskazują, że korelacja jest istotna statystycznie (współczynnik korelacji dla związków pozytywnych może przyjmować wartość od 0, co oznacza brak związku, do 1 – liniowa zależność).

Tabela 4.1. Korelacje średnich wyników szkół oraz efektywności nauczania z miarami kontekstu pracy szkoły

Wskaźniki wyliczone jako średnia dla szkoły	Umiejętność czytania		Świadomość językowa		Umiejętności matematyczne	
	średni wynik testu	wsk. efektywności	średni wynik testu	wsk. efektywności	średni wynik testu	wsk. efektywności
Średnia inteligencja uczniów ¹	0,563	0,208	0,547	0,175	0,569	0,197
Średnie wykształcenie rodziców ²	0,731	0,014	0,737	0,010	0,720	-0,005
Średnia wskaźnika prestiżu wykonywanego przez rodziców zawodu	0,697	0,015	0,694	0,002	0,704	0,027
Średnia zasobność gospodarstwa domowego ³	0,704	0,049	0,699	0,035	0,704	0,034

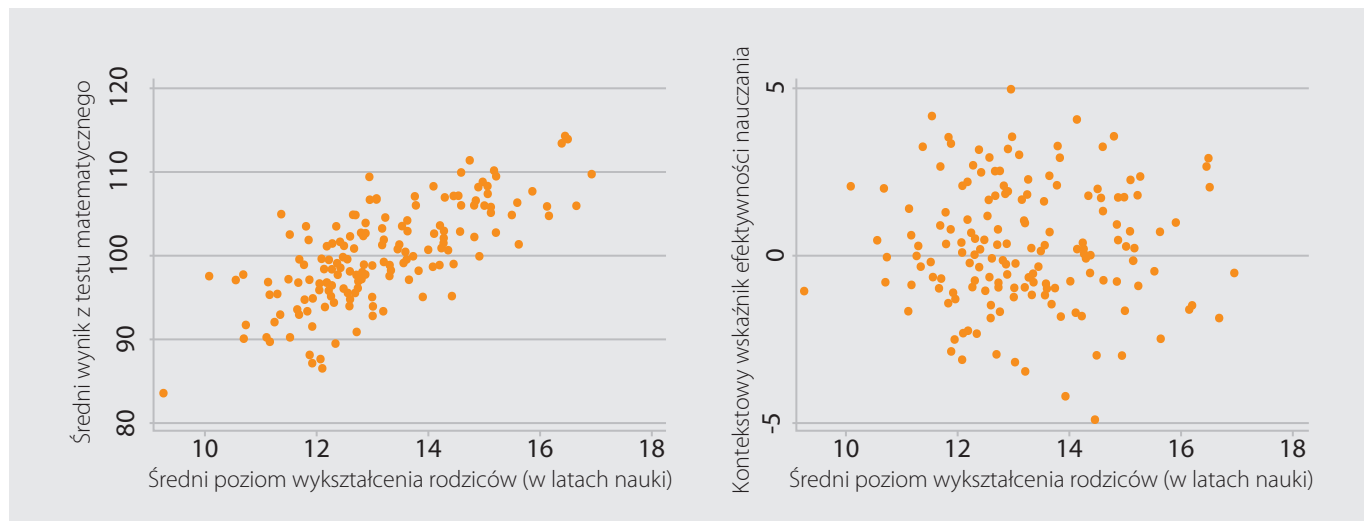
¹ Wykorzystano Test Matryc Ravena.

² Wykształcenie rodziców mierzone w latach nauki potrzebnych do osiągnięcia danego poziomu wykształcenia.

³ Zastosowano skalę wykorzystywaną w badaniach PISA OECD.

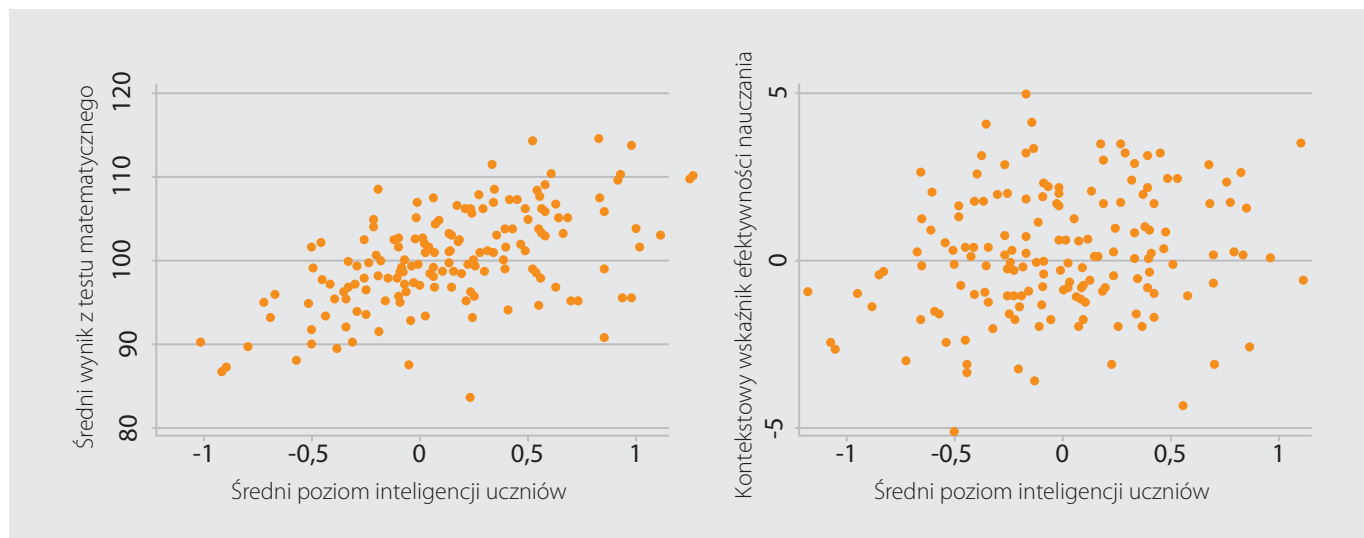
Ilustrację tego, co w praktyce oznaczają podane wartości, zaprezentujemy na przykładzie testu matematycznego. Na rysunku 4.5. przedstawiono zależność wyników testu od poziomu wykształcenia rodziców w szkole (z lewej) oraz relację między wskaźnikiem efektywności nauczania a wykształceniem rodziców uczniów (z prawej). Jedna kropka reprezentuje jedną szkołę.

Rysunek 4.5. Związek średniego poziomu wykształcenia rodziców uczniów ze średnimi wynikami szkół (z lewej) oraz z efektywnością nauczania mierzoną wskaźnikami kontekstowymi (z prawej) dla obszaru umiejętności matematycznych



Wykres z lewej strony bardzo dobrze pokazuje, jak silnie średni wynik w teście osiągnięć szkolnych zależy od środowiska, w którym pracuje szkoła (opisanego tu przez średni poziom wykształcenia rodziców). Szkoły uzyskujące wysokie wyniki to przede wszystkim szkoły pracujące z uczniami pochodzącymi z rodzin o wyższym poziomie wykształcenia rodziców. Szkoły uzyskujące niższe wyniki pracują z uczniami, których rodzice zdobyli średnio niższy poziom wykształcenia. Oceniając więc pracę szkoły tylko na podstawie wyników testowych, oceniamy ją w dużej mierze za to, na co nie ma ona wpływu. Kontekstowe wskaźniki efektywności nauczania natomiast nie zależą od statusu społeczno-ekonomicznego rodzin uczniów. Każda szkoła ma szansę na wysoką wartość wskaźnika, co wyraźnie widać na wykresie prawym, gdzie punkty na wykresie mają kształt rozproszonej chmury. Jak jednak ten obraz wygląda w przypadku inteligencji uczniów, gdzie nawet dla kontekstowej miary efektywności zaobserwowano pozytywną, choć słabą zależność? Przedstawiamy to na poniższym rysunku. Wykresy pokazują, że szkoły, w których średni poziom inteligencji uczniów jest poniżej średniej, uzyskują raczej przeciętne lub poniżej przeciętnej wyniki w testach osiągnięć. Mają one natomiast szansę na dodatnią wartość kontekstowego wskaźnika efektywności nauczania. Wskaźniki efektywności są bowiem znacznie słabiej powiązane ze średnim poziomem inteligencji uczniów.

Rysunek 4.6. Związek średniego poziomu inteligencji uczniów ze średnimi wynikami szkół (z lewej) oraz z efektywnością nauczania mierzoną wskaźnikami kontekstowymi (z prawej) dla obszaru umiejętności matematycznych



Za inny dowód trafności kontekstowych wskaźników efektywności nauczania mogą posłużyć dane przemawiające za tym, że szkoły określone przez proponowane miary jako efektywne, to placówki, w których obserwujemy korzystne z normatywnego punktu widzenia procesy i zjawiska, i nie dostrzegamy czegoś niepokojącego. Dotychczasowe analizy doprowadziły do kilku interesujących wniosków.

W szkołach czy oddziałach, które zostały określone mianem efektywnych (na podstawie proponowanych wskaźników), obserwujemy wyższy poziom integracji szkolnej uczniów w trzech wymiarach:

- integracji społecznej, związanej z satysfakcją z kontaktów z klasowymi kolegami, oceną możliwości zbudowania satysfakcjonujących więzi społecznych;
- integracji emocjonalnej, wyrażającej pozytywny stosunek uczniów do szkoły;
- integracji motywacyjnej, która odnosi się do obrazu siebie jako ucznia i znajduje wyraz w samoocenie szkolnych kompetencji; wyraża to, czy uczeń czuje, że jest w stanie sprostać wymaganiom stawianym mu w szkole i podejmować zadania z pozytywną motywacją.

Istotne zależności obserwuje się dla wszystkich analizowanych obszarów osiągnięć (umiejętności czytania, świadomości językowej i umiejętności matematycznych). Integracja szkolna opisuje wymiary środowiska szkolnego, które mogą być kształtowane przez nauczycieli i pozostałych pracowników szkoły. Pozytywny związek miar integracji z efektywnością nauczania, mierzoną omawianymi wskaźnikami, pokazuje, że w placówkach o wysokich wartościach wskaźnika kontekstowego uczniowie odczuwają ponadprzeciętną satysfakcję z relacji społecznych, mają lepsze nastawienie do szkoły, a także większą wiarę w swoje możliwości uczenia się.

Analiza czynników związanych z nauczycielem pozwoliła na wykrycie tylko części z zakładanych zależności. Stwierdzono związek wykształcenia nauczycieli z efektywnością nauczania w zakresie umiejętności czytania i świadomości językowej. Uczniowie nauczani przez osoby z wykształceniem wyższym magisterskim uzyskali trochę wyższe wyniki w tych obszarach w stosunku do uczniów nauczycieli z tytułem licencjata, jeśli porównujemy dzieci o takim samym poziomie cech uwzględnionych w kontekstowym modelu efektywności nauczania. Dla rozwoju osiągnięć szkolnych w zakresie świadomości językowej i umiejętności matematycznych znaczenie miało także to, czy nauczyciel potrafi utrzymać dyscyplinę na lekcjach. Nauczyciele, którzy lepiej panowali nad sytuacją w klasie, uzyskiwali lepsze wyniki nauczania. Nie stwierdzono natomiast związku efektywności nauczania ze stylem pracy nauczyciela na lekcjach (co może wynikać z deklaratorywnego charakteru tych danych) ani ze stopniem awansu zawodowego.

Analiza znaczenia kultury organizacyjnej szkoły i stylów zarządzania przyniosła niejednoznaczne rezultaty; w większości przypadków stwierdzono brak istotnych związków. Przyczyną może być to, że ewentualne zależności są zbyt subtelne lub zastosowane w badaniu narzędzia okazały się za mało wrażliwe na to, co w zarządzaniu szkołą ma największe znaczenie dla osiągnięć szkolnych. Wykryto jednak m.in. pozytywny związek współpracy rodziców ze szkołą z efektywnością nauczania, ale tylko w obszarze świadomości językowej.

Analizę trafności proponowanych kontekstowych wskaźników efektywności nauczania na pewno trzeba pogłębić. Czytelników zainteresowanych szczegółowymi wynikami omawianych tu analiz odsyłamy do przytaczanej już publikacji (Dolata i in., 2014), dostępnej także na stronie projektu⁴⁵.

Komunikowanie kontekstowych wskaźników efektywności nauczania i możliwości wykorzystania ich przez szkoły

Szkoły, które uczestniczyły w badaniu, otrzymały zindywidualizowane raporty przedstawiające wyniki i efektywność nauczania w zakresie czytania, świadomości językowej i umiejętności matematycznych. Wyniki były omawiane z przedstawicielami szkół na specjalnej konferencji podsumowującej pierwszą część badania⁴⁶. Poniżej prezentujemy przykładowe omówienie wyników dla jednego z obszarów umiejętności dla wybranej szkoły. Można je potraktować jako propozycję analiz, jakie mogłyby wykonywać szkoły, gdyby kontekstowe modele oceny efektywności nauczania były dostępne dla wszystkich szkół. Prezentowane analizy zostały zaczerpnięte z jednego z raportów dla szkół.

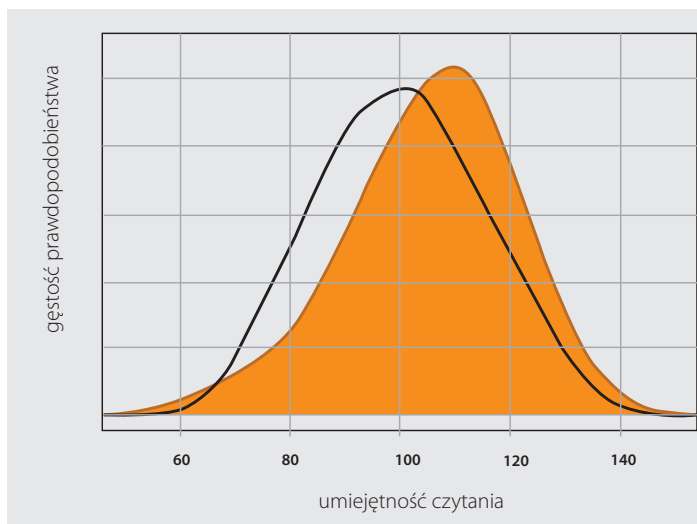
Wyniki uzyskane przez uczniów danej szkoły w każdym z testów osiągnięć zostały zaprezentowane na trzy sposoby. Po pierwsze, zaprezentowany został rozkład wyników w szkole na tle rozkładu wyników w populacji krajowej uczniów (rysunek 4.7). Pomarańczowym kolorem zaznaczono rozkład wyników w teście umiejętności czytania uczniów w szkole. Czarną linią wyrysowano rozkład wyników w populacji krajowej uczniów. Jego średnia wynosi 100 punktów, a odchylenie standardowe 15 punktów. Porównując wzajemne położenie obu tych rozkładów, można ocenić, czy uczniowie szkoły uzyskali lepsze czy słabsze wyniki niż średnio uczniowie w kraju. Jeśli rozkład wyników w szkole jest przesunięty w prawo w stosunku do rozkładu populacyjnego, oznacza to, że uczniowie uzyskali wyniki powyżej przeciętnych. Gdyby rozkład był przesunięty w lewo, oznaczałoby to, że uczniowie uzyskali wyniki poniżej średniej.

Dodatkowo można porównać kształty obu rozkładów. Jeśli rozkład wyników w szkole jest bardziej skupiony wokół średniej niż rozkład populacyjny, oznacza to, że wyniki uczniów w szkole są mniej zróżnicowane niż średnio w populacji. Jeśli jest on silnie rozproszony wokół średniej, oznacza to, że uczniowie w szkole uzyskali wyniki bardziej zróżnicowane niż średnio w populacji. W przypadku prezentowanej szkoły kształt rozkładu jest zbliżony do rozkładu populacyjnego, co oznacza, że zróżnicowanie wyników w szkole jest typowe.

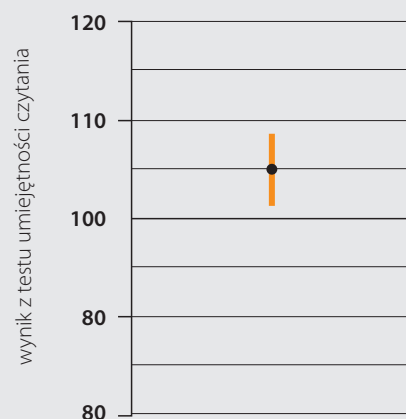
⁴⁵ <http://ewd.edu.pl/publikacje/>

⁴⁶ Informacja o konferencji oraz materiały pokonferencyjne znajdują się na stronie: <http://2013.ewd.edu.pl/szkoły-ewd/sp-2013/>

Rysunek 4.7. Rozkład wyników w teście umiejętności czytania w szkole na tle rozkładu wyników w populacji



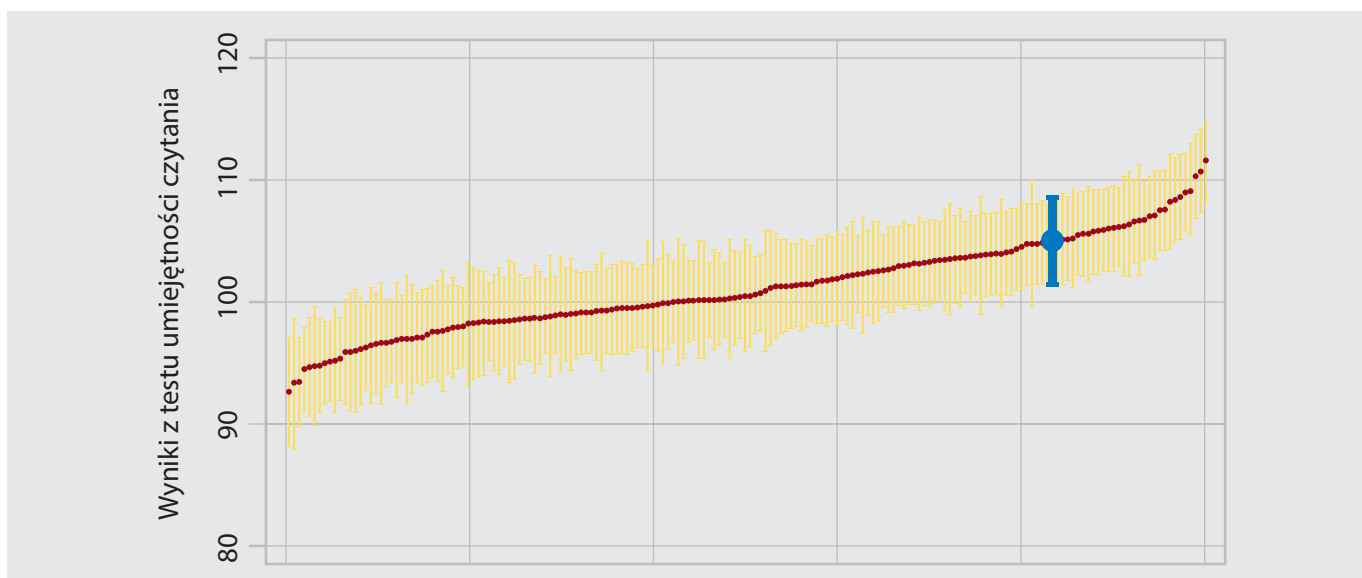
Rysunek 4.8. Średni wynik szkoły w teście umiejętności czytania wraz z przedziałem ufności



Drugi sposób pokazania wyników szkoły to prezentacja średniego wyniku w teście osiągnięć wraz z przedziałem ufności (rysunek 4.8). Pozwala on odpowiedzieć na pytanie, czy średni wynik testu w szkole jest istotnie statystycznie różny od przeciętnej dla kraju. Oś pionowa to skala wyników uzyskanych w teście umiejętności czytania. Średni wynik szkoły jest reprezentowany przez czarny punkt. Przedział ufności (przyjęto poziom ufności 90%) jest zaznaczony niebieskim odcinkiem. Jeśli cały przedział ufności znajduje się ponad wynikiem 100 (tak jak na prezentowanym wykresie), to można z bardzo dużą pewnością powiedzieć, że uczniowie uzyskali w teście wynik wyższy niż średni wynik w populacji. Gdyby cały przedział ufności znajdował się poniżej wyniku 100, można by poprawnie statystycznie lokować średni wynik szkoły poniżej średniej krajowej. Gdyby przedział ufności przecinał poziom 100 punktów, oznaczałoby to, że uczniowie uzyskali wynik porównywalny ze średnim wynikiem w populacji uczniów.

Średni wynik szkoły przedstawiono także na tle wyników wszystkich szkół uczestniczących w badaniu (wykres poniżej). Wyniki szkół zostały uporządkowane rosnąco według wartości średniej. Informacja ta miała pomóc szkołom lepiej zinterpretować położenie ich wskaźnika. Dzięki niej można zobaczyć, ile szkół ma istotnie różne wyniki od danej szkoły.

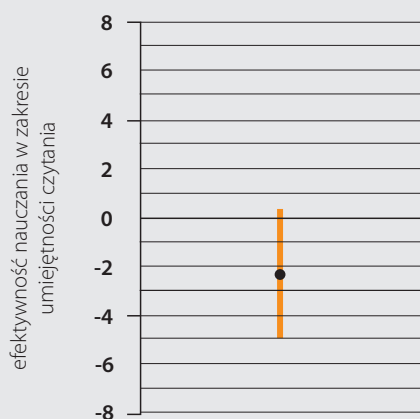
Rysunek 4.9. Średnie wyniki szkoły w teście umiejętności czytania na tle wyników wszystkich badanych szkół



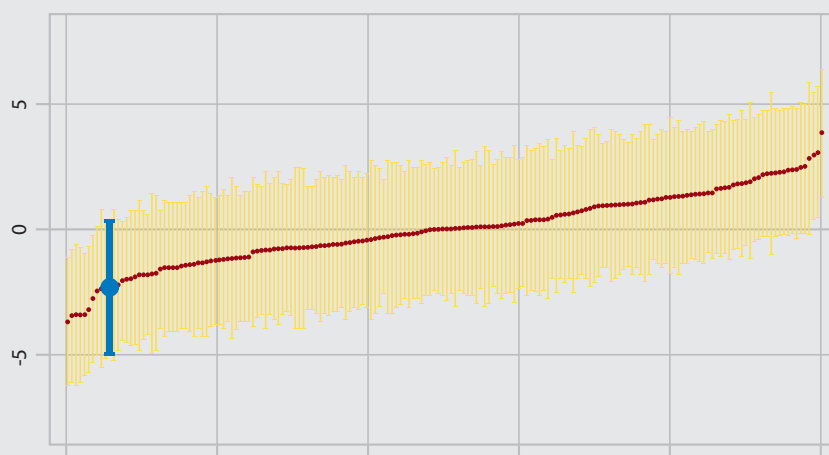
Kontekstowe wskaźniki efektywności nauczania zostały zaprezentowane graficznie wraz z przedziałem ufności. Czarnym punktem oznaczono średnią wartość wskaźnika, a pomarańczowym odcinkiem 90-procentowy przedział ufności. Oś pionowa to skala efektywności nauczania. Im wyższy na niej wynik, tym wyższa efektywność nauczania. Wynik 0 oznacza średnią efektywność nauczania w populacji. Jeśli cały przedział ufności wskaźnika znajduje się powyżej punktu 0, oznacza to, że jest wysoce prawdopodobne, że szkoła naucza z powyżej przeciętną efektywnością. Uzyskuje więc średnio wyniki nauczania wyższe niż inne szkoły pracujące z podobnymi uczniami. Jeśli przedział ufności znajduje się poniżej punktu 0, oznacza to, że szkoła z dużym prawdopodobieństwem naucza z poniżej przeciętną efektywnością, a inne szkoły pracujące z podobnymi uczniami uzyskały średnio lepsze rezultaty. Jeśli przedział ufności przecina wartość 0, oznacza to, że szkoła pracuje z przeciętną efektywnością, czyli taką jak większość szkół w kraju.

Informacja o efektywności nauczania także została przedstawiona na tle wskaźników innych szkół biorących udział w badaniu. Efektywność nauczania w zakresie umiejętności czytania w prezentowanej szkole należałoby ocenić jako przeciętną (przedział ufności zawiera zero), ale fakt, że wskaźnik „ciąży” w stronę wyników ujemnych, powinien skłonić szkołę do szukania przyczyn takiej sytuacji.

Rysunek 4.10. Efektywność nauczania w zakresie umiejętności czytania

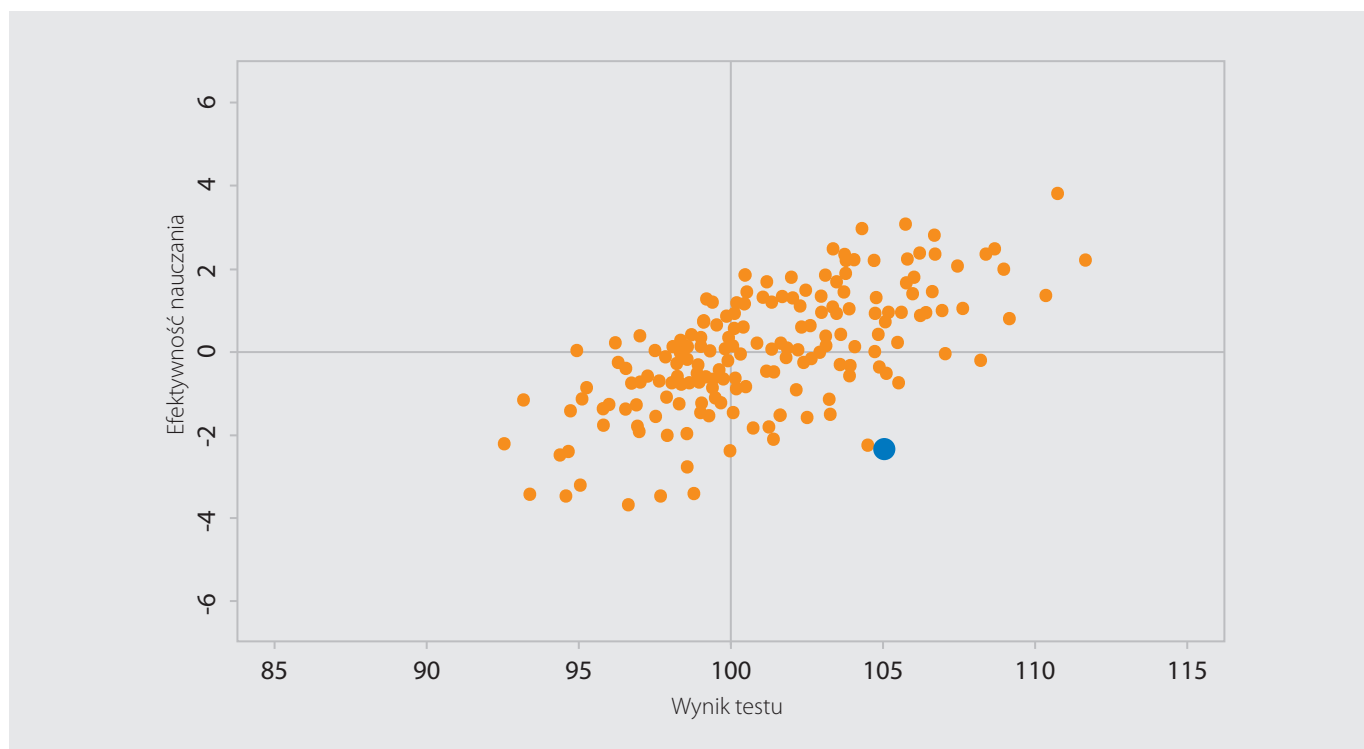


Rysunek 4.11. Efektywność nauczania w zakresie umiejętności czytania na tle wszystkich badanych szkół



Szkołom przedstawiono także łączną informację o średnim wyniku w teście i efektywności nauczania na tle wyników wszystkich szkół biorących udział w badaniu. Dzięki takiemu zestawieniu informacji można zobaczyć, gdzie ze względu na wartości obu wskaźników łącznie znajduje się dana szkoła w porównaniu do innych szkół. Na osi poziomej przedstawiono wyniki uzyskane w teście umiejętności czytania. Na osi pionowej przedstawiono wskaźnik efektywności nauczania w tym obszarze. Na tym wykresie wyniki przedstawiono bez przedziałów ufności, by zwiększyć jego czytelność. Trzeba jednak pamiętać, że są one obarczone niepewnością statystyczną, więc nie należy przeceniać małych różnic.

Rysunek 4.12. Łączna informacja o wyniku w teście osiągnięć i efektywności nauczania w zakresie umiejętności czytania



Szkoły otrzymały informację zarówno o średnich wynikach w testach, jak i efektywności nauczania. Obie te informacje wzajemnie się dopełniają. Wyniki testów są oczywiście w sposób naturalny powiązane z wartościami kontekstowych wskaźników efektywności nauczania, bowiem wyższa lub niższa efektywność nauczania przekłada się na wyższe lub niższe wyniki⁴⁷. Niemniej możemy wyłonić takie szkoły, które osiągają podobne wyniki, ale pracują z różną efektywnością, jak i odwrotnie.

Ograniczenia i rozwój modeli kontekstowych

Przeprowadzone badania pokazały, że kontekstowe wskaźniki efektywności nauczania mogłyby być wartościowym narzędziem wykorzystywanym przez szkoły do ewaluacji własnej pracy na pierwszym etapie edukacyjnym. Jakie warunki musiałyby być spełnione, by można je było powszechnie stosować?

Po pierwsze, niezbędny byłby dobry pomiar osiągnięć szkolnych po trzeciej klasie szkoły podstawowej. Nie jest jednak konieczne, by zostały nim objęte wszyscy uczniowie w kraju. Wystarczające byłoby badanie na reprezentatywnej próbie szkół, którego wyniki posłużyłyby do wyliczenia modeli i ustalenia punktu odniesienia do porównań dla innych szkół. Pozostałe placówki, zainteresowane takimi wskaźnikami, mogłyby dobrowolnie przeprowadzić u siebie takie pomiary i porównać ich rezultaty z wynikami badania reprezentatywnego. W podobnej formule było realizowane badanie OBUT (*Ogólnopolskie Badanie Umiejętności Trzecioklasistów*), co pokazuje, że taki pomiar diagnostyczny jest możliwy i mógłby być podstawą do wyliczania takich wskaźników. Oczywiście, konieczna byłaby krytyczna analiza testów, które byłyby wykorzystane do wyliczenia wskaźników, bowiem od ich trafności i rzetelności zależałaby jakość szacowanych na ich podstawie miar⁴⁸. Poza pomiarem osiągnięć potrzebne byłyby dane o uczniach, które musiałyby być zbierane i przechowywane przez szkoły, a następnie łączone z wynikami testów osiągnięć. Dla proponowanych wskaźników należałoby zebrać informację o wieku, płci dziecka, wykształceniu jego rodziców oraz przeprowadzić

⁴⁷ Korelacja pomiędzy kontekstowymi wskaźnikami a wynikami testów wynosi około 0,64.

⁴⁸ Testy OBUT nie były konstruowane na potrzeby uzyskania całościowego obrazu umiejętności uczniów. W kolejnych edycjach dobierano różne, punktowe obszary opisane w podstawie programowej, tak by rozwiązania poszczególnych zadań były dobrą informacją zwrotną dla szkół i nauczycieli uczestniczących w badaniu.

krótką ankietę wśród rodziców w celu pozyskania informacji o aspiracjach edukacyjnych względem dzieci i liczbie książek posiadanych w domu. W przypadku pytania o aspiracje edukacyjne ważne jest, by było ono skierowane do rodziców na początku edukacji ich dzieci w klasie pierwszej. Aspiracje rodziców mogą się bowiem zmieniać pod wpływem osiągnięć szkolnych dzieci oraz działań szkoły. Dlatego, jeśli chcemy w modelu uwzględnić czynniki niezależne od pracy szkoły, ważne jest, by pomiar aspiracji przeprowadzać już w pierwszej klasie. To jednak rodzi dodatkową komplikację w postaci konieczności zbierania danych kontekstowych dla badanego rocznika już na początku nauki oraz przechowywania ich i ochrony przez szkoły przez dwa kolejne lata. Oczywiście nie jest to nierealne, a ankieta kierowana do rodziców na początku drogi edukacyjnej ich dzieci mogłaby stanowić okazję dla szkół do poznawania środowiska rodzinnego uczniów. Wyniki analiz pokazały, że można by również rozważyć zbudowanie modeli prostszych, wymagających pozyskania tylko informacji o wieku, płci i wykształceniu rodziców. W takiej sytuacji wystarczyłoby zebranie tych danych w okolicach pomiaru osiągnięć. Wskaźniki wyliczone z takich modeli miały tylko trochę gorsze właściwości (szczegóły zostały omówione w przytaczanej już publikacji: Dolata i in., 2014). I na koniec kwestia odpowiednich narzędzi wspomagających analizy z wykorzystaniem wskaźników kontekstowych. Po przeprowadzeniu pomiaru osiągnięć szkolnych na koniec pierwszego etapu edukacyjnego zebrane przez szkołę dane mogłyby być wprowadzane do specjalnie przygotowanej aplikacji (np. podobnej do funkcjonujących już Kalkulatora EWD 100 lub Kalkulatora EWD SP, omówionych w kolejnych częściach rozdziału), w której uprzednio zaimplementowano by modele efektywności nauczania, wyliczone na podstawie danych z badania na reprezentatywnej próbie. W ten sposób osoby zainteresowane mogłyby analizować efektywność nauczania w danej szkole czy w podgrupach uczniów w szkole, bez konieczności wysyłania danych do zarządzającej tym procesem centrali.

4.3. Wykorzystanie metody edukacyjnej wartości dodanej po drugim etapie edukacyjnym

Wskaźniki edukacyjnej wartości dodanej dla drugiego etapu edukacyjnego (klasy IV–VI szkoły podstawowej) są ważnym uzupełnieniem tego typu miar dostępnych już od kilku lat dla szkół gimnazjalnych oraz ponadgimnazjalnych. Dotychczasowe modele EWD wykorzystywały wyniki z systemu egzaminacyjnego. Miary EWD opracowane dla drugiego etapu edukacyjnego wykorzystują jako miarę uprzednich osiągnięć wyniki pomiaru diagnostycznego realizowanego w klasach III szkoły podstawowej w ramach *Ogólnopolskiego Badania Umiejętności Trzecioklasistów* (OBUT). W realizowanym od 2011 roku badaniu bierze udział większość szkół podstawowych w Polsce. Ze względu na diagnostyczny charakter OBUT metoda EWD opracowana dla II etapu edukacyjnego ma wspomagać ewaluację wewnątrzszkolną i nie powinna być używana w ewaluacji zewnętrznej.

Model edukacyjnej wartości dodanej dla drugiego etapu edukacyjnego

Wyniki sprawdzianu w klasie VI nie są wystarczającą przesłanką do oceny efektywności nauczania w szkole podstawowej (Dolata i in., 2013). Aby ocenić nauczanie w danej placówce na podstawie wyników sprawdzianu, należy porównywać wyniki jej uczniów z rezultatami podobnych uczniów w grupie odniesienia. Dzięki danym z OBUT możliwe jest zastosowanie w tym celu metody EWD. Należy jednak pamiętać, że wskaźniki wyliczone na bazie wyników OBUT i sprawdzianu w klasie VI dotyczą tylko efektywności nauczania w klasach IV–VI i grupą odniesienia jest populacja uczniów uczęszczających do szkół, które brały udział w danej edycji OBUT. Wskaźniki EWD dla drugiego etapu edukacji są miarą relatywną, informującą o tym, czy dana szkoła wypracowała z uczniami więcej, mniej czy tyle samo, co inne, statystycznie podobne szkoły.

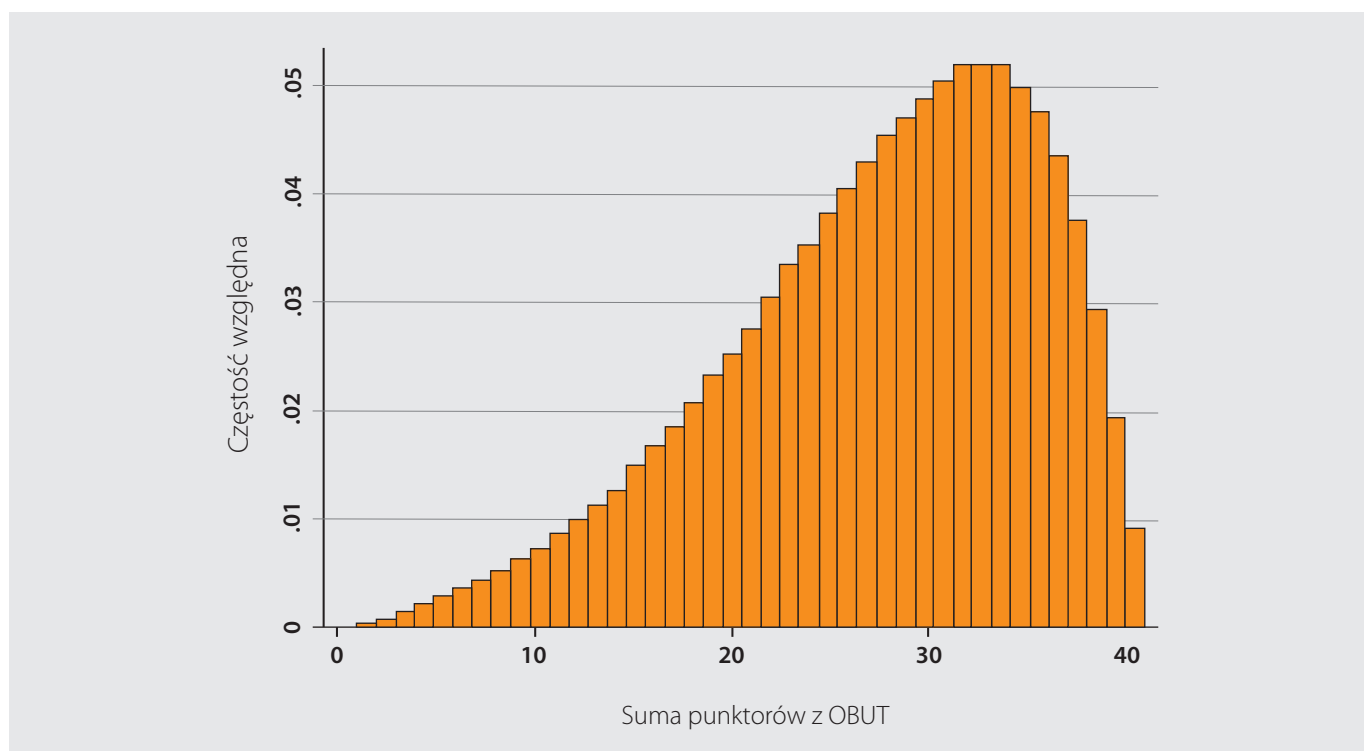
Aby można było wyznaczyć EWD dla szkoły, najpierw należy określić prognozę statystyczną mówiącą, jaki jest najbardziej prawdopodobny wynik ucznia na sprawdzianie w zależności od jego wcześniejszych osiągnięć. Taką prognozę nazywamy modelem EWD. Do obliczenia takiego modelu

4. Metoda edukacyjnej wartości dodanej w Polsce

potrzebne są dwa pomiary osiągnięć szkolnych: na „wejściu” i na „wyjściu” ze szkoły. W dalszej części podrozdziału będziemy zajmować się modelem EWD dla II etapu edukacyjnego, w którym pomiarem na wejściu jest wynik z OBUT 2011, a pomiarem na wyjściu przeprowadzony trzy lata później sprawdzian w klasie VI, czyli w 2014 r.

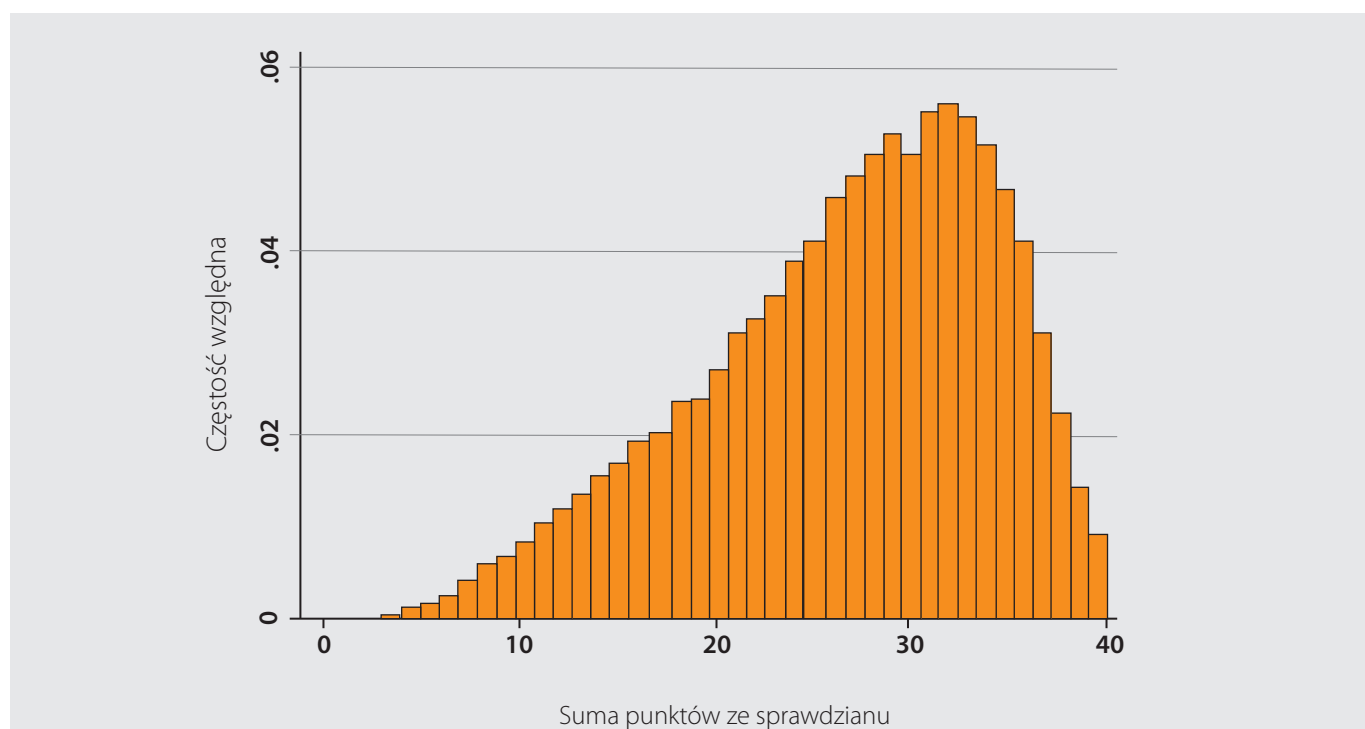
W 2011 roku w badaniu OBUT wzięło udział 9 756 placówek, tj. ponad 75% wszystkich szkół podstawowych. OBUT składał się z dwóch części: polonistycznej i matematycznej. W części z języka polskiego było 11 zadań, za które można było uzyskać łącznie 24 punkty. Część matematyczna składała się z 9 zadań, za ich rozwiązanie maksymalnie można było otrzymać 17 punktów. Łącznie za cały test można było uzyskać 41 punktów. Do głównej bazy badania OBUT przesłano łącznie 273 898 wyników testowania. Średnio badani uczniowie uzyskali 27,8 punktu (przy odchyleniu standardowym 8,1). Rozkład wyników był silnie lewowski, co świadczy o tym, że test był relatywnie łatwy (por. rysunek 4.13).

Rysunek 4.13. Rozkład wyników badania OBUT 2011 w populacji badanych



Pomiarem końcowym osiągnięć szkolnych, a analizowanym modelem EWD jest sprawdzian w klasie VI. W 2014 roku standardowy arkusz egzaminacyjny zawierał 26 zadań, w tym 20 zadań zamkniętych wielokrotnego wyboru oraz 6 otwartych. Arkusz ten rozwiązywali uczniowie bez dysfunkcji i z dysleksją rozwojową. Za poprawne wykonanie wszystkich zadań można było maksymalnie otrzymać 40 punktów. Średni wynik w Polsce na skali wyników surowych to 25,8 punktu (przy odchyleniu standardowym 8,0). Podobnie jak w przypadku OBUT, sprawdzian był relatywnie łatwy, o czym świadczy lewoskośny rozkład wyników (por. rysunek 4.14).

Rysunek 4.14. Rozkład wyników sprawdzianu w 2014 roku w populacji uczniów VI klas szkół podstawowych

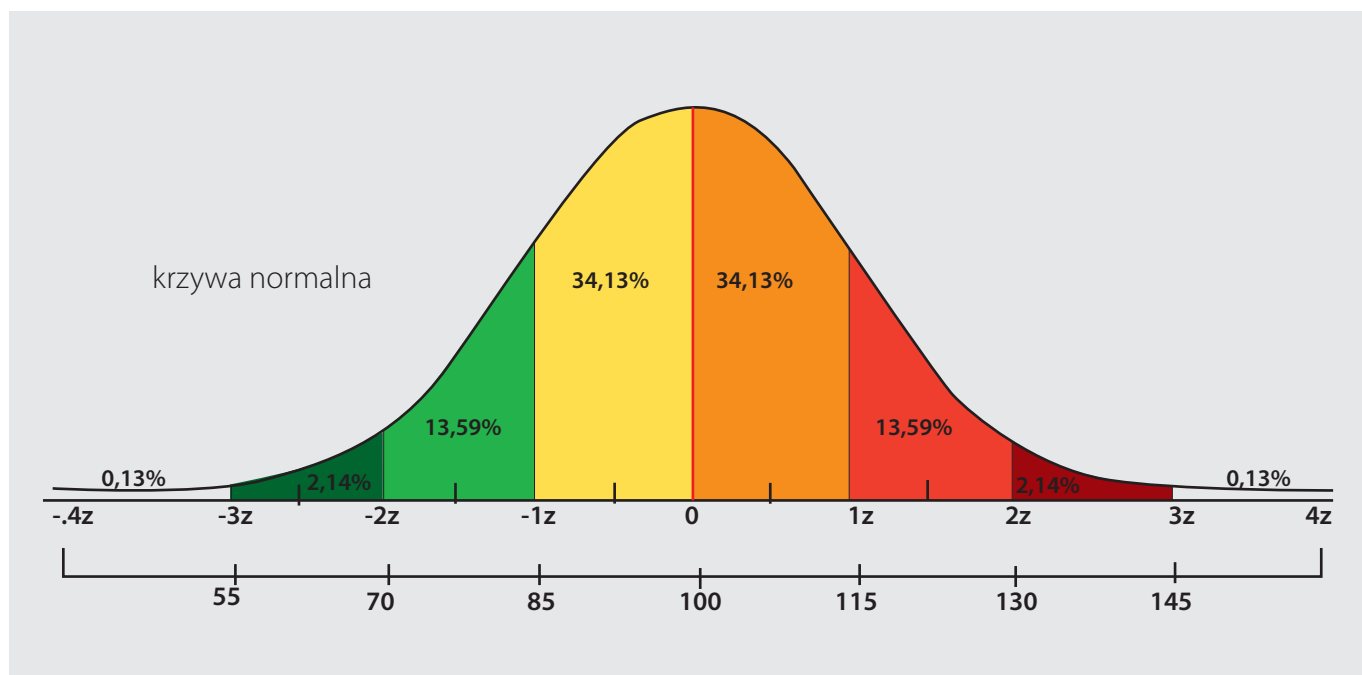


Aby statystyczne modelowanie EWD dało jak najlepsze rezultaty, a wyniki miały ustaloną interpretację ilościową, konieczne jest odpowiednie przygotowanie danych z testowania. Wyniki surowe zostały przekształcone tak, by rozkład był maksymalnie zbliżony do normalnego, i następnie przełożone na skalę standardową o średniej 100 i odchyleniu standardowym równym 15.

Jak widzieliśmy, rozkłady wyników testu OBUT 2011 i sprawdzianu 2014 bardzo odbiegają od normalnego. Konieczne jest zatem takie wyskalowanie obu testów, by maksymalnie zbliżyć rozkład do założonego. W tym celu wykorzystano jedną z nowoczesnych metod skalowania testów – model Rascha⁴⁹. Dzięki temu można w prosty sposób przełożyć wynik surowy testów na skalę dającą rozkład bardziej zbliżony do normalnego. Zastosowanie nowoczesnej metody skalowania pozwoliło równocześnie sprawdzić pomiarową jakość testów użytych w OBUT 2011 i sprawdzianie 2014.

⁴⁹ Więcej informacji na temat tej metody skalowania testów zainteresowany Czytelnik może znaleźć w pracy Aleksandry Jasińskiej i Michała Modzelewskiego (2014) oraz w książce pod redakcją Artura Pokropka (2015).

Rysunek 4.15. Krzywa normalna i skala standardowa 100; 15



Jak przedstawia to rysunek 4.15. przekształcenie wyników do rozkładu normalnego i wyrażenie na skali o średniej 100 i odchyleniu standardowym 15 pozwala w łatwy sposób odnieść wartość na skali do odsetka osób posiadających wskazany poziom umiejętności. Dla przykładu, wiemy, że jeśli uczeń uzyskał wynik równy 100, to połowa uczniów w grupie odniesienia uzyskała wynik niższy, a druga połowa wyższy. Jeśli uczeń uzyskał wynik 115, to gorzej test napisało 84,13% rówieśników, a lepiej 15,87%. Taki sposób komunikowania wyników sprawia, iż nie musimy pamiętać, jaka była w danym roku średnia liczba punktów zdobytych na teście OBUT czy na sprawdzianie, co roku będzie to 100. Odpowiedzi wszystkich uczniów biorących udział w badaniu OBUT w 2011 roku posłużyły do wyskalowania wyników z użyciem modelu Rascha. Analiza dopasowania danych do modelu pozwala przyjąć, że uzyskane w efekcie skalowania wyniki wiarygodnie odzwierciedlają poziom umiejętności uczniów. Każdemu uzyskanemu w OBUT wynikowi sumarycznemu, można przyporządkować wynik na skali o średniej 100 i odchyleniu 15 (por. rysunek 4.16).

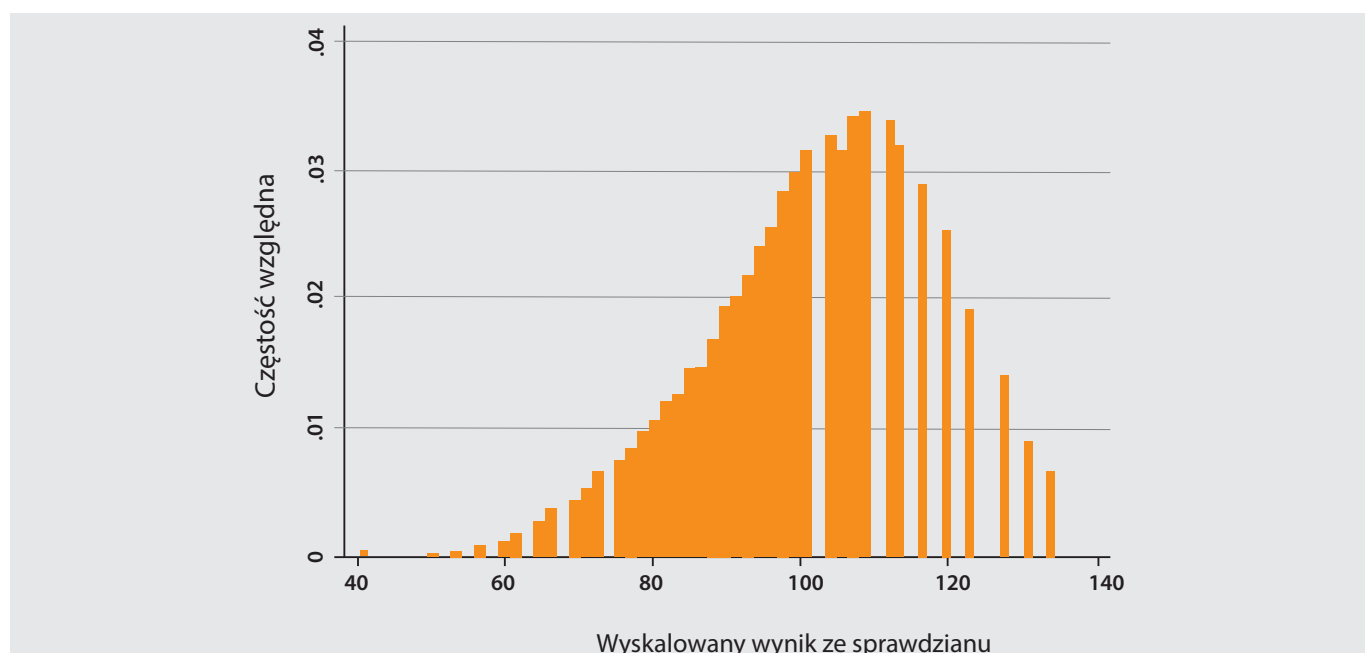
Rysunek 4.16. Przeskalowane wyniki testu OBUT wyrażone na skali 100; 15



Niestety silny efekt pułapu sprawił, że na skali 100;15 widzimy w strefie wyników wysokich dużą „niedociągłość” skali.

Podobnie jak miało to miejsce z wynikami testu OBUT, również wyniki sprawdzianu dla całej populacji uczniów rozwiązujących arkusz standardowy wyskalowany został z wykorzystaniem modelu Rasha. Dopasowanie danych do modelu okazało się dobre. Podobnie jak miało to miejsce w przypadku testów OBUT, oszacowania umiejętności uczniów przekształcono tak, aby w grupie odniesienia (wśród uczniów piszących arkusz standardowy sprawdzianu w 2014 roku) średnia równa była 100, a odchylenie standardowe 15. Przekształcone wyniki sprawdzianu na standardowej skali 100; 15 przedstawia rysunek 4.17.

Rysunek 4.17. Przeskalowane wyniki sprawdzianu wyrażone na skali 100; 15



Podobnie jak w wypadku testu OBUT, w wypadku wyskalowanych wyników sprawdzianu 2014 w strefie wyników wysokich widzimy pewną nieciągłość, ale luki są na szczęście mniejsze.

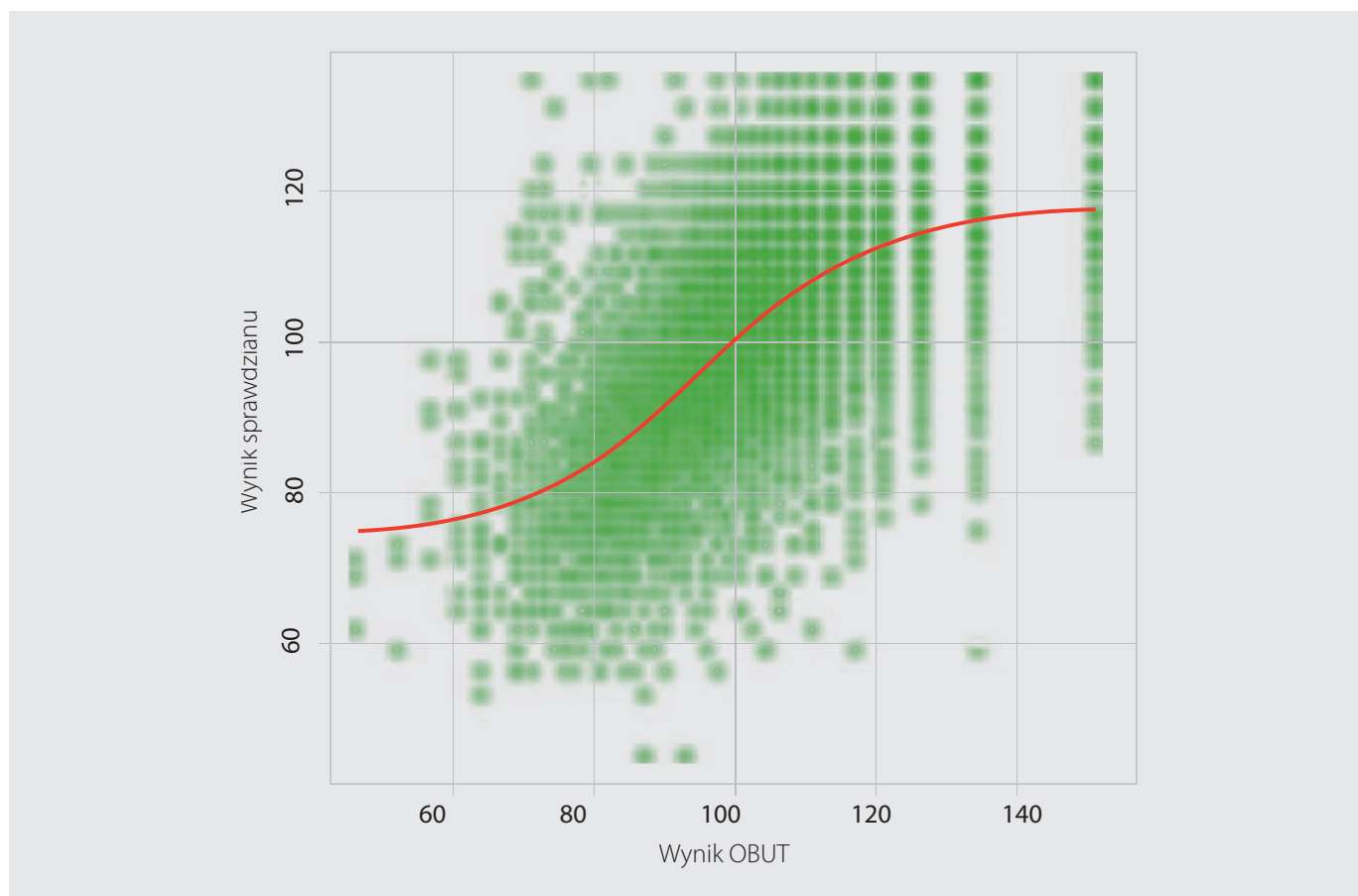
Na potrzeby wyliczenia modelu EWD potrzebne są połączone dla całej populacji uczniów (precyzyjnie mówiąc – grupy odniesienia) wyniki testu OBUT 2011 i sprawdzianu 2014. Ze względu na organizację badania OBUT i konieczność anonimizacji krajowej bazy danych z tego badania nie było jednak możliwości połączenia wyników wszystkich uczniów piszących OBUT z ich wynikami na sprawdzianie. Z tego względu szacowanie modelu EWD przeprowadzono na danych z ogólnopolskiej, reprezentatywnej dla populacji szkół biorących udział w OBUT 2011 próby 200 szkół podstawowych. Dyrektorzy tych placówek, jako osoby posiadające dostęp do nieanonimizowanych wyników uczniów dla obu testów, poproszeni zostali o połączenie tych wyników i przesłanie ich (po anonimizacji) w formie elektronicznej do IBE. Poza samym wynikiem dyrektorów poproszono o przekazanie informacji o płci ucznia oraz zaznaczenie, czy pisząc sprawdzian, miał on opinię o dysleksji rozwojowej. W wyznaczonym terminie udało się zebrać dane ze 181 szkół.

Dla oceny reprezentatywności próby zebrane dane porównano z charakterystykami populacyjnymi. Średnie wyniki uczniów w badanych szkołach okazały się nieco lepsze niż średnie w populacji. W przypadku OBUT uczniowie, których wyniki zostały przesłane przez szkoły, uzyskali średnio 101,1 punktu (na skali o średniej w populacji 100 i odchyleniu standardowym 15). Zróżnicowanie wyników, wyrażone odchyleniem standardowym wyniosło 15,4. Dla sprawdzianu średni wynik w próbie wyniósł 101,4 punktu (na skali 100; 15), a odchylenie standardowe 15,1 (por. tabela 4.2). Można zatem przyjąć, że próba dobrze odzwierciedla wyniki populacyjne.

Tabela 4.2. Porównanie wyników testu OBUT i sprawdzianu pomiędzy populacją a próbą wylosowaną do przygotowania modelu EWD.

Pomiar	Grupa	Średnia	Odchylenie st.
OBUT 2011	populacja	100,0	15,0
	próba	101,1	15,4
Sprawdzian 2014	populacja	100,0	15,0
	próba	101,4	15,1

Na danych z próby przygotowany został model EWD dla II etapu edukacyjnego, czyli określono, jaki jest najbardziej prawdopodobny wynik ucznia na sprawdzianie w zależności od tego, jaki wynik uczeń uzyskał na teście OBUT. Przebieg tej zależności można przedstawić graficznie za pomocą linii przewidywanego wyniku, jak to widać na poniższym wykresie.

Rysunek 4.18. Linia przewidywanego wyniku uczniów na sprawdzianie 2014 w zależności od wyniku na teście OBUT 2011

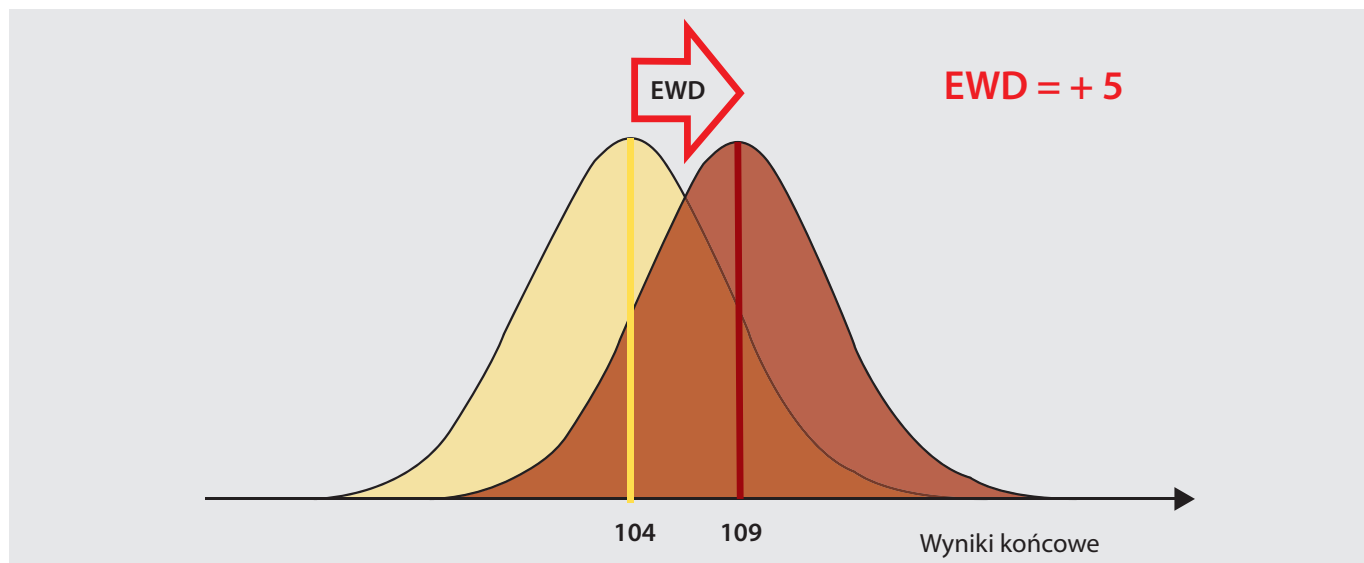
Znajdujące się na nim punkty reprezentują rzeczywiste wyniki uczniów na obu pomiarach: na osi X wynik w teście OBUT, na osi Y wynik na sprawdzianie. Punkty powyżej linii przewidywanego wyniku reprezentują uczniów, którzy na sprawdzianie uzyskali wyniki wyższe, niż można by się tego spodziewać na podstawie wyników z OBUT. I na odwrót, poniżej znajdują się wyniki niższe niż przewidywane przez model. Różnicę między wynikiem faktycznie uzyskanym przez ucznia a wynikiem przewidywanym z modelu EWD nazywamy resztą. Reszta to ważny termin, często wykorzystywany w analizach EWD. Należy zwrócić uwagę na zjawisko zwane regresją ku średniej. Im wynik w pierwszym pomiarze bardziej odbiega od średniej, im minus lub in plus, tym w drugim pomiarze przewidywany statystycznie wynik bardziej ulegał „ściągnięciu” do średniej. Na przykład, dla wyniku na teście OBUT równym 100, model regresji przewiduje na sprawdzianie trzy lata później też 100 punktów. Ale jeżeli uczeń w klasie III uzyskał 140 punktów, to w klasie VI model nie przewiduje rezultatu 140, ale znacznie bliższy średniej (z wykresu można odczytać, że ok 118). Analogiczne zjawisko obserwujemy oczywiście w wypadku bardzo niskich wyników na „wejściu”.

Model EWD jest obliczany ex post i co roku może wyglądać trochę inaczej (trochę inny przebieg linii przewidywanego wyniku). Proces obliczania wskaźników EWD dla szkoły lub jakiejś innej grupy uczniów można w pewnym uproszczeniu (uproszczenie polega na nieuwzględnieniu zmiennych kontrolnych, czyli informacji o płci i dysleksji ucznia) przedstawić następująco:

- (1) na podstawie wyników testu OBUT oraz modelu EWD dla każdego ucznia w szkole/grupie wyznaczamy przewidywany wynik sprawdzianu;
- (2) liczymy dla szkoły/grupy średnią arytmetyczną wyników przewidywanych;
- (3) liczymy dla szkoły/grupy średnią arytmetyczną wyników faktycznie uzyskanych na sprawdzianie;
- (4) w ostatnim kroku odejmujemy od średniej wyników faktycznie uzyskanych średnią wyników przewidywanych; różnica jest wartością wskaźnika EWD.

Proces ten przedstawia rysunek na następnej stronie.

Rysunek 4.19. Ilustracja sposobu obliczania wskaźników EWD dla grupy uczniów. Po lewej rozkład wyników przewidywanych na podstawie modelu EWD, po prawej rozkład wyników faktycznie uzyskanych



Powyższy rysunek przedstawia sytuację, w której wskaźnik EWD jest dodatni, czyli efektywność nauczania w danej szkole/grupie uczniów jest powyżej przeciętnej efektywności w grupie odniesienia. Wartość liczbową +5 oznacza, że w stosunku do rozkładu przewidywanych wyników udało się dzięki efektywnemu nauczaniu uzyskać średni wynik na sprawdzianie o 5 punktów wyższy (na skali indywidualnych wyników o odchyleniu standardowym 15 punktów, czyli przesunięcie jest znaczące). Powyższą procedurę uzupełnia wyznaczenie dla obliczonej wartości przedziału ufności, czyli uwzględnienie statystycznej niepewności oszacowania wskaźnika EWD. Tak jak inne wskaźniki EWD, tak też miary EWD dla II etapu edukacyjnego prezentowane są zawsze z przedziałem ufności. Przypomnijmy, że szerokość przedziału ufności zależy od dwóch zasadniczych czynników: (1) liczby uczniów, na podstawie wyników których obliczamy wartość wskaźnika. Wraz ze wzrostem liczby uczniów wzrasta nasza pewność, co do tego jaka jest rzeczywista wartość wskaźnika, a więc zmniejsza się szerokość przedziału ufności; (2) zróżnicowania reszt wyliczanych na podstawie modelu EWD. Mniejsze zróżnicowanie reszt przekłada się na większą pewność, co do tego, jaka jest efektywność nauczania. W efekcie przedział ufności jest węższy.

Z zaznaczeniem przedziału ufności prezentowane są wskaźniki EWD dla szkół podstawowych w specjalnie przygotowanym do tego programie Kalkulator EWD SP.

Trafność wskaźników EWD dla II etapu edukacyjnego

Ze względu na fakt, że wskaźniki EWD dla II etapu edukacyjnego w 2014 roku zostały obliczone po raz pierwszy, mamy niewiele informacji, które mogłyby służyć jako źródło wiedzy o tym, na ile są one trafne. Porównywać możemy jedynie ich własności z charakterystykami wskaźników obliczanych od dłuższego czasu. W efekcie takich porównań odkryto, że zróżnicowanie wyników testu OBUT, które można przypisać podziałowi uczniów na szkoły (wariancja międzyszkolna), jest większe, niż można by się spodziewać, biorąc pod uwagę wyniki innych badań umiejętności prowadzonych wśród dzieci z klas trzecich (por. tabela 4.3.). Gdyby wszystkie szkoły uczyły tak samo (i dawały takie same warunki do rozwiązania testu) różnice w wynikach uczniów byłyby efektem ich indywidualnych cech. W takiej sytuacji nie byłoby wariancji międzyszkolnej (równa byłaby 0%). Jeśli wyniki w każdej szkole byłyby inne, lecz nie różniłyby się wewnątrz szkoły, to wskaźnik ten byłby równy 100%. W przypadku badania OBUT, biorąc pod uwagę wyniki wszystkich piszących, wariancja międzyszkolna równa jest niemal 18%, jeszcze większa jest wariancja w próbie: niemal 22%. Wartości te są znacznie wyższe niż uzyskane w innych badaniach, w których wariancja międzyszkolna osiągnięć szkolnych na koniec I etapu edukacyjnego wahała się w przedziale od 8,6% do 13,7% (Dolata i in., 2014; Jasińska,

Modzelewski, 2013). Może to przemawiać za hipotezą, że przynajmniej w części szkół uzyskane wyniki nie są precyzyjnym obrazem umiejętności uczniów.

Tabela 4.3. Wariancja międzyszkolna w badaniach umiejętności uczniów na koniec I etapu edukacyjnego

Dane		Odsetek wariancji międzyszkolnej
OBUT 2011	populacja	17,7%
	próba	21,9%
Badanie podłużne EWD (2012), TOS3	czytanie	8,6% ^a
	świadomość językowa	11,8% ^a
	matematyka	10,5% ^a
PIRLS 2011	czytanie	12,2% ^b
TIMSS 2011	matematyka	13,7% ^b

^aDolata i in., 2014;

^bJasińska i Modzelewski, 2013.

Biorąc pod uwagę fakt, że podwyższona wariancja międzyszkolna wyników OBUT może wynikać z uchybień w zakresie standaryzacji testowania w wykorzystaniu wskaźników EWD SP w ewaluacji wewnętrznej, należy zwrócić uwagę, iż niskie wartości EWD przy wysokich wartościach OBUT mogą wymagać pogłębionej refleksji nie tylko nad efektywnością nauczania w klasach IV – VI, ale również nad sposobem przeprowadzania i oceniania testu OBUT.

Możliwości wykorzystania metody EWD przez szkoły podstawowe

Analizy EWD dla II etapu edukacyjnego są możliwe dzięki specjalnie w tym celu przygotowanej aplikacji komputerowej Kalkulator EWD SP. Można ją bezpłatnie pobrać ze strony projektu www.ewd.edu.pl/pobierz-2/.

Aby przeprowadzić analizę dla szkoły należy, w odpowiedniej formie przygotować i wczytać do programu dane. Przykład kilku wierszy danych do analizy przedstawia tabela 4.4. Użytkownik może dodać swoje własne kolumny, np. zawierające informację o dojazdach ucznia, o tym, który nauczyciel uczył go poszczególnych przedmiotów itp. Informacje dodatkowe mogą okazać się kluczowe dla wykorzystania analiz EWD w procesie ewaluacji wewnętrznej.

Tabela 4.4. Przykład zestawienia informacji do wprowadzenia do Kalkulatora EWD SP

kod_ucznia	SPR	OBUT_pl	OBUT_mat	Płeć	Dysleksja
A1	22	20	11	M	N
A2	13	17	9	K	N
A3	35	24	16	K	N
A4	22	10	15	M	N
A5	20	20	9	K	N
A6	10	16	9	K	N
A7	21	22	14	K	N

Przejście pierwszych kroków w obsłudze Kalkulatora EWD SP i w interpretacji wyników ułatwiają krótkie filmiki instruktażowe znajdujące się na stronie <http://ewd.edu.pl/lista-filmow/> oraz instrukcje „Szybki start” ze strony <http://ewd.edu.pl/szybki-start-2/> i „Analizy” (<http://ewd.edu.pl/analizy-2/>). W materiałach tych przedstawiony został sposób importu danych z plików oraz wskazówki jak samodzielnie przeprowadzić pierwsze analizy.

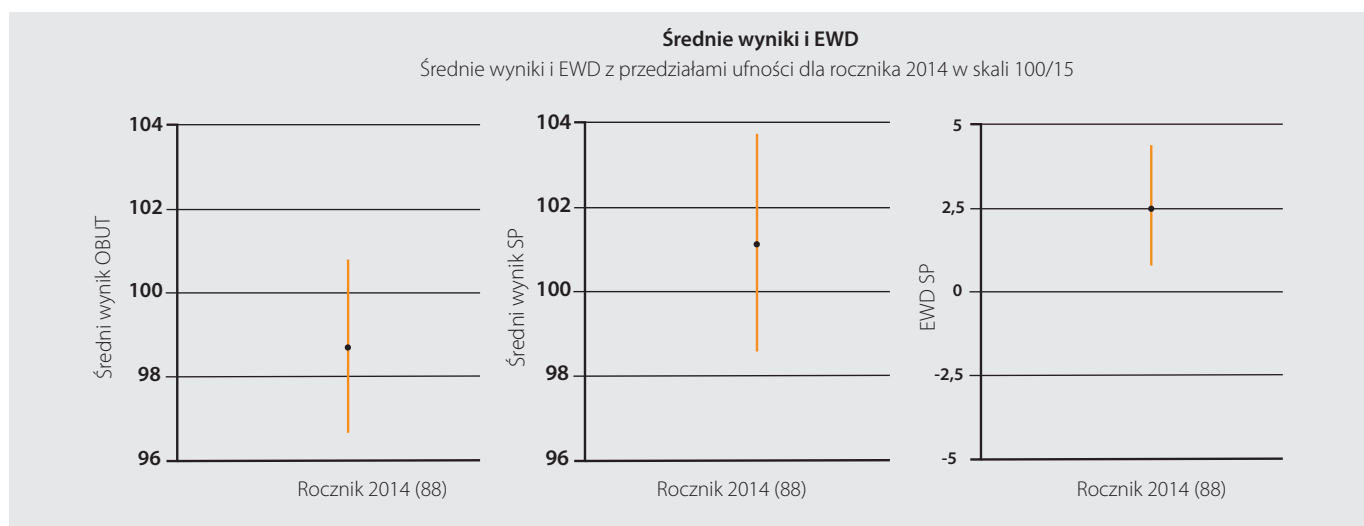
Analiza danych za pomocą kalkulatora służyć powinna przede wszystkim lepszemu zrozumieniu procesów zachodzących w szkole, zwłaszcza w zakresie nauczania przedmiotów objętych sprawdzianem. Korzystanie z Kalkulatora ułatwia wypełnianie wymagań, jakie państwo stawia wobec szkół, zwłaszcza wymagania 11. „Szkoła lub placówka, organizując procesy edukacyjne, uwzględnia wnioski z analizy wyników sprawdzianu (...) oraz innych badań zewnętrznych i wewnętrznych.” Badanie wskaźników EWD jest zatem istotnym elementem procesu autoewaluacji szkoły.

Pracę z Kalkulatorem podzielić można na trzy zasadnicze części. Pierwsza to zapoznanie się z ogólnymi wynikami dla szkoły. Drugi to poszukiwanie takich grup uczniów, dla których obserwujemy istotne różnice w zakresie EWD. Trzeci to stawianie i weryfikowanie hipotez na temat takiego stanu rzeczy. Kalkulator, dzięki zestawowi kilku predefiniowanych wykresów, pozwala na graficzną prezentację analizowanych danych. Możliwe do przygotowania typy wykresów to: wyniki przewidywane, wskaźniki EWD, średnie wyniki, rozkład wyników, uprzednie osiągnięcia, rozkład reszt, wykres rozrzutu, krzywe przewidywanego wyniku. Kalkulator umożliwia również zestawianie analizowanych wskaźników w formie tabelarycznej. Wykresy i tabelę można zestawiać ze sobą w jednym widoku, co znacznie ułatwia interpretację uzyskiwanych wyników. Można je też eksportować do dowolnego edytora tekstu.

Przykładowa analiza danych z wykorzystaniem Kalkulatora EWD SP

Analizę wyników przykładowej szkoły rozpoczniemy od zapoznania się z graficzną prezentacją średnich wyników uczniów w teście OBUT, na sprawdzianie oraz ze wskaźnikiem EWD dla szkoły. Zestawienie takich wykresów przedstawia poniższy rysunek.

Rysunek 4.20. Średnie wyniki w szkole testu OBUT 2011 i sprawdzianu 2014 oraz wskaźniki EWD – przykład

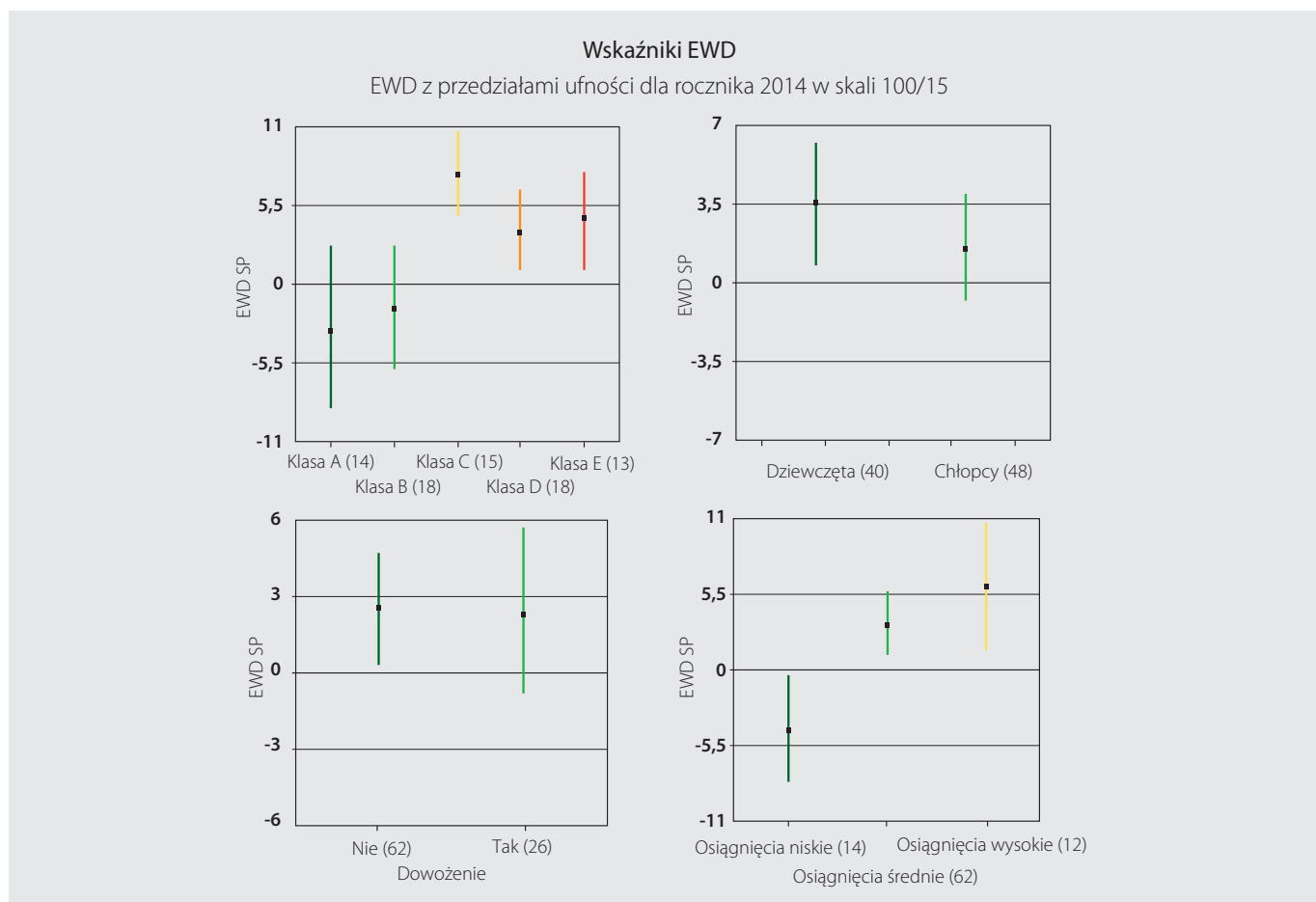


W analizowanej szkole wynik OBUT jest niższy od średniej. Jednak przedział ufności – przedstawiony na wykresie jako pomarańczowy odcinek – przecina linię o wartości 100, co nakazuje zinterpretować wynik szkoły jako nierozróżnialny statystycznie od średniej w grupie odniesienia – w uproszczeniu można przyjąć, że w kraju. Na pisanym trzy lata później sprawdzianie w klasie VI średni wynik nieznacznie przekracza przeciętną dla kraju, choć i w tym wypadku przedział ufności nakazuje uznanie, że jest on nierozróżnialny statystycznie od średniej w kraju. Wskaźnik EWD dla tej placówki wynosi

2,5 i cały przedział ufności znajduje się ponad osią X, która pokazuje średnią wartość EWD w kraju. Można zatem powiedzieć, że efektywność nauczania w tej szkole jest ponadprzeciętna.

Analizy w Kalkulatorze pozwalają określić, czy w różnych grupach uczniów wskaźniki EWD różnią się. W tym celu można obliczyć wskaźniki EWD osobno dla każdego z oddziałów klasowych. Takie porównania można również wykonywać w podziale na płeć, czy poziom uprzednich osiągnięć. Ze-stawienie takich wykresów przedstawia poniższy rysunek.

Rysunek 4.21. Wskaźnik EWD szkoły w podziale na różne grupy uczniowskie



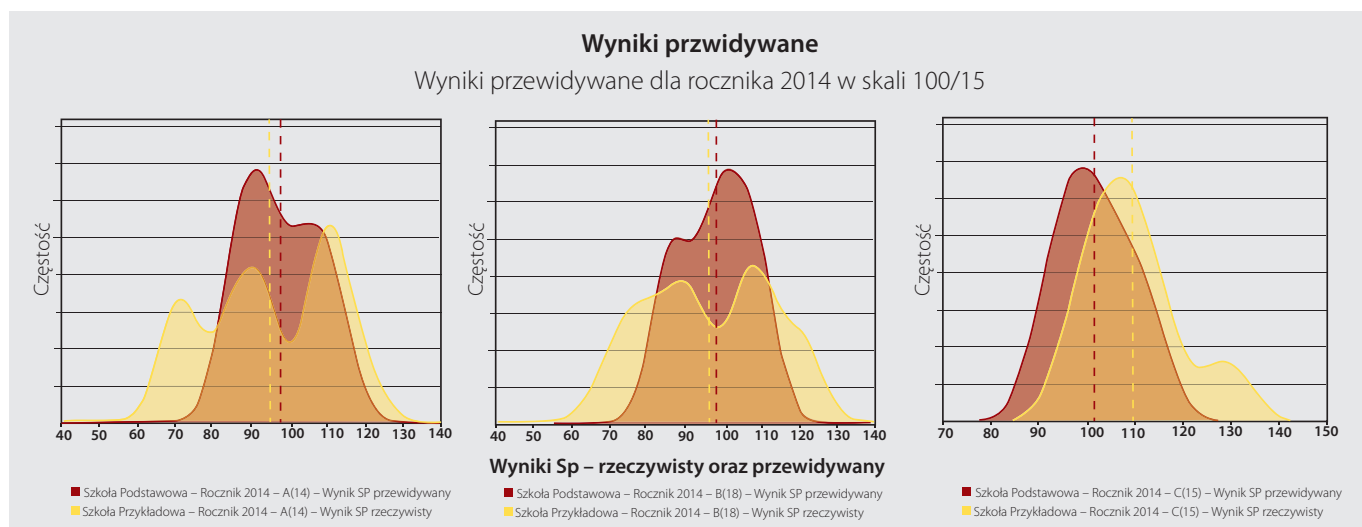
Wykresy sporządzone dla analizowanej szkoły sugerują, że nie ze wszystkimi grupami uczniów pracowano równie efektywnie. Efektywność nauczania w obszarach objętych sprawdzianem w klasach A i B była statystycznie istotnie niższa niż w oddziale C. W grupie chłopców i dziewcząt EWD jest porównywalna, podobnie jak wśród uczniów dowożonych i niedowożonych do szkoły. To, co szczególnie zwraca uwagę, to różne wartości EWD dla grup wyróżnionych ze względu na wyniki w testach OBUT, czyli poziom uprzednich osiągnięć. Uzyskane wyniki wskazują, że efektywność nauczania w grupie uczniów o wysokich uprzednich osiągnięciach jest o ok. 10 punktów wyższa niż w grupie uczniów o niskich wynikach na teście OBUT. To bardzo duża różnica: siłę efektu możemy opisać jako $2/3$ odchylenia standardowego na skali wyników indywidualnych sprawdzianu. Trzeba w tym miejscu przypomnieć, że model EWD jest tak skonstruowany, że w całej grupie odniesienia wyniki testu OBUT nie są skorelowane z wartością wskaźników efektywności.

Zaobserwowane zróżnicowanie EWD dla oddziałów i dla grup uczniów wyróżnionych ze względu na uprzednie osiągnięcia powinny w analizowanej szkole być impulsem uruchamiającym proces autoewaluacyjny. Jakie są przyczyny zaobserwowanych różnic? By ich dociec zwykle trzeba wykroczyć poza dane z testów osiągnięć szkolnych, jednak Kalkulator EWD też może być do tego pomocny.

4. Metoda edukacyjnej wartości dodanej w Polsce

Problemy z efektywnym nauczaniem w klasach A i B można zobrazować za pomocą wykresów wyników przewidywanych i faktycznie uzyskanych. Rysunek 4.22. przedstawia odpowiednie wykresy dla oddziałów A, B i C. W dwóch pierwszych widać wyraźnie, że wyniki faktycznie uzyskane są bardziej zróżnicowane niż wyniki przewidywane. Oznacza to, że istniały duże różnice w przyrostach dla poszczególnych uczniów, co sugeruje, że nie ze wszystkimi uczniami pracowano równie efektywnie. Dla porównania w klasie C osiągnięcia uczniów były bardziej zbliżone do siebie.

Rysunek 4.22. Przewidywane i faktycznie uzyskane wyniki sprawdzianu w wybranych oddziałach



Z przedstawionej powyżej przykładowej analizy wyników OBUT i sprawdzianu w klasie VI wynika, że choć efektywność nauczania w tej szkole jest ponadprzeciętna, to są jednak obszary, w których poprawa mogłaby się przełożyć na jeszcze lepsze wyniki uczniów w szkole. Wykryte w tej szkole problemy to niska efektywność nauczania w oddziałach A i B oraz w grupie uczniów o niskich wynikach po I etapie edukacyjnym. Analiza wskaźników EWD powinna być punktem wyjścia do autoewaluacji szkoły. Prowadzić powinna do stawiania pytań i poszukiwania przyczyn zaobserwowanego stanu rzeczy oraz ich zweryfikowania na podstawie danych ilościowych (np. ocen szkolnych, danych o frekwencji), czy jakościowych (obserwacji lekcji, danych z wcześniejszych ewaluacji wewnętrznych). W takim rozumieniu korzystanie z Kalkulatora EWD SP jest sposobem na wzmocnienie ewaluacji wewnętrznej w szkole.

Ograniczenia i rozwój metody EWD na II etapie edukacyjnym

W przypadku wskaźników EWD dla II etapu edukacyjnego podstawowe ograniczenie rozwoju metody wynika z braku stabilnego, powszechnego pomiaru osiągnięć szkolnych po III klasie. Jeżeli szkoła nie bierze udziału w badaniu OBUT, nie może użyć metody EWD do analizy wyników sprawdzianu w klasie VI. Oznacza to znaczące ograniczenie przydatności wyników sprawdzianu z punktu widzenia ewaluacyjnej funkcji systemu egzaminów krajowych. Kolejnym problemem jest brak spójnej z systemem egzaminacyjnym bazy wyników badania OBUT. Sprawne łączenie wyników OBUT i sprawdzianu w klasie VI pozwoliłoby na obliczanie wskaźników EWD w oparciu o wyniki całej populacji uczniów biorących udział w pomiarze umiejętności na zakończenie I etapu edukacyjnego. Trzecie ograniczenie jest związane z niedoskonałą standaryzacją testów stosowanych w badaniu OBUT. Obecnie w sytuacji, gdy obserwowana jest niska wartość wskaźnika EWD przy równoczesnej wysokiej średniej testu OBUT, należy zastanawiać się nie tylko nad efektywnością nauczania w klasach IV –VI, ale również nad sposobem przeprowadzania i oceniania testu OBUT. Poprawa jakości testowania powinna przełożyć się na podwyższenie rzetelności i trafności wskaźników EWD SP. Jednak należy pamiętać, że testy tworzone na potrzeby OBUT mają też do spełnienia inne funkcje niż

tylko pomiar poziomu początkowego w modelach EWD. Z pewnością w kolejnych edycjach tego badania trzeba starannie przemyśleć jego funkcje i wynikające z tego pożądane własności testów. Wydaje się, że pomimo opisanych ograniczeń wskaźniki EWD mogą być dla szkół podstawowych ważnym źródłem informacji wykorzystywanym w procesie ewaluacji wewnętrznej. Dają unikalną możliwość oparcia refleksji nad działaniami prowadzonymi w szkole na obiektywnych danych.

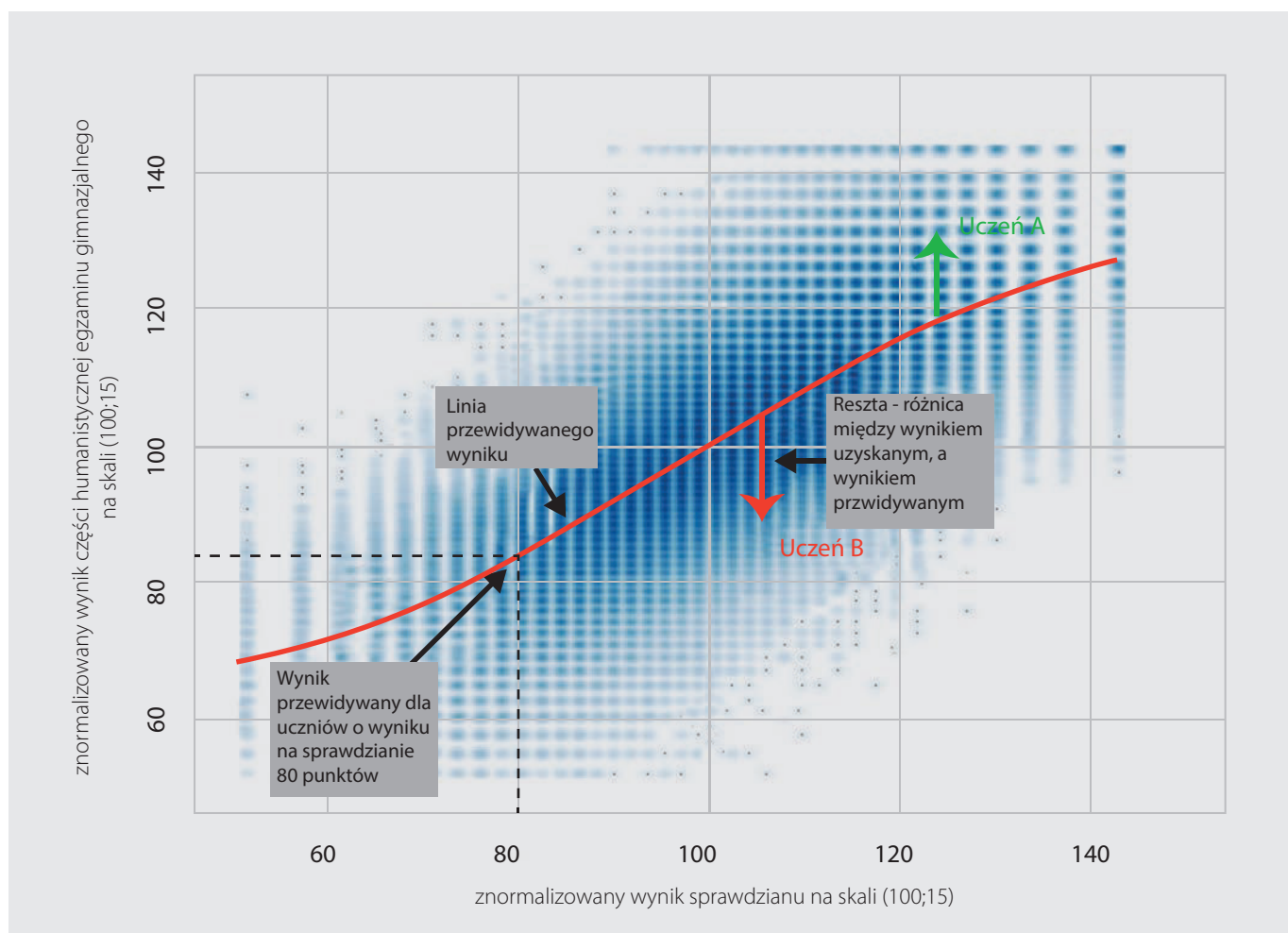
4.4. Wykorzystanie metody edukacyjnej wartości dodanej w gimnazjach

Prace nad modelami edukacyjnej wartości dodanej dla gimnazjów mają najdłuższą tradycję. Opracowaniu metodologii szacowania wskaźników EWD towarzyszyły prace nad narzędziami analizy i materiałami szkoleniowymi dla nauczycieli i dyrektorów gimnazjów oraz innych użytkowników metody. Obecnie (wiosna 2015 roku) dostępne są trzyletnie wskaźniki EWD dla ponad 6 tysięcy gimnazjów, w siedmiu trzyletnich okresach, od okresu 2006–2008 rozpoczynając. Dostępny jest również Kalkulator EWD 100, w którym można prowadzić analizy na danych egzaminacyjnych od 2009 roku.

Jednoroczne i trzyletnie modele edukacyjnej wartości dodanej dla gimnazjów

Jednoroczne modele EWD dla gimnazjów wykorzystywane są w Kalkulatorze EWD 100. Sposób obliczania wskaźników EWD przez Kalkulator EWD jest analogiczny do modelu omówionego w poprzednim podrozdziale. Statystyczną logikę metody pokazuje poniższy wykres.

Rysunek 4.23. Linia przewidywanego wyniku na egzaminie gimnazjalnym w zależności od wyniku na sprawdzianie w klasie VI. Analiza dla przykładowego rocznika



Jeśli w szkole przeważają uczniowie typu A, to efektywność nauczania jest ponadprzeciętna. Jeśli w szkole przeważają uczniowie typu B, czyli uczniowie, którzy uzyskali wyniki poniżej przewidywania, szkoła naucza z efektywnością poniżej przeciętnej, a wskaźnik EWD szkoły będzie ujemny. Wskaźniki EWD dla gimnazjów, tak jak dla szkół podstawowych, są miarami względnymi, a grupą odniesienia są gimnazja o podobnym składzie uczniowskim ze względu na uprzednie osiągnięcia. Metodę szacowania trochę komplikuje uwzględnienie w modelu zmiennych kontrolnych. Zmienne kontrolne to te czynniki, które mogą mieć wpływ na postęp uczniów, a nie są zależne od szkoły. Jeśli chcemy, aby wskaźniki EWD pokazywały efektywność pracy szkół, takie czynniki powinny być dodatkowo kontrolowane tak, aby różnice pomiędzy uczniami uczącymi się w różnych szkołach nie miały wpływu na wartości wyliczanych dla tych szkół wskaźników EWD. Generalnie przyjmuje się, że uwzględnienie w modelach zmiennych kontrolnych pozytywnie wpływa na własności wyliczanych wskaźników EWD. Można tu jednak sformułować dwa zastrzeżenia. Po pierwsze, zakres kontrolowanych zmiennych powinien być dostosowany do celu, jakiemu mają służyć wskaźniki. Po drugie, w sytuacji nielosowego przydziału uczniów do szkół – co ma miejsce właściwie w każdym systemie szkolnym, choć z różnym nasileniem – część wpływu wynikającego z działań szkół może być w procesie estymacji modelu przypisywana oddziaływaniu zmiennych kontrolnych. Poszerzenie zakresu czynników kontrolowanych w modelu z jednej strony chroni więc przed „niesprawiedliwym” ocenianiem szkół, jednak z drugiej może utrudniać rozróżnianie od siebie szkół pracujących efektywnie i nieefektywnie (Ballou, Sanders i Wright, 2004; McCafrey i in., 2003; McCafrey i in., 2004; OECD, 2008; Raudenbush i Willms, 1995).

W gimnazjalnych modelach edukacyjnej wartości dodanej uwzględnia się dość ubogi zestaw zmiennych kontrolnych. Podstawowym problemem związanym z uwzględnianiem w modelu zmiennych kontrolnych jest słabość polskiego systemu informacji oświatowej. W praktyce niemożliwe jest dołączenie do wyników egzaminacyjnych żadnych dodatkowych informacji o uczniach ponad te, które przechowywane są w bazach danych okręgowych komisji egzaminacyjnych. Zakres zmiennych uwzględnianych w modelach obejmuje płeć i informacje o dysleksji (podczas sprawdzianu, podczas egzaminu gimnazjalnego oraz interakcję tych dwóch zmiennych).

Włączenie do modeli płci oznacza, że jeśli w skali ogólnokrajowej występują różnice w średnich postępach dziewcząt i chłopców, to nie będą one mieć wpływu na wskaźniki EWD szkół – punktem odniesienia dla dziewcząt będą inne dziewczęta, a dla chłopców inni chłopcy. Analogiczna sytuacja występuje w przypadku informacji o korzystaniu z dostosowań dla dyslektyków, z tym że inny punkt odniesienia jest tu ustalany w ramach każdej z czterech grup: uczniów, którzy nie korzystali z takiego dostosowania, uczniów, którzy korzystali z dostosowania na sprawdzianie, uczniów, którzy korzystali z dostosowania tylko na egzaminie gimnazjalnym i wreszcie uczniów, którzy korzystali z dostosowań podczas obu egzaminów. Istnieją powody, by przypuszczać, że posiadanie opinii poradni o dysleksji nie jest zbyt dobrym wskaźnikiem występowania specyficznych trudności w nauce czytania i pisanie (Dolata, 2007). Jest za to jasne, że wyniki egzaminów takich osób nie powinny być bezpośrednio porównywane z wynikami innych zdających, gdyż uczniowie korzystający z dostosowań dla dyslektyków mają wydłużony czas pisania egzaminu i stosuje się wobec nich zmodyfikowane kryteria oceny w zakresie wybranych typów błędów.

Na uwagę zasługuje również grupa laureatów konkursów przedmiotowych. Nie podchodzą oni do egzaminu (w przypadku egzaminu gimnazjalnego – do jednej z jego części, odpowiadającej tematyce konkursu, którego są laureatami), lecz mają przypisywane maksymalne możliwe do uzyskania wyniki. Jednocześnie laureaci stanowią grupę niezwykle zróżnicowaną wewnętrznie. Liczba konkursów dających możliwość zwolnienia z egzaminu jest znaczna, a do tego różna na terenie różnych województw, ponieważ organizatorami konkursów przedmiotowych są kuratoria oświaty. Brak też standaryzacji co do trudności różnych konkursów. Sprawia to, że sensowność porównywania laureatów do innych laureatów jest mocno problematyczna. W związku z tym zdecydowano się nie uwzględniać w modelach informacji o byciu laureatem, uczniowie tacy traktowani są tak samo jak inni, którzy otrzymali z egzaminu maksymalną liczbę punktów. Z punktu widzenia EWD najlepszym

rozwiązaniem problemu związanego z laureatami byłoby, gdyby pisali oni egzaminy tak, jak wszyscy uczniowie, a wyniki konkursów ujawniane były dopiero po napisaniu egzaminów. W ten sposób wyniki konkursów mogłyby być uwzględniane w rekrutacji, jednak w danych egzaminacyjnych dysponowalibyśmy również dobrym oszacowaniem bardziej ogólnych umiejętności laureatów, porównywalnym z wynikami innych uczniów.

Poza modelami statystycznymi pozwalającymi obliczać wskaźniki jednoroczne tworzone są dla gimnazjów modele wieloletnie, które obejmują wyniki trzech kolejnych roczników absolwentów gimnazjów. Wskaźniki jednoroczne i trzyletnie różnią się metodami statystycznymi, ale główną różnicę stanowi ich przeznaczenie i sposób udostępniania odbiorcom.

Tabela 4.5. Porównanie cech jednorocznych i trzyletnich wskaźników EWD.

Wskaźniki jednoroczne	Wskaźniki trzyletnie
obliczane na podstawie wyników jednego rocznika absolwentów	obliczane na podstawie wyników trzech kolejnych roczników absolwentów
dla bardzo małych szkół wskaźniki są niewiarygodne, ze względu na duży wpływ pojedynczych uczniów	mogą zostać opublikowane także dla małych szkół (więcej uczniów)
mogą podlegać dużym wahaniom z roku na rok	mało podatne na krótkookresowe wahania, ale w większym stopniu oparte na danych historycznych, nieodzwierciedlających aktualnej efektywności szkoły
pozwalają przyrzeć się szczegółom pracy szkoły	pozwalają na bardziej ogólną ocenę pracy szkoły
dostępne tylko dla dyrektorów szkół (lub za ich pośrednictwem) – wymagają dostępu do połączonych danych sprawdzian–egzamin gimnazjalny	powszechnie dostępne
możliwość prowadzenia wielu rodzajów analiz w Kalkulatorze EWD	tylko ściśle określony zestaw wskaźników udostępniany na stronie internetowej
analizy głównie w ramach jednej szkoły	łatwe porównanie między szkołami

Jednoroczne gimnazjalne wskaźniki EWD można było wyznaczać dla szkół po raz pierwszy w 2006 r. Początkowo ich wyliczanie umożliwiały odpowiednio przygotowane arkusze kalkulacyjne, udostępniane odbiorcom pod nazwą Kalkulator EWD. Od 2009 r. udostępniana była szkołom samodzielna aplikacja służąca do wyliczania wskaźników Kalkulator EWD Plus, zastąpiona w 2012 r. przez Kalkulator EWD 100. Do 2011 r. dla każdego roku można było wyznaczać dwa wskaźniki: humanistyczny i matematyczno-przyrodniczy. W 2012 r., w związku ze zmianą formuły egzaminu gimnazjalnego, zakres wskaźników rozszerzono o cztery nowe wskaźniki, które jako miary umiejętności „na wyjściu” wykorzystują wyniki poszczególnych testów, z których składa się nowy egzamin gimnazjalny. Są to wskaźniki z zakresu: języka polskiego, historii i wiedzy o społeczeństwie, matematyki oraz przedmiotów przyrodniczych. Wszystkie jako miarę umiejętności „na wejściu” wykorzystują wyniki sprawdzianu (Pokropek i Żółtak, 2012b). Zmiany wprowadzone w 2012 r., oprócz rozszerzenia liczby wyliczanych wskaźników, objęły zastąpienie w regresji „surowych” wyników egzaminacyjnych wynikami wyrażonymi na skali standardowej oraz uwzględnienie uczniów, których tok kształcenia w gimnazjum był dłuższy o jeden rok (Pokropek i Żółtak, 2012b). Wykorzystanie wyników wyrażonych na standardowej skali 100;15 miało na celu zarówno polepszenie własności tak przekształconych zmiennych w modelach regresji, jak i ułatwienie interpretacji wyników, między innymi poprzez uzyskanie względnej porównywalności między latami. Wadą wyników surowych polskich egzaminów

jest bowiem zróżnicowany pomiędzy latami poziom trudności i rozkład wyników – normalizacja pozwala te problemy w znacznej mierze usunąć, co sprawia, że wyniki z różnych lat mogą być interpretowane tak samo.

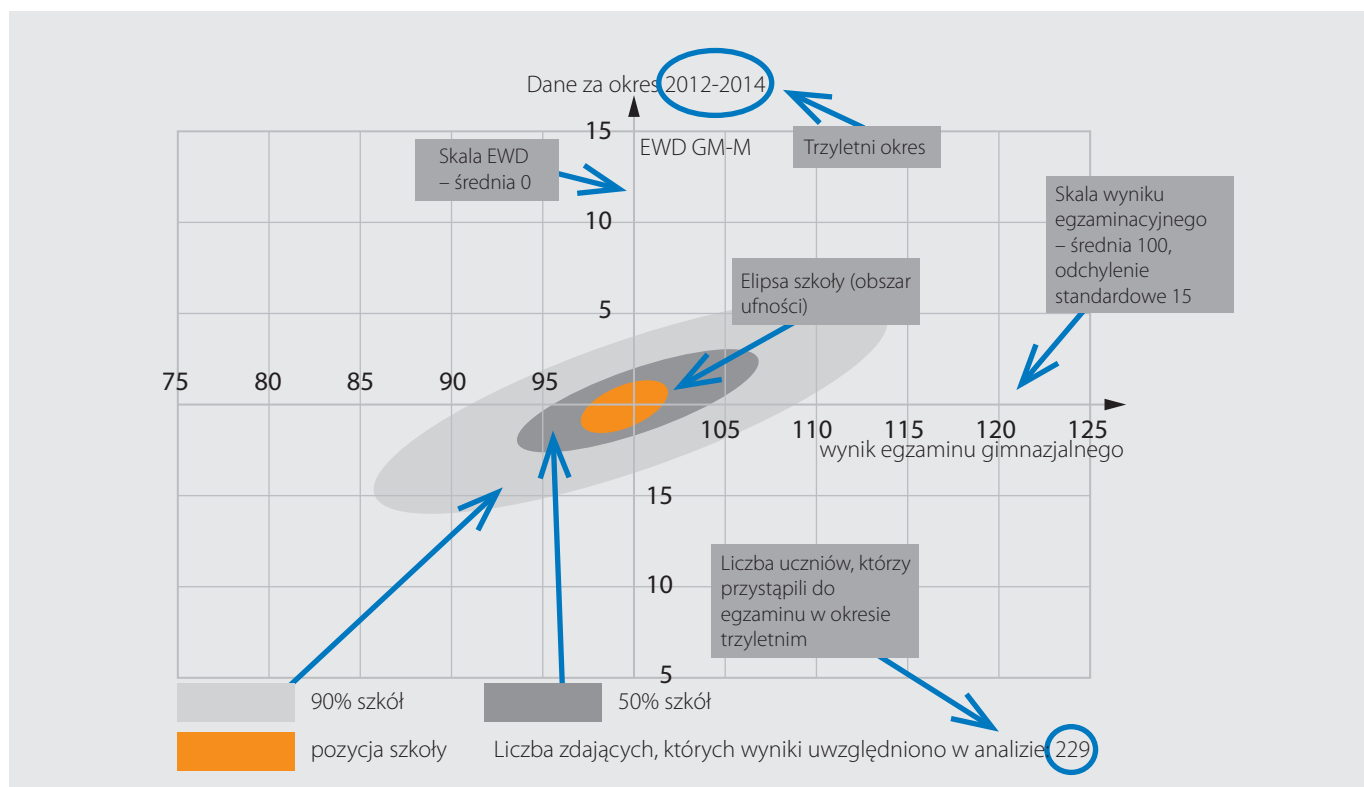
Wskaźniki jednoroczne EWD przeznaczone są przede wszystkim na potrzeby ewaluacji wewnętrznej. Odbiorcom nie są przy tym udostępniane gotowe wskaźniki, wyliczone dla poszczególnych szkół, lecz aplikacja pozwalająca wyliczyć wartość EWD dla dowolnej grupy, określonej przez użytkownika na podstawie danych, wczytanych przez niego do tej aplikacji. Pozwala to na prowadzenie bardziej szczegółowych analiz nad zróżnicowaniem efektywności nauczania, jednak stawia użytkownikom dużo wyższe wymagania zarówno w zakresie technicznych umiejętności posługiwania się dostarczanym im narzędziem, jak i w zakresie interpretacji uzyskiwanych wyników. Grupa osób mogących korzystać ze wskaźników jednorocznych jest przy tym ograniczona do tych, którzy dysponują połączonymi wynikami sprawdzian–egzamin gimnazjalny dla analizowanej grupy uczniów, w praktyce głównie nauczycieli i dyrektorów.

Gimnazjalne wskaźniki trzyletnie opublikowane zostały po raz pierwszy w 2009 r. Analogicznie jak w przypadku wskaźników jednorocznych były to dwa wskaźniki: humanistyczny i matematyczno-przyrodniczy. Do rozszerzenia zakresu wskaźników o cztery kolejne, odnoszące się do nowej struktury egzaminu gimnazjalnego, doszło jednak dopiero w 2014 r., gdy wszystkie trzy sesje egzaminacyjne objęte wskaźnikiem trzyletnim (2012–2014) były przeprowadzone według nowej formuły. Wskaźniki trzyletnie publikowane są w ogólnodostępnym serwisie internetowym (<http://ewd.edu.pl/wskazniki/gimnazjum>), z myślą o bardzo szerokiej grupie odbiorców, od nauczycieli i dyrektorów, poprzez organy prowadzące szkoły i nadzór pedagogiczny. W związku z tym zdecydowano się na wyliczanie wskaźników na podstawie wyników z trzech kolejnych sesji egzaminu gimnazjalnego tak, by były one bardziej stabilne, zdawały sprawozdanie z przeciętnej efektywności pracy szkoły w dłuższym okresie i nie prowokowały do wyciągania zbyt pochopnych wniosków na podstawie różnic, jakie mogą występować pomiędzy pojedynczymi latami.

Trzyletnie wskaźniki egzaminacyjne dla gimnazjów są łączną informacją o wyniku egzaminacyjnym i edukacyjnej wartości dodanej (EWD) i są prezentowane w formie graficznej. Przy konstrukcji trzyletnich wskaźników EWD zadbano o poprawność statystyczną prezentacji wyniku (Pokropek i Żółtak, 2012a). Do 2011 r. w modelach wykorzystywane były wyniki sprawdzianu i egzaminu gimnazjalnego w prosty sposób przekładane na skalę standardową (tzw. normalizacja ekwikutylowa), uwzględniano przy tym wyłącznie wyniki uczniów, którzy uczyli się w gimnazjum przez trzy lata (Pokropek i Żółtak, 2012a). Począwszy od wskaźników dla okresu 2012–2014 wyniki egzaminów skalowane są w bardziej zaawansowany sposób (metodologia IRT przy pomocy modelu 2PL/graded response), a w modelach regresji uwzględniono też uczniów, których tok kształcenia był dłuższy o jeden rok. Więcej informacji na temat sposobu modelowania statystycznego trzyletnich gimnazjalnych wskaźników EWD można znaleźć w dokumentacji technicznej, dostępnej na stronie internetowej EWD (Pokropek i Żółtak, 2012b).

Trzyletni wskaźnik egzaminacyjny uwzględnia niepewność pomiarową (95% przedziały ufności) zarówno dla wyniku egzaminacyjnego, jak i dla edukacyjnej wartości dodanej, a jego reprezentacją jest elipsa (obszar ufności). Wielkość obszaru ufności zależy od liczby zdających. Pozycja elipsy jednocześnie informuje o średnim wyniku egzaminacyjnym oraz efektywności nauczania w zakresie sprawdzanym przez egzamin zewnętrzny.

Na poniższym rysunku pokazano najważniejsze elementy graficznej prezentacji trzyletnich wskaźników egzaminacyjnych. Dwie środkowe (szare) elipsy, zwane warstwicami, wskazują obszary, w których koncentruje się odpowiednio około 50% i około 90% wyników szkół. Innymi słowy, warstwice określają obszar najbardziej typowych wyników.

Rysunek 4.24. Podstawowe elementy wykresu trzyletnich wyników egzaminacyjnych/wskaźników EWD


Trafność wskaźników EWD dla gimnazjów

By wskaźniki EWD można było odpowiedzialnie polecać szkołom, musimy mieć pewność, że rzeczywiście odzwierciedlają one efektywność nauczania. Wyniki wstępnych badań, prowadzonych w latach 2005–2007, dawały powody do optymizmu, były jednak zbyt skromne pod względem wielkości próby i zakresu zgromadzonych danych, by wyciągać ostateczne wnioski. Dlatego w latach 2009–2012 wśród gimnazjalistów przeprowadzono szeroko zakrojone, podłużne badanie, którego celem była odpowiedź na pytania o niektóre aspekty trafności wskaźników EWD. Uczestniczyli w nim uczniowie, rodzice, nauczyciele i dyrektorzy szkół gimnazjalnych. Dzięki objęciu badaniami próby dobrze reprezentującej populację publicznych gimnazjów, możliwe było wnioskowanie o wszystkich tego typu szkołach w Polsce.

Ramka 4.2. Opis podłużnego badania uwarunkowań wyników nauczania w gimnazjach

Badanie podłużne w gimnazjach przeprowadzono w latach 2009–2012 i objęto nim reprezentatywną, ogólnopolską losową próbę 292 oddziałów klas pierwszych ze 150 szkół gimnazjalnych (ok. 5 tys. uczniów). Zostało ono zrealizowane w ramach projektu systemowego *Badania dotyczące rozwoju metodologii szacowania wskaźnika edukacyjnej wartości dodanej (EWD)*. Śledzono w nim losy uczniów od pierwszej do trzeciej klasy gimnazjum, włączając do badania także ich rodziców oraz nauczycieli i dyrektorów szkół.

Głównym celem badania była weryfikacja, czy obliczane obecnie wskaźniki EWD dla gimnazjów trafnie oddają efektywność nauczania szkół. Zebrane dane dostarczyły również cennej wiedzy dotyczącej indywidualnych, rodzinnych i szkolnych czynników odpowiedzialnych za osiągnięcia szkolne uczniów. Dodatkowo w badaniu monitorowano opinie kadry pedagogicznej na temat egzaminów zewnętrznych.

Podczas pierwszego etapu badania, który odbył się, gdy uczniowie uczęszczali do pierwszych klas, zebrano m.in. dane opisujące uczniów (płeć, data urodzenia, poziom inteligencji,

wcześniejsze doświadczenia edukacyjne, aspiracje) i ich rodziny (różne miary statusu społeczno-ekonomicznego, aspiracje edukacyjne względem dzieci, struktura rodziny). Dyrektorów i nauczycieli pytano o opinie na temat egzaminów zewnętrznych. W kolejnym etapie badania, który odbył się na początku III klasy, sprawdzono osiągnięcia szkolne uczniów w zakresie czytania oraz matematyki. W etapie trzecim ponownie dokonano pomiaru inteligencji uczniów, sprawdzono ich motywację do nauki, integrację szkolną oraz zapytano o przygotowania do egzaminu gimnazjalnego i korzystanie z dodatkowej pomocy w nauce. Od dyrektorów i nauczycieli pozyskano cenne informacje nt. funkcjonowania szkoły (przywództwa, kultury organizacyjnej) oraz stosunku do egzaminów zewnętrznych i EWD. Więcej informacji na temat badania znajduje się na stronie: www.ewd.edu.pl

Istnieje wiele sposobów weryfikacji trafności wskaźników EWD. W badaniu skoncentrowano się na trzech z nich. Po pierwsze sprawdzono, czy dane wykorzystywane do ich wyliczenia, tj. wyniki testów egzaminacyjnych, są odpowiedniej jakości. Ponadto zbadano, czy modele EWD skutecznie kontrolują czynniki niezależne od szkoły, rzutujące na postępy uczniów. Tylko wtedy bowiem będziemy mogli zasadnie przypisać wartość wskaźnika EWD działaniom szkoły. Rozpatrywano czynniki indywidualne (status społeczno-ekonomiczny i inteligencja uczniów, korzystanie z korepetycji) oraz lokalne (społeczno-gospodarcze charakterystyki gmin). Po trzecie zbadano, czy szkoły o wysokim EWD, to szkoły, w których zgodnie z naszą normatywną wiedzą o funkcjonowaniu szkoły dobrze się dzieje. Dlatego zweryfikowano m.in. to, czy w szkołach o wysokim EWD uczniowie szybciej rozwijają się poznawczo. Poniżej przedstawiono wybrane wyniki analiz. Pełny opis wyników badania trafności wskaźników EWD można znaleźć w książce *Trafność metody edukacyjnej wartości dodanej dla gimnazjów* (Dolata i in., 2013).

Jeśli pomiar osiągnięć szkolnych nie jest wystarczająco dobry, nawet najbardziej wyrafinowane metody statystyczne nie dostarczą wartościowych miar. W badaniu zweryfikowaliśmy więc, czy wykorzystane egzaminy spełniają cztery podstawowe kryteria, konieczne, by można było je wykorzystać do obliczania EWD. Były to: rzetelność, odzwierciedlanie realizacji szeroko rozumianych celów kształcenia, pomiar na „wejściu” i „wyjściu” podobnych umiejętności oraz dobrze zdefiniowana jednostka pomiaru.

Przeprowadzone analizy pokazały, że rzetelność testów egzaminacyjnych jest wystarczająca dla tworzenia miar charakteryzujących szkoły. W przypadku wykorzystanego w badaniu sprawdzianu w 2009 roku wyniosła ona 0,83. W przypadku egzaminu gimnazjalnego w 2012 roku wahała się od 0,78 do 0,91 w zależności od części egzaminu. Wartości te są typowe dla wszystkich edycji zarówno sprawdzianu, jak i egzaminu gimnazjalnego.

Drugą cechą, którą powinny charakteryzować się egzaminy wykorzystywane do szacowania EWD, jest dobre odzwierciedlanie przez ich wyniki stopnia, w jakim szkoła zrealizowała najważniejsze cele kształcenia. Krytycy testowania wskazują, że egzaminy pozwalają mierzyć jedynie skuteczność rozwiązania danego zestawu zadań i że można maksymalizować wynik egzaminu przez różne praktyki nauczania opatrywane zbiorczą nazwą „nauczanie pod testy”. Praktyki te, choć pozwalałyby osiągać dobre wyniki egzaminacyjne miałyby wątpliwą wartość edukacyjną, bo budowałyby niekorzystne wzorce motywacji, zawężałyby program kształcenia do tego, co sprawdzane na egzaminach, czy też ograniczałyby wpływ potencjału intelektualnego dziecka na szkolne osiągnięcia.

By zweryfikować te zarzuty, sprawdzono, czy uczniowie, którzy dobrze wypadają na egzaminie gimnazjalnym, wykazują także inne pożądane cechy, świadczące o tym, że wysoki wynik egzaminacyjny nie jest okupiony licznymi efektami ubocznymi. Wyniki analiz pokazują, że uczniowie ci otrzymują wyższe oceny w szkole, wykazują się korzystnym wzorcem motywacji (wysoka, wewnętrzna), mają pozytywne nastawienie do nauki i szkoły oraz wierzą we własne możliwości uczenia się. Zaobserwowano również negatywny związek wyników egzaminów z bezradnością intelektualną, czyli

stanem, który powstaje w sytuacji, w której uczeń mimo długotrwałego wysiłku umysłowego nie jest w stanie zrozumieć danego materiału. Ujawnił się również pozytywny, dość silny związek wyników egzaminacyjnych z inteligencją uczniów. Podsumowanie uzyskanych wyników przedstawiono w poniższej tabeli.

Tabela 4.6. Związki między wynikami egzaminów a zewnętrznymi kryteriami

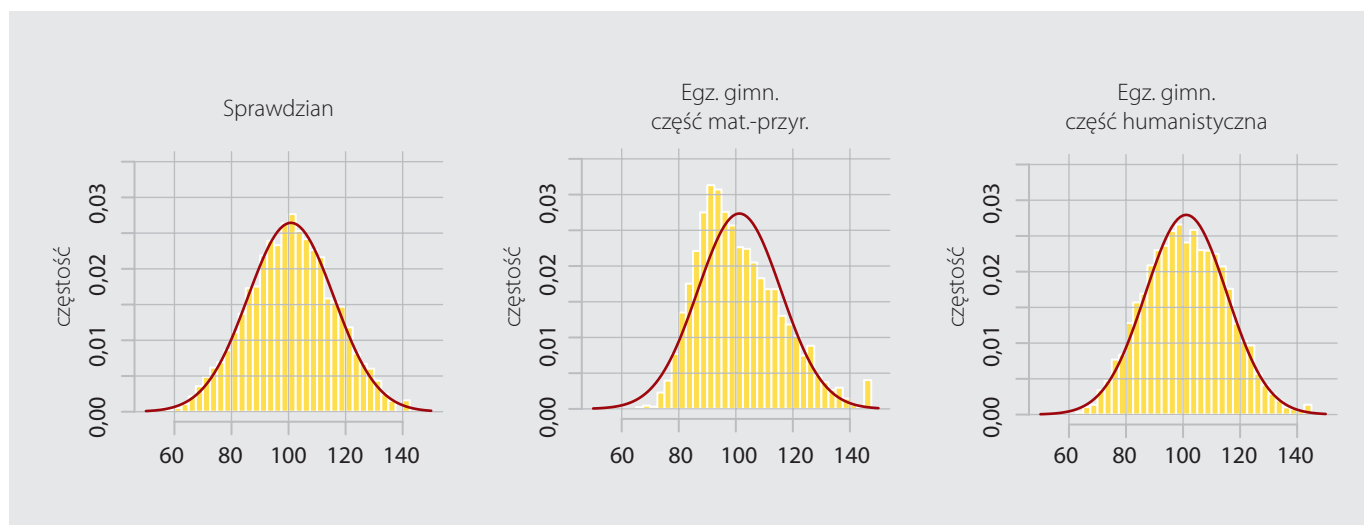
oceny szkolne	+
motywacja niska	x
motywacja wysoka	+
motywacja zewnętrzna	x
motywacja wewnętrzna	+
pozytywne nastawienie do nauki i szkoły	+
wiara we własne możliwości uczenia się	+
bezzadność intelektualna	-
inteligencja	+

+ związek pozytywny; - związek negatywny; x brak związku

Wykorzystywane w modelowaniu EWD testy powinny mierzyć te same lub zbliżone umiejętności. Im będą one sobie bliższe, tym lepiej miara EWD będzie odzwierciedlała rzeczywisty wkład szkoły w osiągnięcia uczniów. Obliczając wskaźniki dla gimnazjów, wykorzystuje się wyniki sprawdzianu oraz egzaminu gimnazjalnego. Mimo iż sprawdzian jest testem wiedzy ogólnej, korelacje między nim a poszczególnymi częściami egzaminu gimnazjalnego są dość wysokie i wynoszą od ok. 0,71 do 0,78. Oznacza to, że wyniki na sprawdzianie w wystarczającym stopniu pozwalają przewidzieć wyniki na egzaminie gimnazjalnym. Innymi słowy, mogą być wartościową podstawą wyliczania wskaźników EWD.

Innym ważnym kryterium, które powinny spełniać testy egzaminacyjne wykorzystywane do obliczania EWD, jest dobrze zdefiniowana jednostka pomiaru skali, na której prezentowane są wyniki egzaminacyjne. W szczególności wzrost o jedną jednostkę na skali (np. o jeden punkt) powinien oznaczać równoważny przyrost wiadomości i umiejętności w każdym zakresie skali, tj. zarówno wśród uczniów o niskich, jak i przeciętnych oraz wysokich wynikach. W wypadku skal zdefiniowanych na podstawie odchylenia standardowego – a takie stosujemy w metodzie EWD – oznacza to, że rozkłady wyników powinny mieć kształt rozkładu normalnego. Dlatego do wyliczania wskaźników EWD nie wykorzystuje się wyników surowych. Wyniki egzaminów najpierw się skaluje metodami IRT i przekształca na skalę o średniej 100 i odchyleniu standardowym 15. Na poniższym rysunku pokazano rozkłady wyników wyskalowanych. Widoczne jest, że ich kształty są bliskie rozkładowi normalnemu (dla czterech testów z egzaminu gimnazjalnego mamy podobny obraz). Oznacza to, że możemy przyjąć, że wyniki testów wykorzystywane do wyliczania wskaźników EWD są przedstawione na skali o dobrze zdefiniowanej jednostce pomiaru.

Rysunek 4.25. Rozkłady wyskalowanych wyników testów egzaminacyjnych: sprawdzian 2009, egzamin gimnazjalny 2012



Podsumowując, przeprowadzone analizy pokazały, że wyniki egzaminów zewnętrznych wykorzystywane do konstrukcji gimnazjalnych wskaźników EWD mają wystarczająco dobre właściwości, by mogły stanowić podstawę wyliczania trafnej miary EWD.

Kolejnym problemem rzutującym na trafność wskaźników EWD jest wpływ na wyniki egzaminacyjne czynników pozaszkolnych, w szczególności pochodzenia społecznego, i inteligencji uczniów. Brak uwzględnienia przy wyliczaniu wskaźników EWD informacji o statusie społecznym rodziny i inteligencji ucznia jest często podnoszonym argumentem krytycznym. Wskazuje się, że status rodziny i poziom zdolności ucznia mogą wpływać na postępy w nauce, a tym samym wyznaczać szansę szkół na wysokie EWD. Jeżeli szkoły różnią się między sobą ze względu na te czynniki, to placówkom rekrutującym uczniów bardziej uzdolnionych lub z rodzin o wyższym statusie może być łatwiej uzyskać wysokie EWD niezależnie od efektywności nauczania.

Analizy pokazują, że gimnazja rzeczywiście znacząco różnią się statusem społecznym rodziców uczniów, którzy do nich trafiają. Największe różnice stwierdzono dla poziomu wykształcenia rodziców i indeksu statusu społeczno-ekonomicznego (ISEI). Około 20% zmienności tych wskaźników to zróżnicowanie międzyszkolne. Również w zakresie poziomu inteligencji uczniów zanotowano znaczne, choć mniejsze niż w wypadku wskaźników statusowych, różnice między gimnazjami. Około 10% zmienności wyników testu inteligencji to zróżnicowanie międzyszkolne.

Status społeczny i inteligencja uczniów mają znaczenie dla osiągnięć szkolnych uczniów. Spośród zbadanych aspektów statusu społecznego ucznia wyniki egzaminu gimnazjalnego najsilniej wyznacza wykształcenie rodziców. W wypadku części humanistycznej przeciętny dystans między dziećmi rodziców o wykształceniu zasadniczym zawodowym i wyższym wynosi prawie 15 punktów, czyli jedno odchylenie standardowe. Dla części matematyczno-przyrodniczej znaczenie wykształcenia rodziców jest troszkę mniejsze. Natomiast siła zależności między wynikami egzaminu gimnazjalnego a inteligencją jest prawie dwukrotnie większa niż obserwowana dla wykształcenia rodziców. Wzrost inteligencji o jedno odchylenie standardowe oznacza wzrost wyników o ok. 8–9 pkt (w zależności od egzaminu).

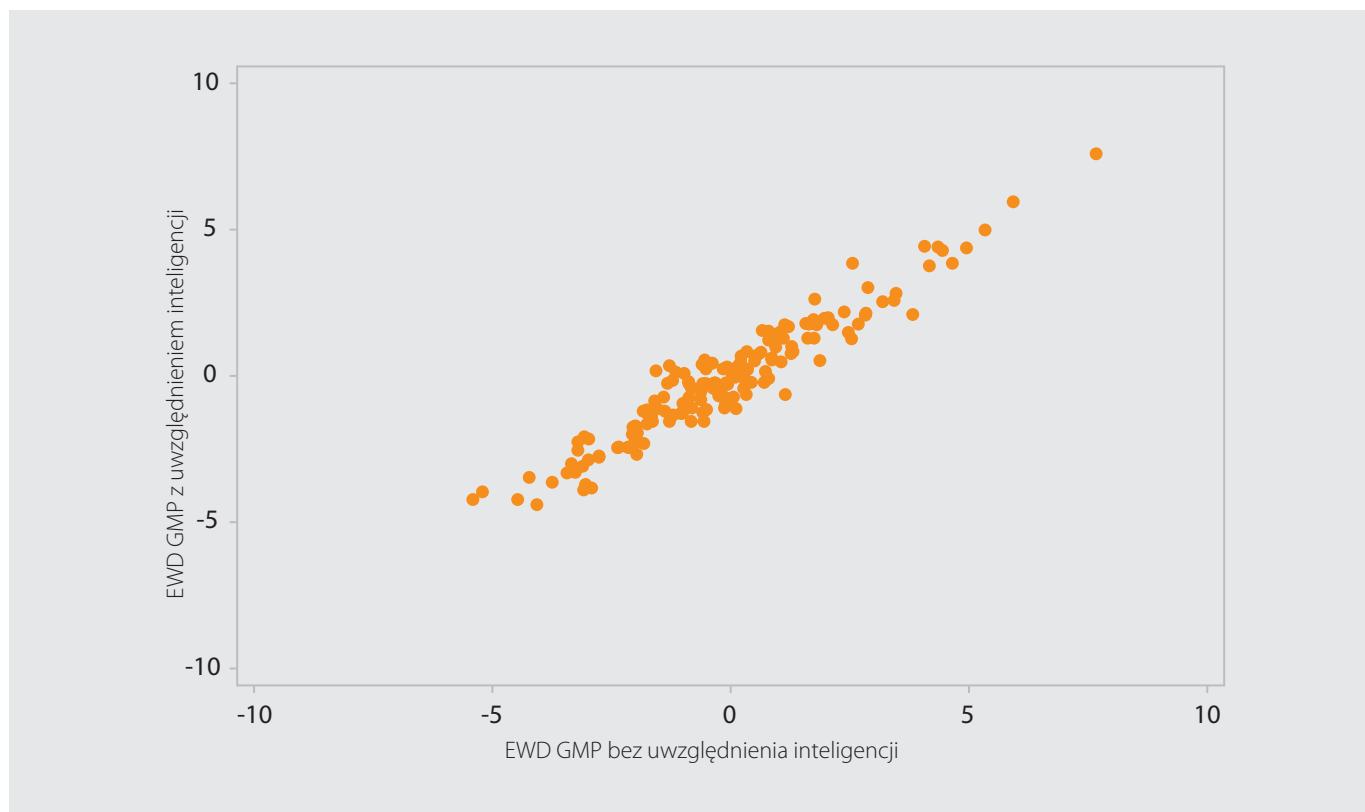
Jednakże uwzględnienie w analizie uprzednich osiągnięć, czyli wykorzystanie modelu EWD, zmienia siłę związku między czynnikami statusowymi, inteligencją a wynikami egzaminu gimnazjalnego. Siła związku wykształcenia rodziców z przyrostem umiejętności staje się ponad dwukrotnie niższa dla obszaru humanistycznego i ponad czterokrotnie niższa dla obszaru matematyczno-przyrodniczego (w porównaniu z siłą związku z wynikami egzaminu). Siła zależności między inteligencją a przyrostami umiejętności dla obszaru humanistycznego jest czterokrotnie słabsza, natomiast dla części matematyczno-przyrodniczej – dwukrotnie. Efekt ten pozostaje jednak znaczący: 4 punkty na jedno odchylenie standardowe inteligencji. Sugeruje to, że wskaźniki EWD nie do końca są niezależne od

tych cech. Dla badania trafności metody EWD ważniejsze jednak od wyników analiz na poziomie ucznia jest zróżnicowanie międzyszkolne. Okazuje się, że gdy kontrolujemy uprzednie osiągnięcia szkolne, zmienne statusowe i inteligencja mają znaczenie dla przyrostu umiejętności w analizach wewnątrz szkół, w znacznie mniejszym zaś stopniu rzutują na międzyszkolną zmienność wyników egzaminacyjnych.

Oszacowane z wykorzystaniem informacji o wykształceniu rodziców uczniów wskaźniki EWD dla szkół są niemal identyczne, jak te wyliczane na podstawie stosowanych obecnie modeli (korelacja na poziomie 0,98–0,99). Oznacza to, że brak w modelu EWD zmiennych statusowych w bardzo niewielkim stopniu obciąża oszacowania wskaźników, a tym samym ma marginalne znaczenie dla trafności metody. Podobny rezultat uzyskano dla inteligencji w obszarze humanistycznym. Trochę niższą korelację uzyskano natomiast dla obszaru matematyczno-przyrodniczego: 0,96. Oznacza to, że w skrajnych sytuacjach notujemy w tym wypadku pewne rozbieżności. Skalę problemu ilustruje poniższy wykres rozrzutu.

Podsumowując, brak uwzględnienia w obliczaniu wskaźników EWD informacji o statusie społecznym rodziny i inteligencji ucznia nie zagraża znacząco trafności metody EWD. Jedynie w wypadku egzaminu w części matematyczno-przyrodniczej brak w modelu EWD informacji o inteligencji ucznia może nieznacznie obciążać szacunki.

Rysunek 4.26. Zależność między wskaźnikami EWD dla szkół wyliczonymi bez uwzględnienia informacji o inteligencji uczniów oraz z uwzględnieniem tej informacji



Gdy myślimy o EWD jako mierze efektywności nauczania w szkole, w sposób naturalny pojawia się pytanie, czy jej wartość zależy od tego, czy uczniowie korzystają z korepetycji. Korepetycje bowiem służą lepszemu opanowaniu materiału, jednak nie stanowią wkładu szkoły w osiągnięcia ucznia. Jednak by korepetycje miały znaczenie dla EWD, musi z nich korzystać znacząca liczba dzieci. Przeprowadzone badanie pokazało, że jest to prawdą w przypadku przedmiotów matematyczno-przyrodniczych – korzystanie z korepetycji z tych przedmiotów choć raz w ciągu 3 lat nauki w gimnazjum

zadeklarowało 22% gimnazjalistów. Co istotne, odsetek uczniów korzystających z korepetycji z przedmiotów matematyczno-przyrodniczych różnił się znacząco między szkołami. W niektórych placówkach uczniowie w ogóle nie korzystali z korepetycji, w innych robiło to ponad 50% z nich. Jeżeli EWD zależy od korepetycji, pierwsze z nich mogą znajdować się w niekorzystnej sytuacji, gdyż osiągnięcia szkolne ich uczniów nie otrzymują dodatkowego „zastrzyku” w postaci płatnych zajęć. Korzystanie z korepetycji z przedmiotów humanistycznych zadeklarowało niespełna 5% gimnazjalistów, skala zjawiska jest więc bardzo mała.

Przyjrzyjmy się zatem, jak korzystanie z korepetycji z przedmiotów matematyczno-przyrodniczych powiązane jest z EWD. Pamiętajmy jednak, że EWD jest cechą szkoły. Konieczna jest zatem weryfikacja zależności między odsetkiem uczniów korzystających z korepetycji z przedmiotów matematyczno-przyrodniczych a matematyczno-przyrodniczym EWD szkoły. Korelacja ta jest słaba, dodatnia i wynosi 0,19. Wiadomo jednak, że z korepetycji częściej korzystają uczniowie z rodzin o wyższym statusie społecznym. W związku z tym odsetek korzystających z korepetycji może nieść informację o statusie społecznym uczniów. Gdy uwzględnimy to w analizie, korelacja między odsetkiem uczniów pobierających korepetycje z przedmiotów matematyczno-przyrodniczych a EWD szkoły spada do zera.

W wielu badaniach wskazuje się, że czynniki lokalne wpływają na poziom umiejętności uczniów. W projekcie EWD zbadaliśmy, czy i jak poszczególne cechy otoczenia szkoły powiązane są z wynikami egzaminacyjnymi i czy wskaźniki EWD są wrażliwe na wpływ tych czynników. W analizach uwzględniono takie czynniki charakteryzujące otoczenie społeczno-gospodarcze szkoły jak: poziom bezrobocia, wielkość gminy, nowoczesność gospodarki (ilość firm high-tech z siedzibą w danej gminie w stosunku do liczby mieszkańców), powszechność opieki przedszkolnej (odsetek dzieci w wieku 3–5 lat z terenu gminy uczęszczających do przedszkola), wydatki na oświatę, zasoby kulturowe (liczba bibliotek publicznych na tysiąc mieszkańców gminy) oraz zamożność (dochody mieszkańców gminy oraz zakres i skala pomocy społecznej udzielanej mieszkańcom). Informacje te pozyskano z Banku Danych Lokalnych Głównego Urzędu Statystycznego. Dotyczyły one gmin, w których zlokalizowane są szkoły, i dotyczyły roku 2010, czyli środkowego roku dla cyklu kształcenia badanej grupy gimnazjalistów (początek nauki w gimnazjum: 2009, koniec: 2012)

Spośród analizowanych charakterystyk gmin sześć istotnie statystycznie wiąże się z wynikami nauczania: wielkość gminy, gospodarka nowoczesnych technologii, powszechność opieki przedszkolnej, poziom bezrobocia, wydatki na oświatę oraz zasoby kulturowe. Trzy pierwsze okazały się skorelowane zarówno z umiejętnościami humanistycznymi, jak i matematyczno-przyrodniczymi. Wysokość gminnych wydatków na oświatę i poziom bezrobocia korelowały jedynie z poziomem umiejętności humanistycznych, natomiast zasoby kulturowe – z osiągnięciami matematyczno-przyrodniczymi. Wyniki analiz pokazały, że im bardziej nowoczesna jest lokalna gospodarka oraz im więcej dzieci objętych jest opieką przedszkolną, tym wyższe są osiągnięcia uczniów. Z kolei wyższy poziom bezrobocia związany był z obniżeniem tych osiągnięć. Znaczenie pozostałych czynników (wielkości gminy, jej wydatki na oświatę i zasoby kulturowe) okazało się statystycznie istotne, ale w praktyce nieznaczące (minimalna siła efektu).

Czy czynniki te powiązane są z efektywnością pracy szkoły? Okazało się, że ze wskaźnikami EWD, zarówno w zakresie przedmiotów humanistycznych, jak i matematyczno-przyrodniczych, istotnie powiązany jest jedynie poziom bezrobocia w gminie. Im większe bezrobocie w gminie, tym trudniej szkole uzyskać wysoką efektywność nauczania. Zależność ta jest jednak w porównaniu z zależnością z wynikami egzaminacyjnymi znacznie słabsza. Oznacza to, że modele statystyczne EWD dla gimnazjów pozwalają częściowo kontrolować ten efekt. Dodatkowo dla efektywności pracy szkoły w zakresie umiejętności matematyczno-przyrodniczych znaczenie ma także wielkość gminy oraz usytuowanie w gminie przedsiębiorstw zawansowanych technologii. Oba te czynniki również znacznie słabiej oddziałują na efektywność pracy szkoły niż na same osiągnięcia uczniów. Podsumowanie wyników analiz przedstawiono w poniższej tabeli.

Tabela 4.7. Powiązanie czynników społeczno-gospodarczych z osiągnięciami uczniów i efektywnością nauczania (EWD) w zakresie przedmiotów humanistycznych oraz matematyczno-przyrodniczych

Czynnik	Humanistyczne		Mat-przyr.	
	osiągnięcia	EWD	osiągnięcia	EWD
poziom bezrobocie	-	-	x	-
wielkość gminy	-	x	-	-
przedsiębiorstwa zaawansowanych technologii	+	x	+	+
powszechność opieki przedszkolnej	+	x	+	x
wydatki na oświatę	-	x	x	x
zasoby kulturowe	x	x	-	x
zamożność gminy	x	x	x	x

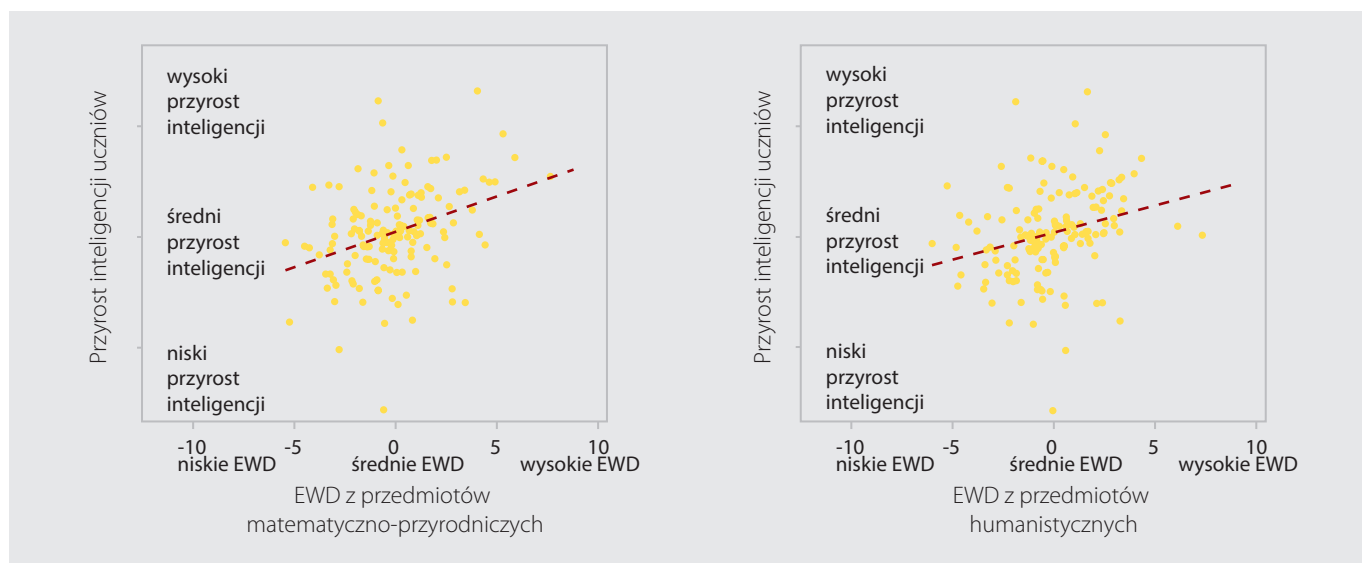
+ związek pozytywny, - związek negatywny, x brak związku

Okazuje się zatem, że choć niektóre cechy społeczno-gospodarczego otoczenia gimnazjum mają znaczenie dla wyników uzyskiwanych przez uczniów na egzaminach, to zdecydowanie mniej czynników tego typu związanych jest z efektywnością pracy szkoły, mierzoną wskaźnikami EWD. Można więc powiedzieć, że miary EWD są zdecydowanie mniej wrażliwe na wpływ środowiska społeczno-gospodarczego, w którym działa szkoła.

Krytycy EWD zarzucają niekiedy, iż wysoka wartość wskaźnika jest dowodem testomanii panującej w szkole i świadczy o silnym nacisku na przygotowanie uczniów do egzaminów, kosztem ich ogólnego rozwoju intelektualnego. Gdyby teza ta była prawdziwa, w szkołach o wysokim EWD obserwowalibyśmy wolniejszy rozwój poznawczy uczniów. Czy rzeczywiście ma to miejsce? W celu weryfikacji tego zarzutu sprawdziliśmy zależność między EWD szkoły a przyrostem inteligencji uczniów w toku nauki w gimnazjum. By możliwe było obliczenie średnich przyrostów inteligencji uczniów w każdej badanej szkole, mierzono ją dwukrotnie – w pierwszej i trzeciej klasie.

Przeprowadzone analizy pokazały, że szkoły rzeczywiście różnią się między sobą średnią wielkością przyrostów inteligencji swoich uczniów (podział na szkoły wyjaśnia ok. 15% tego zróżnicowania). Jednak zależność okazała się odwrotna od przewidywanej przez krytyków EWD. Zaobserwowano statystycznie istotną pozytywną korelację między EWD szkoły a wielkością średnich przyrostów inteligencji jej uczniów. Im wyższa wartość EWD szkoły, tym większe przyrosty inteligencji. Związek okazał się silniejszy w przypadku EWD w obszarze matematyczno-przyrodniczym (korelacja wynosi 0,32) niż humanistycznym (korelacja równa 0,24). Zależność tę zobrazowano na poniższym wykresie. Kropki reprezentują badane szkoły, niebieską linią pokazano najlepsze przybliżenie dla zaobserwowanego związku.

Rysunek 4.27. Związek EWD z przyrostami inteligencji w szkołach



Wniosek płynący z analiz zaprzecza więc wysuwanemu pod adresem EWD zarzutowi. W szkołach o wysokiej wartości wskaźników uczniowie robią nie tylko większy niż przeciętnie postęp w zakresie przyswajania wiadomości i umiejętności szkolnych, ale także szybciej rozwijają się poznawczo. Wynik ten stanowi ważki argument na rzecz tezy o trafności wskaźników.

Przeprowadzone analizy dostarczają wiele argumentów na rzecz trafności wskaźników EWD, ale wskazują też na słabsze ich strony:

- Egzaminy zewnętrzne wykorzystane do konstrukcji gimnazjalnych wskaźników EWD mają wystarczająco dobre właściwości, by mogły stanowić podstawę wyliczania trafnej miary EWD.
- Brak uwzględnienia w obliczaniu wskaźników EWD dla szkół informacji o statusie społecznym rodziny i inteligencji ucznia nie zagraża znacząco trafności metody EWD. Jednak w wypadku egzaminu w części matematyczno-przyrodniczej brak w modelu EWD informacji o inteligencji ucznia może nieznacznie obciążać szacunki.
- Gimnazja, w których uczniowie uczęszczają na korepetycje, nie mają wyższego EWD. Korepetycje nie mają znaczenia dla EWD jako miary na poziomie szkoły.
- Choć wiele cech społeczno-gospodarczego otoczenia gimnazjum ma znaczenie dla wyników uzyskiwanych przez uczniów na egzaminach, to zdecydowanie mniej czynników tego typu związanych jest z efektywnością pracy szkoły mierzoną wskaźnikami EWD.
- Powiązanie wskaźników EWD z przyrostami inteligencji dowodzi, że miara EWD nie wskazuje jedynie na to, na ile dobrze szkoła radzi sobie z przygotowaniem uczniów do egzaminów zewnętrznych. W szkołach o wysokim EWD szybciej przebiega ogólny rozwój poznawczy uczniów.

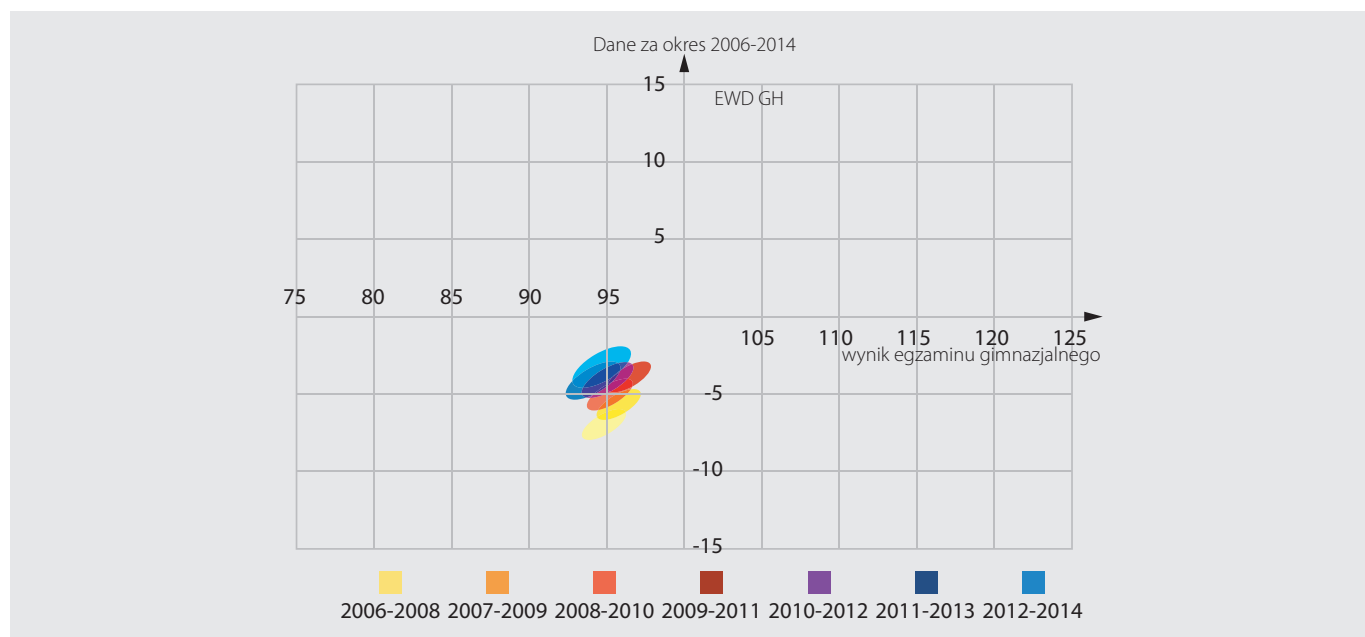
Możliwości wykorzystania przez gimnazja trzyletnich wskaźników EWD

Trzyletnie wskaźniki egzaminacyjne dają ogólny obraz wyników i efektywności nauczania w szkole. Ich przydatność dla szkół jest związana przede wszystkim z możliwością śledzenia zmian w czasie i porównywania swojej szkoły z innymi. Śledzenie zmian w czasie pozwala z większą pewnością wnioskować o zależnościach przyczynowo-skutkowych: o ile w otoczeniu szkoły nie zaszły jakieś znaczące zmiany, to zmiany w czasie wskaźników EWD można przypisać procesom zachodzącym w szkole. Możliwości wykorzystania przez szkoły trzyletnich miar EWD zilustrujemy kilkoma przykładami.

Gimnazjum A osiąga w analizowanym okresie 2006–2014 średnie wyniki egzaminacyjne niższe od średniej w kraju o około 1/3 odchylenia standardowego. Natomiast w zakresie EWD obserwujemy dość systematyczny trend wzrostowy. Oznacza to, że szkoła rekrutuje do klas pierwszych uczniów o coraz niższych uprzednich osiągnięciach, ale pracując z nimi coraz bardziej efektywnie, i uzyskuje

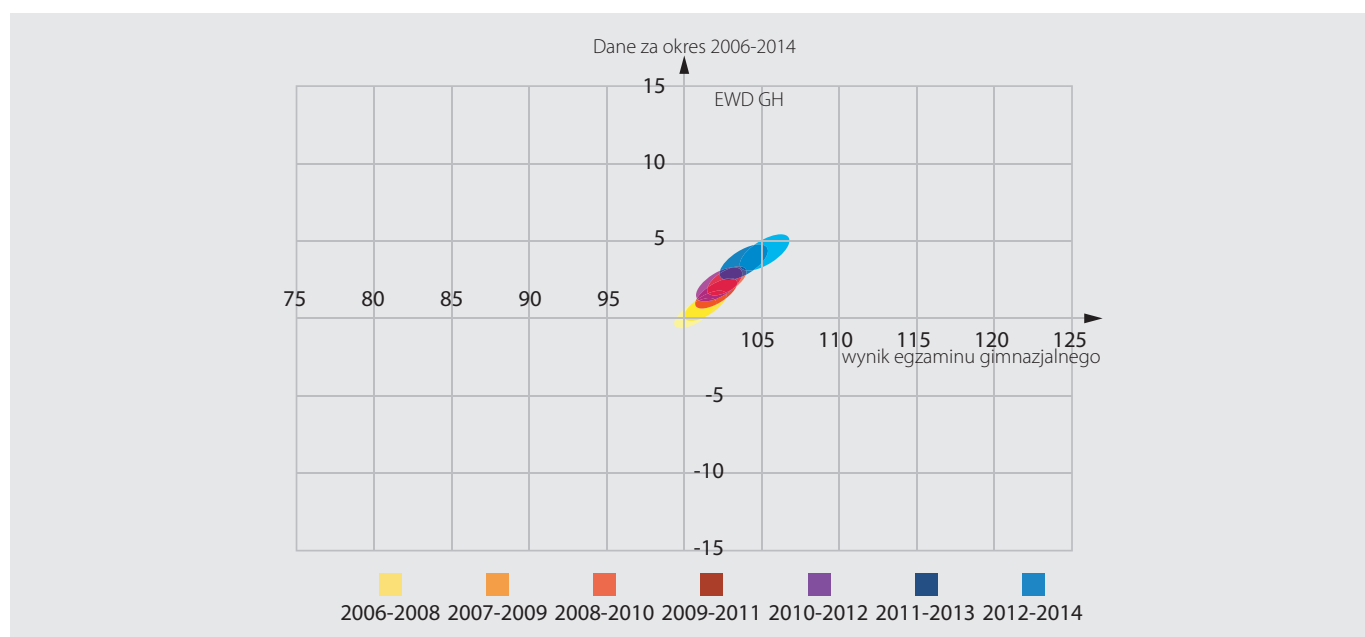
stabilne w czasie wyniki końcowe. Utrzymanie wyniku egzaminacyjnego na stałym poziomie możliwe jest tylko dzięki zwiększaniu efektywności nauczania, która wyraża się w rosnącym wskaźniku EWD. Ponadto rosnąca wielkość elipsy pokazuje, że szkoła rekrutuje do klas pierwszych coraz mniej uczniów: w latach 2006–2008 do egzaminu gimnazjalnego przystąpiło 536 uczniów, a w ostatnim okresie trzyletnim 2012–2014 już tylko 397 uczniów.

Rysunek 4.28. Zmiany trzyletnich egzaminacyjnych wskaźników dla gimnazjum A



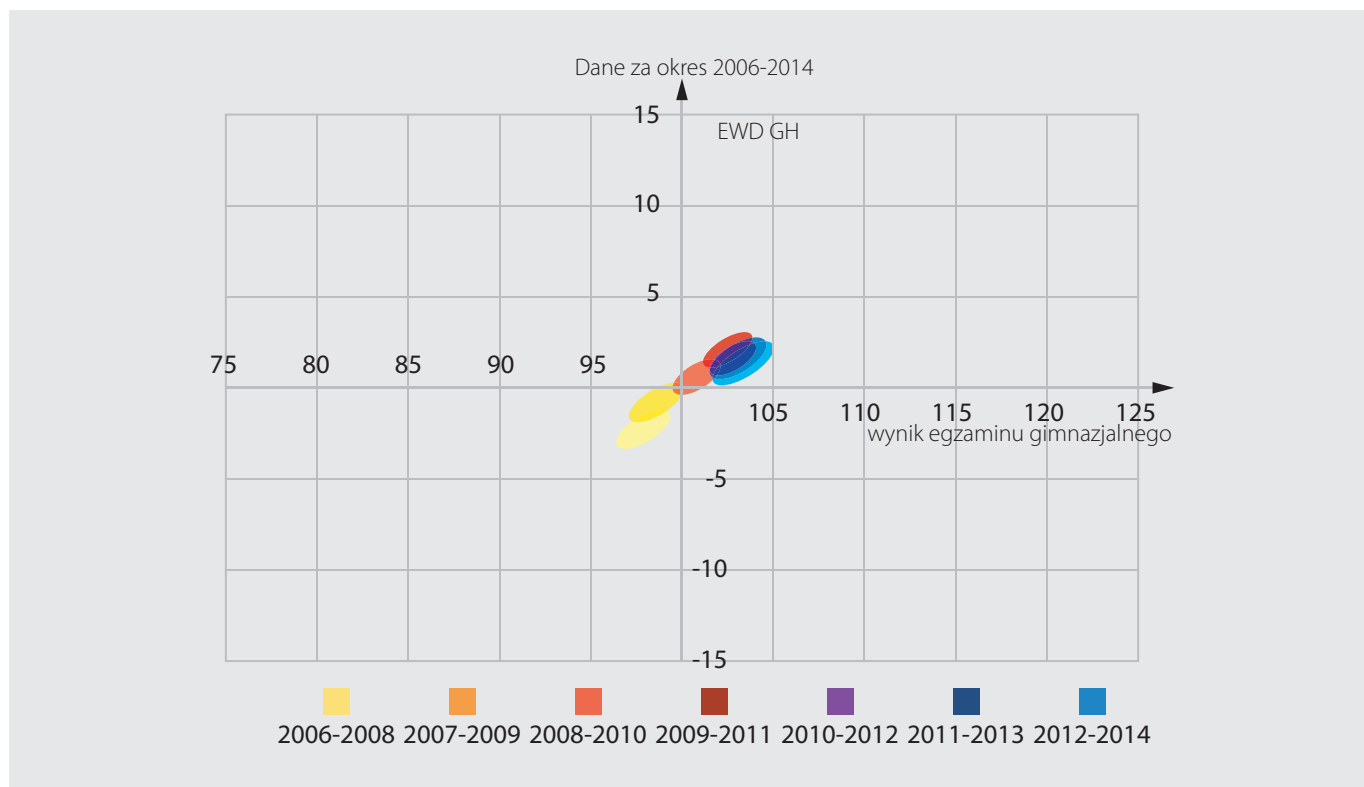
Gimnazjum B osiąga coraz wyższe średnie wyniki egzaminacyjne, które w ostatnim okresie trzyletnim (2012–2014) były wyższe od średnich wyników w kraju o $1/3$ odchylenia standardowego. Jest to efektem stale rosnącej efektywności nauczania przedmiotów humanistycznych mierzonej wskaźnikiem EWD.

Rysunek 4.29. Zmiany trzyletnich wskaźników egzaminacyjnych dla gimnazjum B



Gimnazjum C osiąga ponadprzeciętne wyniki egzaminacyjne w części humanistycznej egzaminu gimnazjalnego, przy dodatnim wskaźniku EWD. W latach 2006–2011 obserwujemy wyraźny trend wzrostowy, od 2011 roku wyniki i efektywność stabilizują się na poziomie trochę przekraczającym średnią krajową.

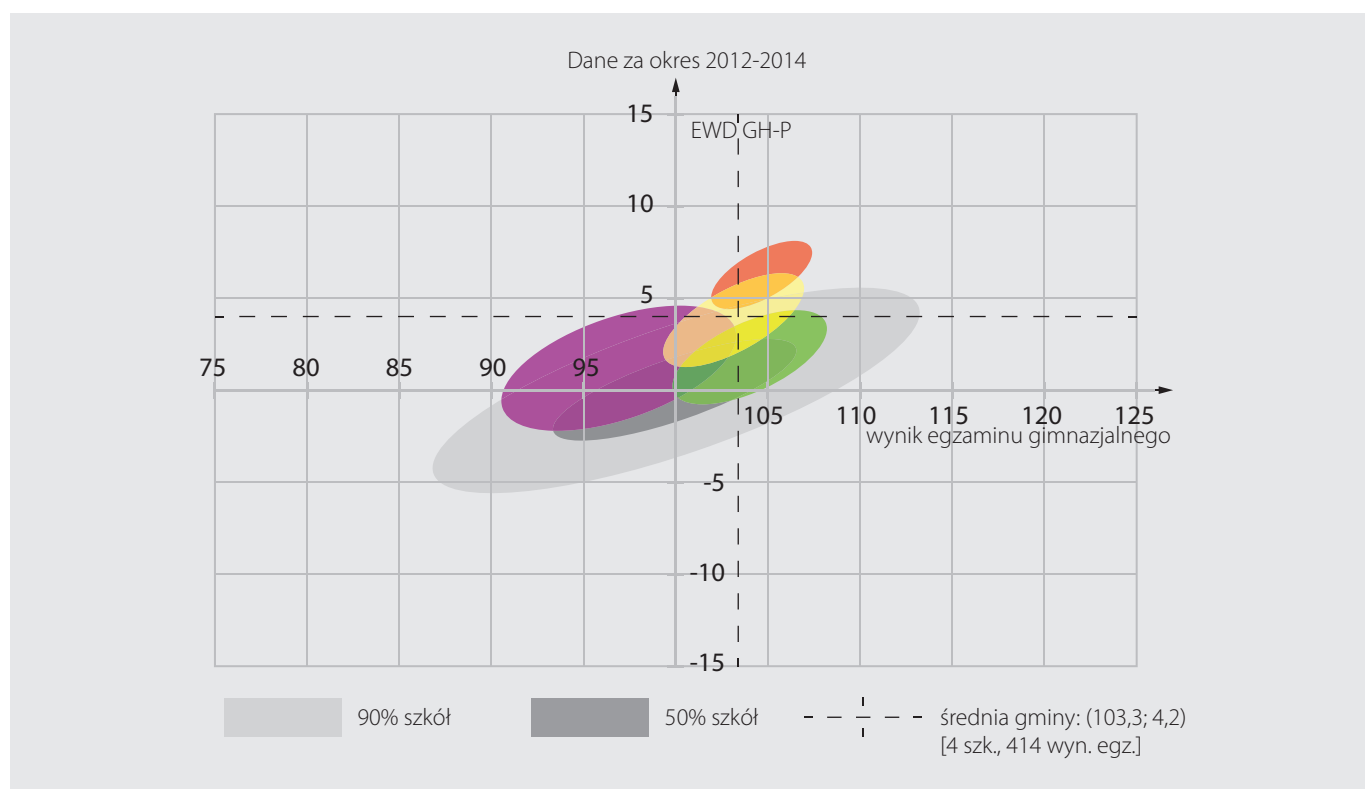
Rysunek 4.30. Zmiany trzyletnich wskaźników egzaminacyjnych dla gimnazjum C



Wskaźniki EWD są miarami względnymi. Ważne jest, aby móc określić pozycję szkoły nie tylko względem krajowego, ale także względem lokalnych układów odniesienia, takich jak gmina, powiat czy województwo. Dzięki możliwościom strony internetowej, na której prezentowane są trzyletnie wskaźniki EWD, można interaktywnie wybrać potrzebny w danej analizie punkt widzenia.

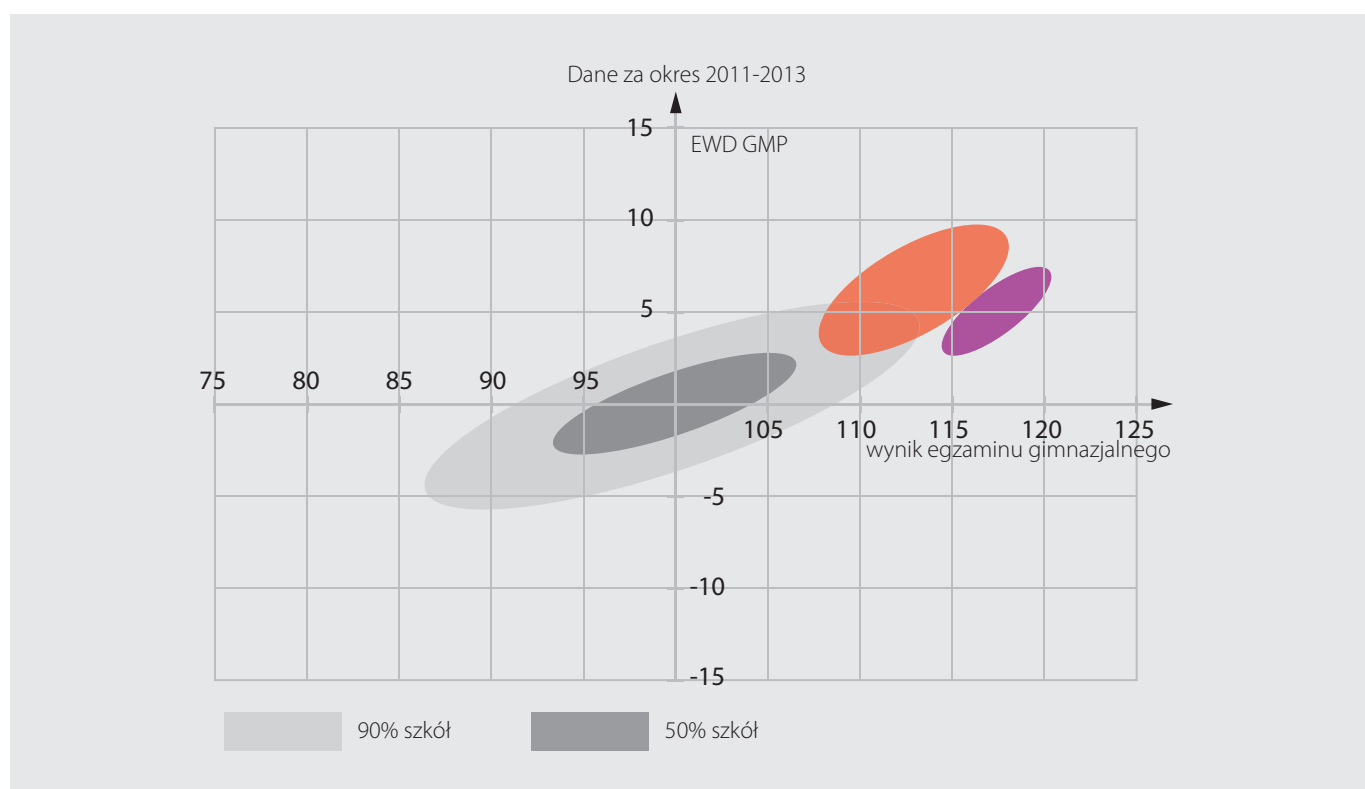
Gmina Z. Lokalny układ odniesienia dla gminy Z dla języka polskiego (zaznaczony na poniższym rysunku przerywaną linią) jest wyraźnie przesunięty w kierunku wyższych wyników i ponadprzeciętnego EWD. W tej gminie większość uczniów uzyskuje wyniki powyżej średniej krajowej, jednak różnie się to rozkłada między szkołami. Gimnazjum, którego pozycję w układzie współrzędnych opisuje fioletowa elipsa, osiąga na tle kraju przeciętne wyniki i EWD, natomiast w gminnym układzie odniesienia jego wyniki i efektywność nauczania języka polskiego mierzona EWD są statystycznie niższe od gimnazjum, którego pozycję określa pomarańczowa elipsa.

Rysunek 4.31. Lokalny układ odniesienia dla gminy Z



Gimnazja żeńskie. Nie zawsze lokalny punkt odniesienia jest najlepszy do porównań. Jeśli znamy szkoły pracujące w podobnych warunkach i chcemy się z nimi właśnie porównywać, warto zobaczyć swoją szkołę na tle tych właśnie placówek.

Rysunek 4.32. Porównanie dwóch żeńskich gimnazjów.



Na powyższym wykresie pokazano dwie żeńskie szkoły o podobnej efektywności nauczania i porównywalnych wynikach końcowych.

Możliwości wykorzystania przez gimnazja jednorocznych wskaźników EWD

Trzyletnie wskaźniki EWD wyliczane są dla całych szkół. Ten poziom analizy jest niewystarczający do zrozumienia procesów zachodzących w szkole. Do tego celu potrzebne jest narzędzie, które wykaże zmiany w krótszych niż trzyletnie okresach czasu, pozwoli wyliczyć wskaźniki EWD dla mniejszych i specyficznych dla szkoły grup uczniów, umożliwi porównanie między latami nie tylko efektów pracy szkoły jako całości, ale również efektywności nauczania w wybranych grupach uczniowskich. W dużych szkołach, o zwykle dużym zróżnicowaniu między oddziałami klasowymi (m.in. pod względem składu uczniowskiego, uczących nauczycieli lub profilu klas – klasy integracyjne, klasy sportowe) do opisu pracy szkoły lepiej nadają się jednoroczne wskaźniki.

W analizach wewnątrzszkolnych standardowo oblicza się jednoroczne wskaźniki dla oddziałów, analizuje się efekt płci, efektywność nauczania w grupach o różnych uprzednich osiągnięciach. Jeśli dysponujemy dodatkowymi informacjami o uczniach, to można je wykorzystać jako kryteria podziału na specyficzne grupy uczniowskie. Można np. analizować wskaźniki EWD w grupie uczniów korzystających z dodatkowych zajęć, uczniów dojeżdżających do szkoły, uczniów-absolwentów różnych szkół podstawowych. Ten katalog można dowolnie poszerzać i będzie on inny dla każdej szkoły. Nie chodzi bowiem jedynie o samo wyznaczenie wskaźnika EWD – ważniejsze jest wyodrębnienie grup uczniowskich o istotnie różnej efektywności nauczania mierzonej EWD. Próba zrozumienia, dlaczego te grupy różnią się efektywnością nauczania, powinna poszerzyć wiedzę o czynnikach efektywności specyficznych dla danej szkoły.

Analizy z wykorzystaniem jednorocznych wskaźników EWD są możliwe w Kalkulatorze EWD 100. Do kalkulatora zostały zaimplementowane modele jednorocznej EWD uwzględniające uczniów drugorocznych. W Kalkulatorze EWD 100 można od rocznika 2012 wyliczyć wskaźniki odpowiadające czterem arkuszom gimnazjalnym: GH-P, GH-H, GM-M, GM-P oraz syntetyczne wskaźniki dla całego obszaru humanistycznego (GH) i całego obszaru matematyczno-przyrodniczego (GMP). Dla wcześniejszych lat Kalkulator EWD oblicza wskaźniki dla części humanistycznej (GH) oraz matematyczno-przyrodniczej (GMP). Syntetyczne wskaźniki GH i GMP pozwalają porównywać wyniki od 2012 roku z wcześniejszymi rocznikami.

W Kalkulatorze EWD można w prosty sposób wykonywać różne analizy statystyczne wyników egzaminacyjnych. Można zrobić:

- **wykres rozrzutu** – pozwala na jednym wykresie zobaczyć wynik ucznia na egzaminie z poprzedniego etapu kształcenia i na egzaminie końcowym oraz krzywą przewidywanego wyniku;
- **wykres prezentujący wskaźniki EWD wraz z przedziałami ufności** – pozwala on porównać ze względu na EWD różne grupy uczniów i śledzić zmiany EWD w czasie;
- **wykres prezentujący średnie wyników egzaminacyjnych wraz z przedziałami ufności** – pozwala on porównać różne grupy uczniów i śledzić zmiany w czasie;
- **wykres procentowy skumulowany dla uprzednich osiągnięć szkolnych** – pozwala porównywać między sobą grupy uczniów ze względu na uprzednie osiągnięcia;
- **rozkład wyników przewidywanych i rzeczywiście uzyskanych** – wykres jest jednym ze sposobów graficznej prezentacji wskaźników EWD;
- **rozkłady reszt** – pozwala porównać zróżnicowanie reszt w szkole z rozkładem krajowym.

Uzupełnieniem powyższych narzędzi graficznych jest możliwość sporządzania tabel, które mogą zawierać wiele wskaźników liczbowych, dotyczących zarówno EWD, jak i wyników egzaminacyjnych. Kalkulator EWD 100 można pobrać bezpłatnie ze strony <http://ewd.edu.pl/kewd100/pobierz>

Analizę wyników egzaminacyjnych wzbogacając o wskaźniki EWD można wykorzystać w szkole zarówno na etapie opisu (jak jest?), jak i na etapie identyfikowania obszarów do zmiany (jak powinno być?) i/lub określenia przedmiotu ewaluacji wewnętrznej. W szkole podejmuje się różnorodne działania, które mają sprzyjać poprawie efektywności nauczania – naturalnym sposobem weryfikacji

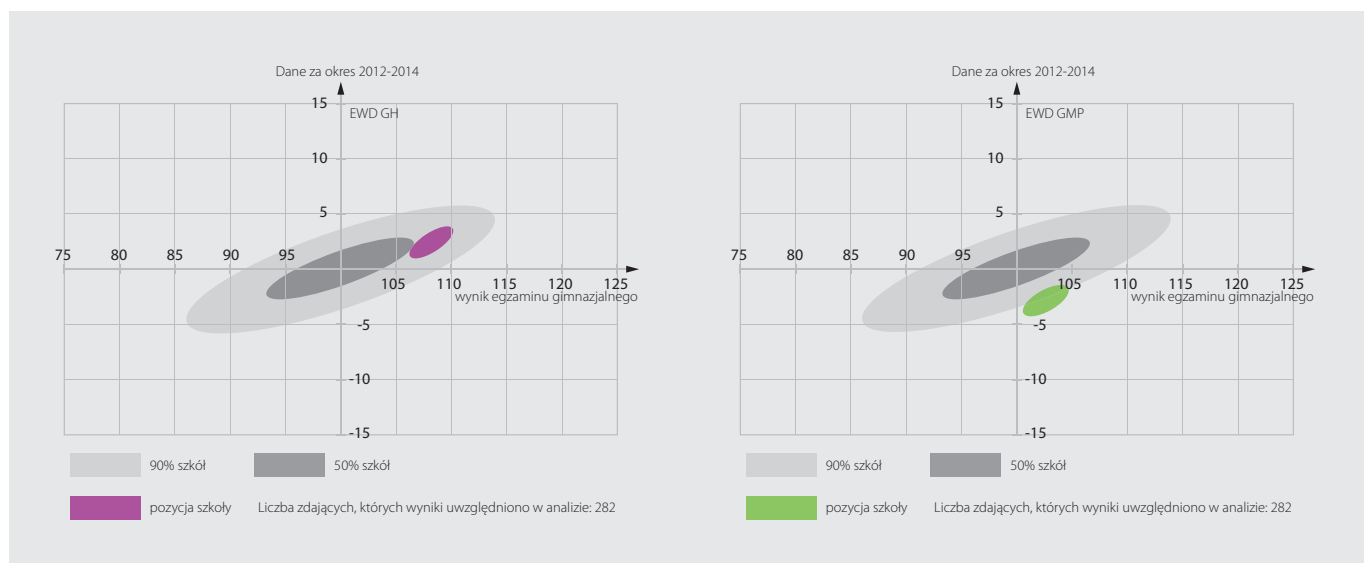
skuteczności tych działań jest analiza wskaźników EWD. W tym przypadku należy pamiętać o tym, że efekty podejmowanych działań rzadko kiedy będą widoczne w krótkim okresie czasu.

Można dokonywać zarówno analiz w czasie, jak i analiz przekrojowych w podziale na różne grupy uczniowskie. W analizach poszukujemy takich grup, które statystycznie różnią się wskaźnikami EWD, bo pozwoli to poszukiwać czynników wpływających na efektywność nauczania. Jeśli kryterium podziału jest czas, stawiamy pytanie: Co spowodowało wzrost lub spadek wskaźników EWD w czasie? Jeśli kryterium podziału jest grupa uczniowska, to stawiamy pytanie: Dlaczego dany podział generuje różnicę we wskaźnikach EWD?

Prześledźmy możliwości tkwiące w analizach EWD na przykładzie dwóch gimnazjów: Gimnazjum SZ oraz Gimnazjum P. Więcej przykładów zastosowania wskaźników EWD do analiz wewnątrzszkolnych można znaleźć w materiałach pomocniczych przygotowywanych dla dyrektorów i nauczycieli gimnazjów (Stożek, 2008, 2010).

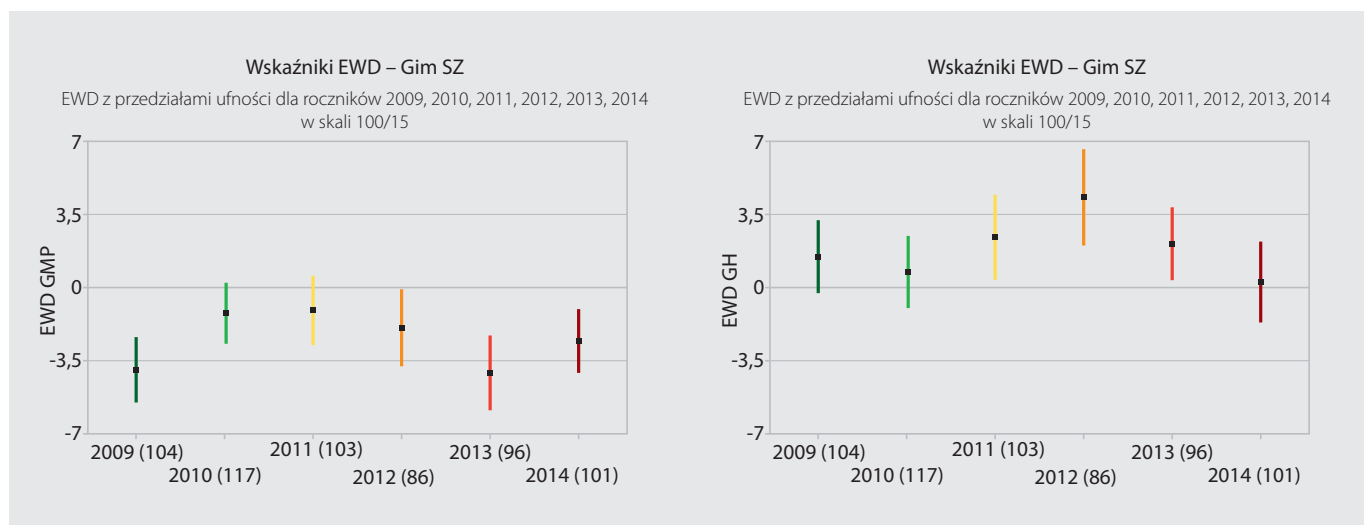
Gimnazjum SZ. Uczniowie gimnazjum SZ uzyskują na egzaminach gimnazjalnych ponadprzeciętne wyniki: w trzyletnim okresie 2012–2014 w części humanistycznej są one o ok. 1/2 odchylenia standardowego wyższe od średniej krajowej, natomiast w części matematyczno-przyrodniczej przekraczają średnią o ok. 1/5 odchylenia standardowego. Jak pokazują poniższe wykresy EWD w obszarze humanistycznym jest dodatnie, ale w zakresie przedmiotów matematyczno-przyrodniczych jest ujemne. Znacząca różnica we wskaźnikach EWD wskazuje na niską efektywność nauczania w zakresie przedmiotów matematyczno-przyrodniczych i powinna być powodem do niepokoju.

Rysunek 4.33. Trzyletnie wskaźniki egzaminacyjne dla gimnazjum SZ za lata 2012–2014



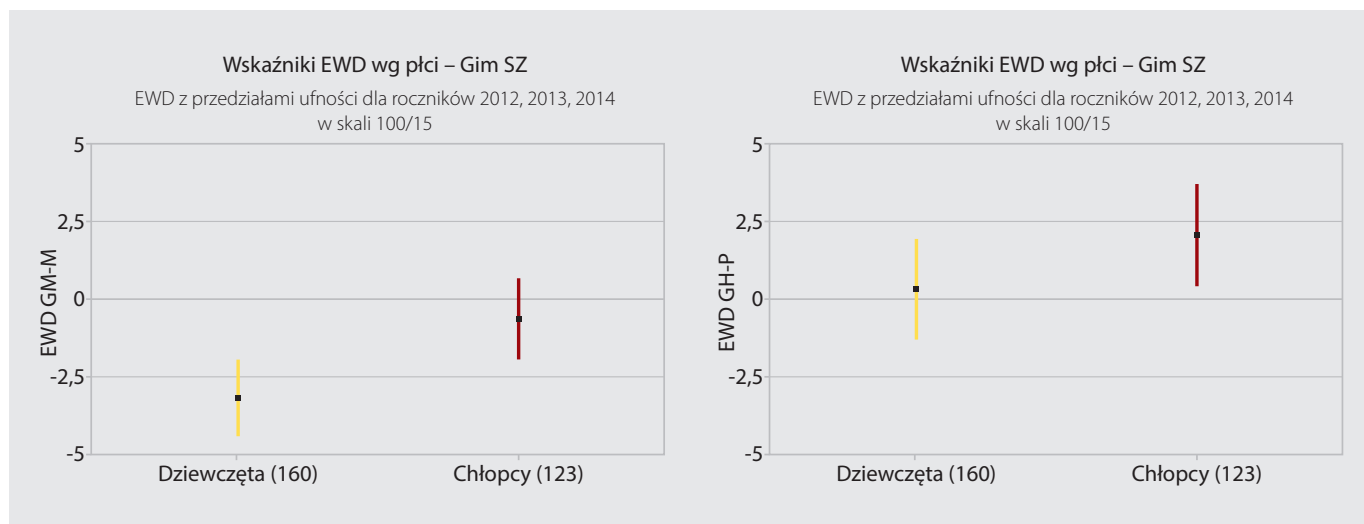
Bardziej szczegółowy obraz można uzyskać dzięki analizom wskaźników jednorocznych wykonanym za pomocą Kalkulatora EWD. Znaczącą (i utrzymującą się w czasie) różnicę w efektywności nauczania przedmiotów matematyczno-przyrodniczych i przedmiotów humanistycznych potwierdza analiza jednorocznych wskaźników EWD. Warto zatem przemyśleć w szkole czynniki, które mogą mieć na to wpływ.

Rysunek 4.34. Jednoroczne wskaźniki EWD za lata 2009–2014 dla gimnazjum SZ

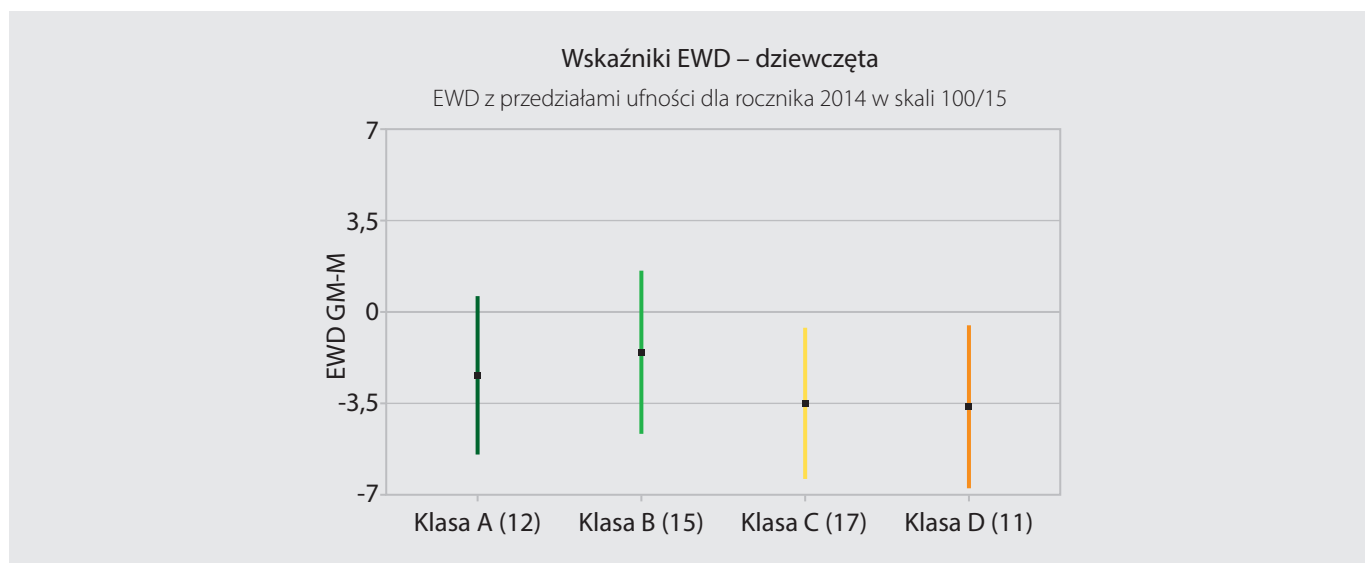


Kalkulator EWD zasadniczo przeznaczony jest do analizowania wskaźników jednorocznych, ale można obliczać wskaźniki z kilku lat. Tak zbudowany wskaźnik EWD dla matematyki za lata 2012–2014 pokazuje znaczącą statystycznie różnicę między efektywnością nauczania dziewcząt i chłopców, natomiast w przypadku wskaźnika EWD dla języka polskiego różnica ta nie jest istotna statystycznie. Efektywność nauczania matematyki w przypadku dziewcząt jest niska, znacząco poniżej średniej krajowej. Dla chłopców natomiast efektywność nauczania matematyki jest bliska przeciętnej.

Rysunek 4.35. Wskaźniki EWD dla dziewcząt (K) i chłopców (M) w gimnazjum SZ za lata 2012–2014



Wróćmy do danych tylko z 2014 roku. Poniższy wykres pokazuje EWD matematyczne dla dziewcząt, ale w rozbiciu na oddziały.

Rysunek 4.36. Jednoroczne wskaźniki EWD dla dziewcząt w podziale na oddziały klasowe w gimnazjum SZ w 2014 roku.

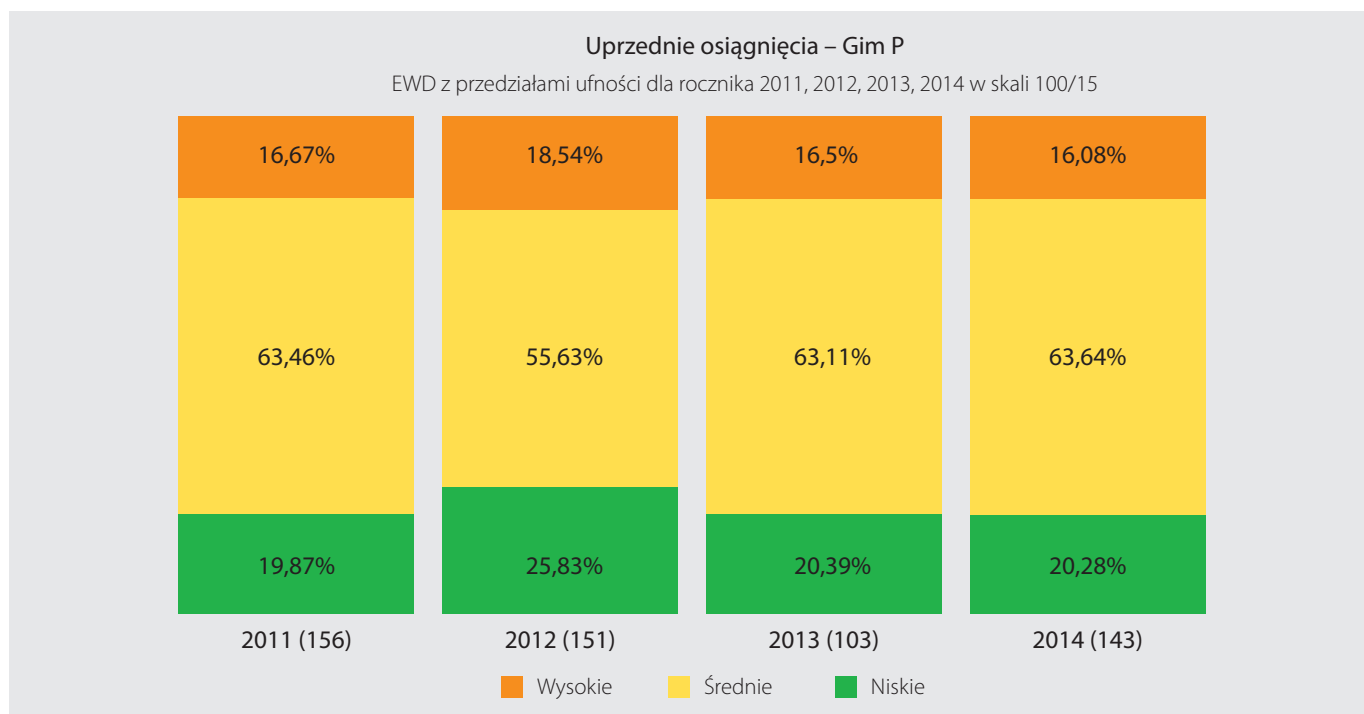
W roku 2014 w żadnym z oddziałów klasowych nie udało się osiągnąć z dziewczętami ponadprzeciętnej efektywności nauczania z matematyki. Problem niskiej efektywności nauczania matematyki w grupie dziewcząt jest trwałą charakterystyką tej szkoły i warto go uczynić przedmiotem ewaluacji wewnętrznej w szkole.

Gimnazjum P. W Gimnazjum P przed czterema laty wdrożono nowy model nauczania oparty na nauczaniu przez doświadczanie. Metodę zaczęli stosować przede wszystkim nauczyciele przedmiotów matematyczno-przyrodniczych. Metoda wpływa na zwiększone zaangażowanie uczniów w proces uczenia się, wzmacnia wiarę we własne możliwości uczniów, sprzyja samodzielnej pracy. W 2014 roku do egzaminu gimnazjalnego przystąpili uczniowie, którzy przez 3 lata nauki w gimnazjum byli nauczani z zastosowaniem tego modelu nauczania. Sprawdźmy zatem, czy wskaźniki EWD pokażą wzrost efektywności nauczania w zakresie przedmiotów matematyczno-przyrodniczych w tej szkole.

Szkoła przez lata uzyskiwała niskie wyniki egzaminacyjne przy ujemnym wskaźniku EWD. Rekrutowała uczniów o przeciętnych uprzednich osiągnięciach, co pokazano na poniższym wykresie.

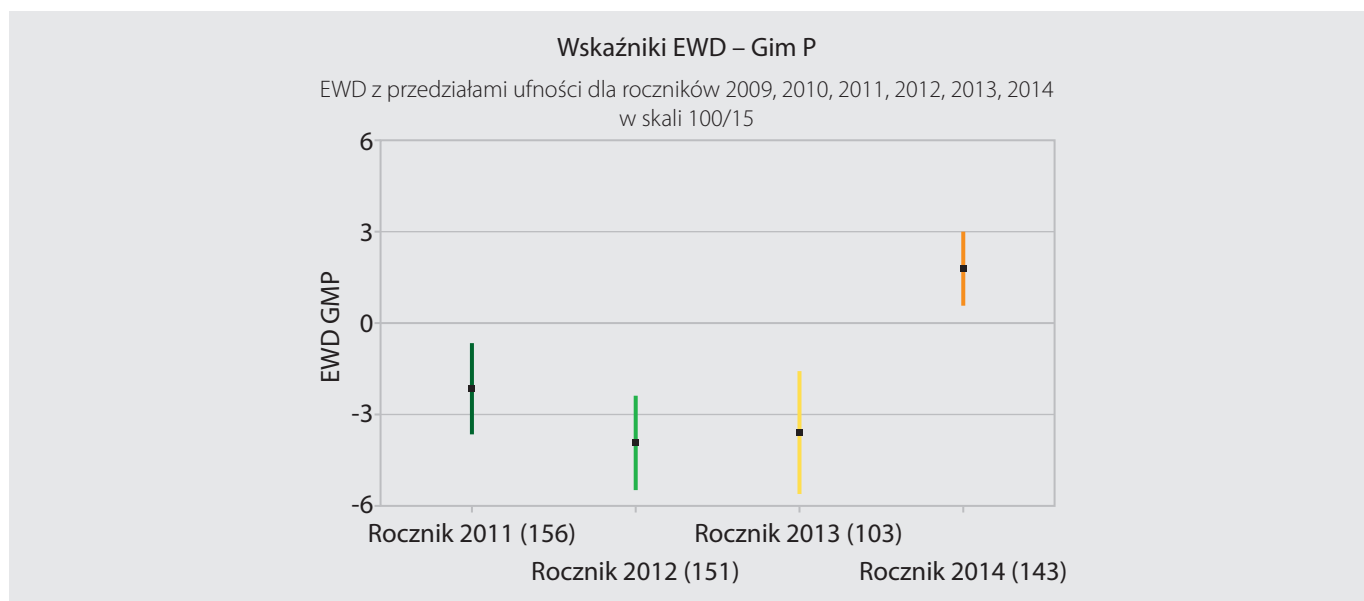
4. Metoda edukacyjnej wartości dodanej w Polsce

Rysunek 4.37. Struktura uczniów gimnazjum P wg wyników uzyskanych na sprawdzianie po szóstej klasie. W opisach podano rok przystąpienia do egzaminu gimnazjalnego, a w nawiasach liczbę uczniów. Niskie wyniki odpowiadają wynikom na sprawdzianie z 1., 2., 3. przedziału staninowego, wyniki wysokie – wynikom na sprawdzianie z 7., 8., 9 przedziału staninowego



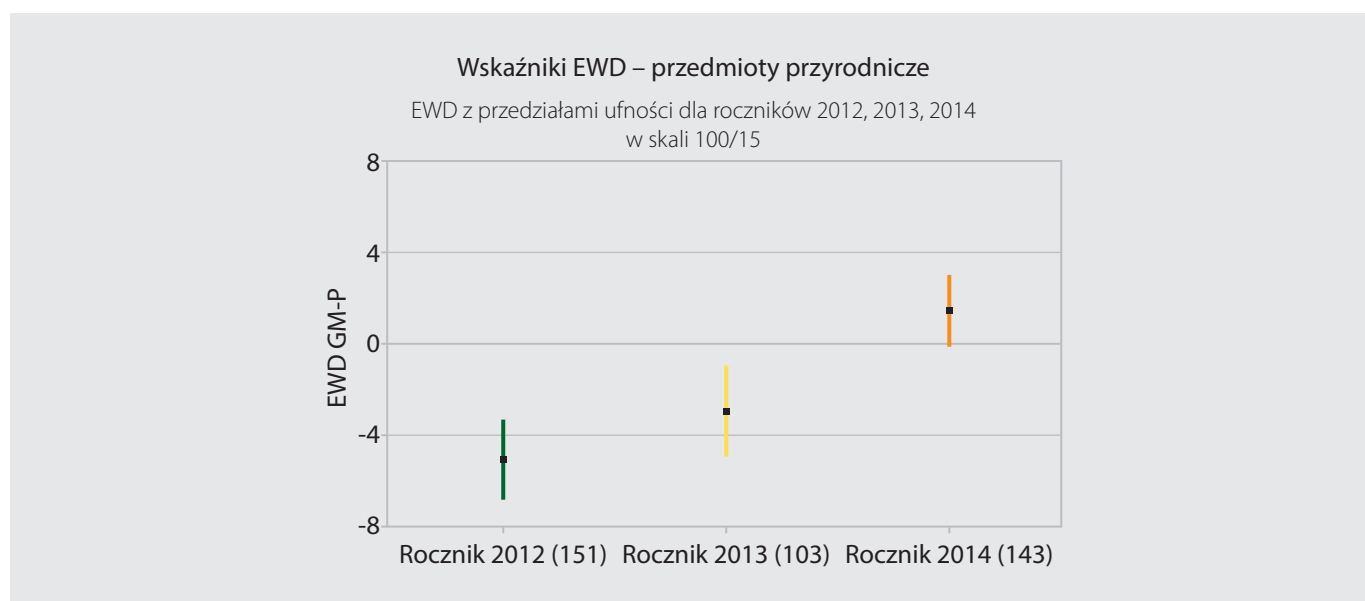
Roczniki, które w latach 2011–2014 przystępowały do egzaminu gimnazjalnego, były do siebie podobne ze względu na upřednie osiągnięcia szkolne. Około 20% uczniów to uczniowie o niskich wynikach na sprawdzianie szóstoklasisty, a 16–18% to uczniowie o wysokich wynikach na sprawdzianie.

Rysunek 4.38. Efektywność nauczania w zakresie przedmiotów matematyczno-przyrodniczych w latach 2011–2014.



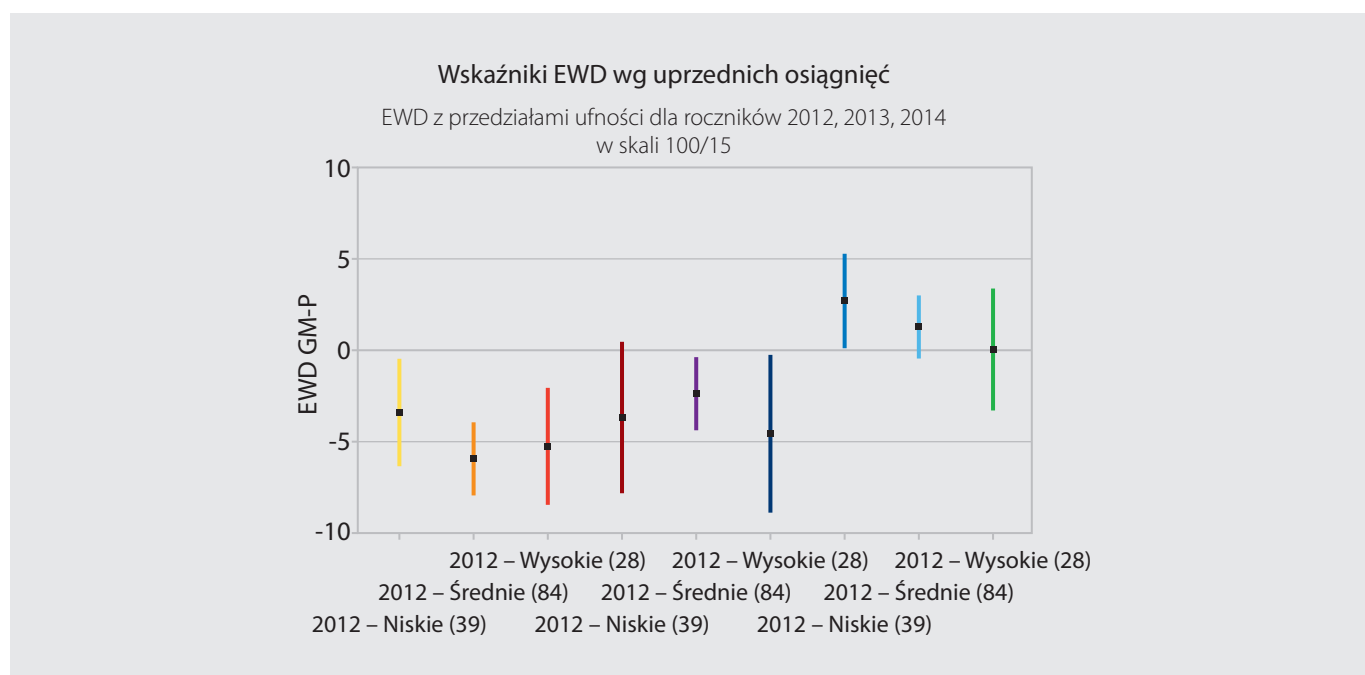
W latach 2011–2013 EWD była ujemna, w roku 2014 ponadprzeciętna.

W dalszych analizach ograniczymy się do roczników 2012, 2013 i 2014, co pozwoli nam rozpatrywać wskaźniki EWD tylko dla przedmiotów przyrodniczych.

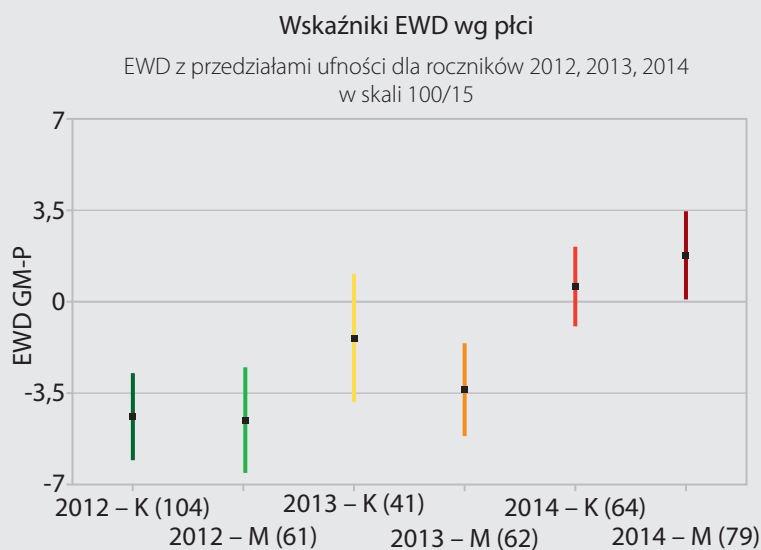
Rysunek 4.39. Efektywność nauczania w zakresie przedmiotów przyrodniczych w latach 2012–2014

We wskaźnikach EWD z przedmiotów przyrodniczych obserwujemy wzrost od niższej niż przeciętna efektywności nauczania w 2012 roku do ponadprzeciętnej efektywności w roku 2014. Sprawdźmy, w jakiej grupie uczniów uzyskano największy względny przyrost umiejętności. Czy wyższe EWD szkoły uzyskano dzięki lepszej pracy z uczniami słabszymi, a może ponadprzeciętne EWD uzyskali tylko uczniowie o wysokich wynikach na sprawdzianie? A może wdrożona w szkole metoda skuteczniej wpływa na postępy chłopców niż dziewczynek?

Odpowiedzi na te pytania dostarcza analiza wskaźników EWD wg uprzednich osiągnięć szkolnych oraz wg płci.

Rysunek 4.40. Efektywność nauczania w zakresie przedmiotów przyrodniczych w latach 2012–2014 w podziale na uprzednia osiągnięcia szkolne

Rysunek 4.41. Efektywność nauczania w zakresie przedmiotów przyrodniczych w latach 2012-2014 w podziale na płeć (K – dziewczęta, M – chłopcy)



W 2014 roku uzyskano porównywalną, wysoką efektywność nauczania we wszystkich analizowanych grupach uczniowskich.

Te wyniki przemawiają za tym, że udało się poprzez zastosowanie nowej metody dydaktycznej podnieść efektywność nauczania przedmiotów przyrodniczych. Należy spodziewać się, że w kolejnych latach wskaźniki EWD szkoły utrzymają się na ponadprzeciętnym poziomie.

Ograniczenia i rozwój metody EWD dla gimnazjów

Nowa formuła egzaminu gimnazjalnego w 2012 roku spowodowała, że wzrosła liczba wskaźników możliwych do wyliczenia dla każdego gimnazjum – nie tylko zbiorcze wskaźniki humanistyczne i matematyczno-przyrodnicze, ale również wskaźniki dla języka polskiego, historii i WOS-u, matematyki oraz przedmiotów przyrodniczych. Podobnie nowa formuła sprawdzianu w 2015 roku da impuls do prac nad szacowaniem nowych wskaźników EWD gimnazjalnych w 2018 roku, między innymi dla języków obcych. Odejście od ponadprzedmiotowego charakteru sprawdzianu zmusi do rewizji modeli szacowania EWD. Może to na przykład oznaczać, że inaczej będą liczone gimnazjalne wskaźniki EWD dla matematyki i polskiego (jako miary uprzednich osiągnięć będzie można używać wyników z wyodrębnionych na sprawdzianie testów przedmiotowych). Konstrukcji nowych wskaźników powinny towarzyszyć badania nad porównywalnością tych wskaźników z miarami liczonymi przed 2018 rokiem.

Ważnym problemem, który warto w tym miejscu zasygnalizować, jest sposób traktowania laureatów konkursów przedmiotowych. Wskaźniki EWD mają ograniczoną wartość dla szkół skupiających dużą liczbę laureatów konkursów przedmiotowych. Stanowią oni specyficzną grupę uczniów. Nie przystępują do egzaminu (w przypadku egzaminu gimnazjalnego – do jednej z jego części, odpowiadającej tematyce konkursu, którego są laureatami), lecz mają przypisywaną maksymalną liczbę punktów. Jednocześnie laureaci stanowią grupę niezwykle zróżnicowaną wewnętrznie. Liczba konkursów dających możliwość zwolnienia z egzaminu jest znaczna, a do tego różna w zależności od województwa. Ponadto, konkursy opracowywane lokalnie nie są porównywalne co do stopnia trudności. Sprawia to, że nie ma sensu w modelu EWD traktowanie bycia laureatem jako ziemnej kontroli. Z punktu widzenia wskaźników EWD najlepszym rozwiązaniem problemu związanego z laureatami byłoby, gdyby pisali oni egzaminy tak jak wszyscy uczniowie, a wyniki konkursów mogłyby być uwzględniane w rekrutacji do szkół ponadgimnazjalnych w inny sposób.

W trzyletnich wskaźnikach EWD do skalowania wyników wykorzystywana jest informacja o rozwiązaniu każdego zadania w teście, natomiast w kalkulatorze EWD 100 wykorzystuje się, głównie ze względu na prostotę obsługi aplikacji, skalowanie wyników na podstawie sumy punktów uzyskanych przez ucznia za rozwiązane zadań z arkusza egzaminacyjnego. Ujednolicenie podejścia byłoby możliwe, gdyby Kalkulator EWD 100 mógł pobierać dane z centralnej bazy połączonych i wyskalowanych wyników, co wymaga jednak rozwiązań dotyczących integracji egzaminacyjnych baz danych i opracowania systemu autoryzowanego dostępu do tych danych.

4.5. Wykorzystanie metody edukacyjnej wartości dodanej w liceach ogólnokształcących i technikach

W wypadku liceów ogólnokształcących i techników najlepszą, dostępną miarą zasobów „na wejściu” są wyniki uczniów na egzaminie gimnazjalnym. W maturalnych modelach EWD jako zmienne kontrolne wykorzystywane są informacje o płci i dysleksji ucznia oraz o tym, czy uczeń miał na etapie ponadgimnazjalnym wydłużony tok nauczania. Wskaźniki EWD są więc liczone w taki sposób, że szansa na ich wysoką wartość nie zależy od tego, z uczniami o jakich cechach (przede wszystkim wynikach egzaminu gimnazjalnego) pracowała szkoła. Szkoła, która świetnie pracowała z uczniami o przeciętnych wynikach na egzaminie gimnazjalnym, będzie charakteryzować się dodatnią wartością EWD, mimo że jej średnie wyniki matury będą prawdopodobnie niższe niż szkoły, która słabo pracowała z uczniami, którzy osiągnęli wysokie wyniki po gimnazjum.

Modele EWD dla liceów ogólnokształcących i techników

Wskaźniki EWD na poziomie ponadgimnazjalnym uwzględniają wyłącznie osiągnięcia w zakresie kształcenia ogólnego. Wskaźniki EWD wyliczane są oddzielnie dla liceów ogólnokształcących i dla techników. Oznacza to, że licea ogólnokształcące porównywane są do innych liceów ogólnokształcących, a technika do innych techników. W związku z tym nie ma możliwości dokonania bezpośredniego porównania dwóch szkół, z których jedna to liceum ogólnokształcące, a druga to technikum. Rozwiązanie takie przyjęto ze względu na istotne różnice pomiędzy tymi dwoma typami szkół. Licea ogólnokształcące to szkoły o profilu ogólnym, ukierunkowanym na przygotowanie do podjęcia studiów. Technika z kolei łączy kształcenie ogólne z kształceniem zawodowym. Duża część uczniów kończących te szkoły w ogóle nie decyduje się na przystąpienie do matury, a przystępujący często ograniczają się do zdawania jedynie obowiązkowego zestawu przedmiotów. Dodatkowo szkoły te różnią się długością cyklu kształcenia (trzyletni w LO i czteroletni w technikach). Są więc istotne powody, by twierdzić, że rozwój umiejętności ogólnych, mierzonych na maturze, przebiega odmiennie w obu tych typach szkół. Co więcej, potwierdzają to wyraźne rozbieżności w wynikach matur – zdecydowanie niższe wśród uczniów techników. W związku z tym istotne wydaje się uwzględnienie w modelowaniu EWD zróżnicowania zdających, związanego z typem szkoły, do której uczęszczali.

Jednoroczne i trzyletnie wskaźniki EWD dla liceów ogólnokształcących i techników

Wskaźniki EWD dla szkół maturalnych, podobnie jak dla gimnazjów, wyliczane są w dwóch podstawowych wariantach: jako tak zwane wskaźniki jednoroczne i trzyletnie. Pierwsze z nich uwzględniają wyniki jednego rocznika absolwentów, podczas gdy przy wyliczaniu tych drugich brane są pod uwagę osiągnięcia trzech kolejnych roczników absolwentów. Oba typy wskaźników różnią się metodami statystycznymi wykorzystywanymi w modelowaniu EWD, ale za ważniejszą różnicę należy uznać ich przeznaczenie i sposób udostępniania odbiorcom.

Wskaźniki trzyletnie dla liceów i techników, tak jak gimnazjalne, publikowane są w ogólnodostępnym serwisie internetowym, z myślą o bardzo szerokiej grupie odbiorców, od nauczycieli i dyrektorów, poprzez organy zarządzające i nadzór pedagogiczny, po rodziców i uczniów. Wskaźniki jednoroczne przeznaczone są przede wszystkim na potrzeby ewaluacji wewnętrznej. Grupa

osób mogących korzystać ze wskaźników jednorocznych jest ograniczona do tych, którzy mogą otrzymać z okręgowych komisji egzaminacyjnych dane z wynikami uczniów – w praktyce głównie nauczycieli i dyrektorów szkół.

Ze względu na zakres treści nauczania objętych danym wskaźnikiem wyróżniamy cztery różne typy wskaźników trzyletnich dla szkół maturalnych:

- w zakresie języka polskiego;
- w zakresie matematyki;
- w zakresie przedmiotów humanistycznych – obejmujące wyniki z języka polskiego, historii i wiedzy o społeczeństwie;
- w zakresie przedmiotów matematyczno-przyrodniczych – obejmujące wyniki z matematyki, biologii, chemii, fizyki, geografii i informatyki.

Syntetyczne miary osiągnięć maturzystów tworzone są z uwzględnieniem wyników z wszystkich przedmiotów, które zdawali poszczególni uczniowie (czy to na poziomie podstawowym, czy rozszerzonym), a które przypisane są do danego wskaźnika. Stworzenie takich złożonych miar osiągnięć jest możliwe dzięki zastosowaniu odpowiednich metod statystycznych, a konkretnie teorii odpowiedzi na zadanie testowe (IRT). W dalszej części tekstu opisane zostały wybrane problemy, z którymi można sobie dzięki tej metodzie poradzić. W przypadku wskaźników EWD w zakresie języka polskiego i przedmiotów humanistycznych jako miara osiągnięć „na wejściu” wykorzystywane są wyniki części humanistycznej egzaminu gimnazjalnego. W przypadku wskaźników EWD w zakresie matematyki i przedmiotów matematyczno-przyrodniczych są to z kolei wyniki części matematyczno-przyrodniczej egzaminu gimnazjalnego.

Wskaźniki jednoroczne EWD dla LO i techników, pozwalające przy użyciu Kalkulatora EWD na prowadzenie analiz wewnątrzszkolnych w zakresie nauczania matematyki, zostały udostępnione po raz pierwszy dopiero w 2013 r., przy czym wyliczono i udostępniono poprzez Kalkulator EWD modele również dla lat wcześniejszych, do roku 2010 włącznie. Pierwotnie zostały one wyliczone w oparciu wyłącznie o wyniki matury z matematyki na poziomie podstawowym (jako wskaźnik osiągnięć na wejściu wykorzystano wyniki części matematyczno-przyrodniczej egzaminu gimnazjalnego), co pozwoliło wykorzystać analogiczną metodologię jak w przypadku jednorocznych wskaźników gimnazjalnych. Wskaźniki jednoroczne uwzględniające wyłącznie poziom podstawowy matury z matematyki miały jednak istotną wadę. Ze względu na występujący w teście maturalnym z matematyki na poziomie podstawowym efekt sufitowy model statystyczny zaniżał EWD liceów, których uczniowie mieli bardzo wysokie umiejętności matematyczne. Aby przezwyciężyć ten problem, w roku 2014 zastosowano skalowanie wyników maturalnych modelem Rascha. Pozwoliło to uwzględnić wyniki matury z matematyki zarówno na poziomie podstawowym, jak i na poziomie rozszerzonym, jednocześnie pozostawiając możliwość posługiwania się w Kalkulatorze EWD wyłącznie informacją o sumie punktów zdobytych w każdej z tych części egzaminu. Obecnie prowadzone są prace nad zaadaptowaniem tej metodologii do wyliczenia jednorocznych wskaźników EWD w zakresie języka polskiego, wymaga to jednak rozwiązania dodatkowych trudności, związanych z wyborem tematu wypracowania.

Skalowanie wyników matury

Polska matura jest egzaminem o bardzo złożonej strukturze. W latach 2010–2014 składała się z dwudziestu różnych przedmiotów, w tym sześciu języków obcych, nie wliczając egzaminów z języków mniejszości narodowych i etnicznych. Dodatkowo dla zdecydowanej większości przedmiotów układane są w każdym roku arkusze na dwóch różnych poziomach trudności: podstawowym i zaawansowanym.

Liczba piszących poszczególne części egzaminu maturalnego jest bardzo zróżnicowana. Wszyscy zdający maturę – obecnie około 300 tys. uczniów szkół ponadgimnazjalnych rocznie – podchodzą do przedmiotów obowiązkowych: języka polskiego i matematyki na poziomie podstawowym.

Spośród języków obcych, z których obowiązkowo trzeba wybrać jeden na poziomie podstawowym, zwykle około 85% zdających wybiera język angielski. Najpopularniejsze przedmioty zdawane jako dodatkowe to w ostatnich latach geografia i biologia na obu poziomach oraz chemia na poziomie rozszerzonym i WOS na poziomie podstawowym – wybiera je od około 20 tys. do około 45 tys. zdających. Są też przedmioty zdawane rokrocznie przez nie więcej niż kilkaset osób, jak filozofia, historia muzyki, wiedza o tańcu czy język łaćski i kultura antyczna.

Takie „rozdrobienie” egzaminu maturalnego stanowi istotne wyzwanie w sytuacji, gdy chcemy wykorzystać jego wyniki do analizy np. efektywności nauczania w szkole. Ze względu na to, że w bardzo wielu szkołach poszczególne przedmioty (a dokładniej poszczególne przedmioty na poszczególnych poziomach) zdawane są przez bardzo niewielu uczniów (wyjawszy oczywiście przedmioty obowiązkowe), wydaje się, że wykorzystanie wyników z tych przedmiotów do konstrukcji oddzielnych wskaźników efektywności pracy szkół w zakresie pojedynczych przedmiotów byłoby rozwiązaniem mało użytecznym, bowiem wskaźniki uzyskiwane na poziomie szkół byłyby w większości wypadków obarczone ogromną niepewnością statystyczną. W związku z tym dużo odpowiedniejsze do wykorzystania w tym kontekście wydają się bardziej ogólne miary umiejętności zdających egzamin maturalny, które byłyby w stanie uwzględnić jednocześnie wyniki różnych części tego egzaminu.

Takie złożone miary umiejętności możemy określić mianem kompozycyjnych, jako że łączą one ze sobą wyniki wielu bardziej szczegółowych pomiarów w ogólniejsze, bardziej syntetyczne wskaźniki. Wykorzystanie metody IRT pozwala skonstruować takie miary osiągnięć nawet na podstawie egzaminu o bardzo złożonej strukturze. Można powiedzieć, że w efekcie zastosowania metody IRT otrzymujemy pewien rodzaj łączenia ze sobą wyników różnych części egzaminu, a więc sprowadzania ich wyników do wspólnej skali. Oczywiście, aby było to możliwe, konieczne jest spełnienie warunku zdawania przez te same osoby różnych części egzaminu. W kontekście polskiej matury zasadnicze znaczenie dla możliwości konstruowania takich miar ma oczywiście występowanie dwóch przedmiotów obowiązkowo pisanych przez wszystkich: języka polskiego i matematyki na poziomie podstawowym.

Jak wspomniano już wcześniej, na potrzeby wyliczania wskaźników EWD konstruowane są na podstawie wyników matury cztery różne miary osiągnięć:

- w zakresie języka polskiego – na poziomie podstawowym i rozszerzonym;
- w zakresie matematyki – na poziomie podstawowym i rozszerzonym;
- w zakresie przedmiotów humanistycznych – na podstawie wyników z języka polskiego, historii i wiedzy o społeczeństwie na poziomie podstawowym i rozszerzonym;
- w zakresie przedmiotów matematyczno-przyrodniczych – na podstawie wyników z matematyki, biologii, chemii, fizyki, geografii i informatyki na poziomie podstawowym i rozszerzonym.

Dwie pierwsze są dużo prostsze – uwzględniają wyłącznie oba poziomy trudności jednego przedmiotu. Dwie kolejne, a zwłaszcza miara matematyczno-przyrodnicza, są bardzo skomplikowanymi miarami kompozycyjnymi. Każda z tych miar prezentowana jest przy pomocy skali standardowej o średniej 100 i odchyleniu standardowym 15, z tym że przeliczenie parametrów skali dokonywane jest oddzielnie dla uczniów techników i oddzielnie dla uczniów liceów ogólnokształcących. Tak więc w przypadku techników grupę odniesienia stanowią wszyscy uczniowie techników, którzy zdają maturę po raz pierwszy, a w przypadku liceów, wszyscy uczniowie liceów, którzy zdają maturę po raz pierwszy.

Poniżej omówiony zostanie tylko jeden szczególny problem, z jakim musimy się zmierzyć przy konstrukcji kompozycyjnych miar osiągnięć na podstawie wyników matury, a mianowicie kwestia wyboru przez uczniów zestawu zdawanych przedmiotów⁵⁰.

⁵⁰ Czytelników zainteresowanych bardziej szczegółowymi analizami odsyłamy do publikacji Tomasza Żółtaka (Żółtak, 2013: s. 13) i książki opisującej problemy modelowania cech ukrytych (Pokropek, 2015).

Problem autoselekcji przy wyborze przedmiotów zdawanych na maturze

Jak wspomniano, specyfika egzaminu maturalnego polega na tym, że uczniowie sami decydują, jakie przedmioty i na jakim poziomie będą zdawać. Oczywiście wyborów tych nie dokonują losowo, lecz w sposób strategiczny, związany z preferowanymi przedmiotami (w dużej mierze zależy to od wymagań rekrutacyjnych na kierunkach studiów, na które chce się dostać maturzysty) i poziomem swoich umiejętności (należy się spodziewać, że przedmioty na poziomie rozszerzonym wybierają uczniowie o wyższym poziomie umiejętności). Aby uzyskać dobre oszacowania parametrów modelu skalowania, a potem także dobre oszacowania poziomu umiejętności zdających, konieczne jest więc uwzględnienie w jakiś sposób w modelu występowania takich procesów wyborów.

W tym celu zdecydowano się na włączenie do modelu dodatkowych zmiennych (po jednej dla każdej części egzaminu) traktowanych tak jak zadania, a opisujących, czy uczeń zdawał daną część egzaminu, czy też nie (wzorowano się na rozwiązaniu zaproponowanym w publikacji Korobko i współpracowników – 2008). Na podstawie charakterystyk takich „zadań”, które określamy mianem parametrów selekcji, możemy potem określić, w jakim stopniu wybór poszczególnych części egzaminu powiązany jest z poziomem mierzonych umiejętności.

Problemem bardzo zbliżonym do wyboru zdawanych przedmiotów jest wybór tematu wypracowania z języka polskiego, a także wypowiedzi pisemnych wchodzących w skład arkuszy z WOS-u i historii na poziomie rozszerzonym. Może on zostać rozwiązany w analogiczny sposób, tj. poprzez potraktowanie tych samych kryteriów oceny wypracowania jako różnych zadań – w zależności od wybranego tematu – i dodanie do modelu dodatkowego „zadania” opisującego dokonany wybór tematu (Pokropek, 2011).

W oparciu o wyliczenia dokonane w procesie konstruowania naszych kompozycyjnych miar osiągnięć możemy przyjrzeć się „elitarności” poszczególnych przedmiotów maturalnych, to jest temu, jak silne są zależności pomiędzy poziomem umiejętności a skłonnością do zdawania danego przedmiotu na danym poziomie. Jak opisano wyżej, mówią nam o tym tzw. parametry selekcji. Ich wartości z poszczególnych modeli skalowania z lat 2012–2014 przedstawione zostały na poniższych rysunkach. Wartości parametrów większe od zera wskazują, że wybór danego przedmiotu (na danym poziomie) jest statystycznie pozytywnie powiązany z poziomem osiągnięć. Inaczej mówiąc, dany przedmiot wybierali raczej zdający o wysokich umiejętnościach w danym obszarze nauczania. Im wartość parametru większa, tym częściej dany przedmiot (na danym poziomie) wybierają uczniowie o wysokich umiejętnościach, a tym rzadziej uczniowie o niskich umiejętnościach. Ujemne wartości parametrów selekcji wskazują z kolei, że dany przedmiot, na danym poziomie wybierali raczej zdający o poziomie umiejętności niższym od średniej wśród wszystkich zdających.

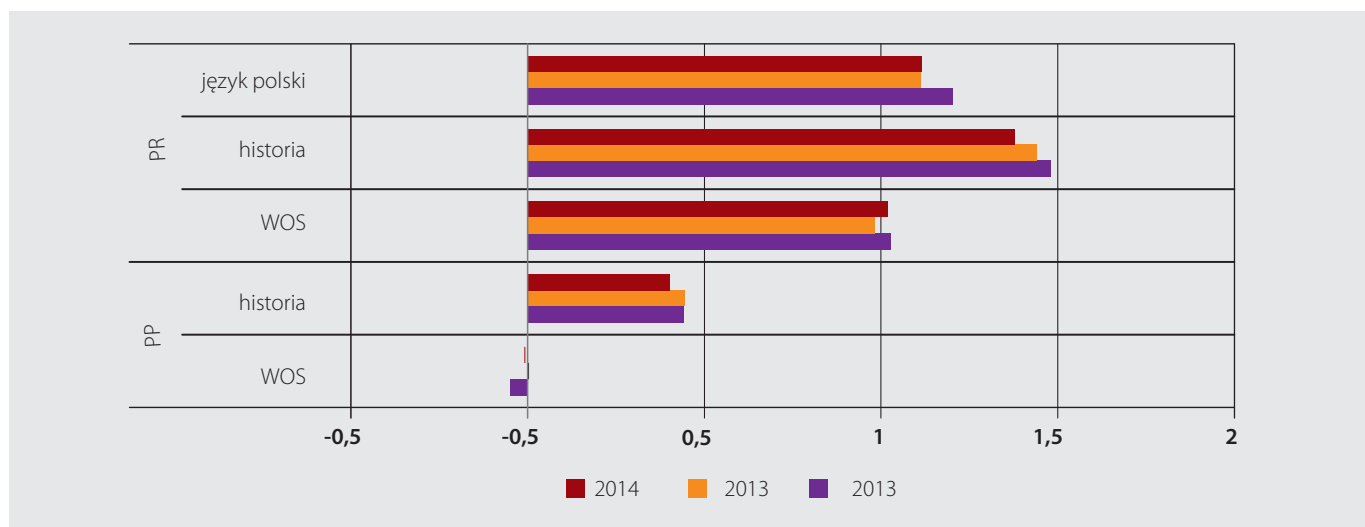
Biorąc pod uwagę konstrukcję egzaminu maturalnego, oczekivalibyśmy dodatnich wartości współczynników selekcji związanych z wyborem przedmiotów na poziomie rozszerzonym. Oczekivalibyśmy również, że w ramach każdego przedmiotu wartość parametru selekcji będzie większa dla poziomu rozszerzonego niż dla poziomu podstawowego. Okazuje się, że oczekiwania te znajdują potwierdzenie w danych. Wartości parametrów selekcji dla poziomu rozszerzonego wyniosły od 0,63 (geografia w 2014 r.) do 2,2 (matematyka w 2012 r.) dla przedmiotów matematyczno-przyrodniczych i od 0,98 (WOS w 2013 r.) do 1,48 (historia w 2012 r.) dla przedmiotów humanistycznych. W przypadku każdego przedmiotu wartości parametrów selekcji dla poziomu rozszerzonego są też wyraźnie wyższe niż dla poziomu podstawowego.

Zdecydowanie najbardziej poziomem umiejętności wyróżniają się zdający rozszerzoną matematykę i fizykę. Mniej „elitarnie” okazały się rozszerzona biologia i geografia. Pozostałe przedmioty przyrodnicze oraz przedmioty humanistyczne na poziomie rozszerzonym plasowały się pomiędzy tymi dwoma grupami przedmiotów. Jeśli chodzi o wybór przedmiotów na poziomie podstawowym, to możemy wyróżnić takie, które wybierane są przez raczej słabszych uczniów (biologia i geografia – wartości parametrów selekcji około $-0,2$), takie, których wybór jest właściwie bez związku z poziomem umiejętności (WOS), oraz takie, które wybierają raczej lepsi zdający (chemia, informatyka,

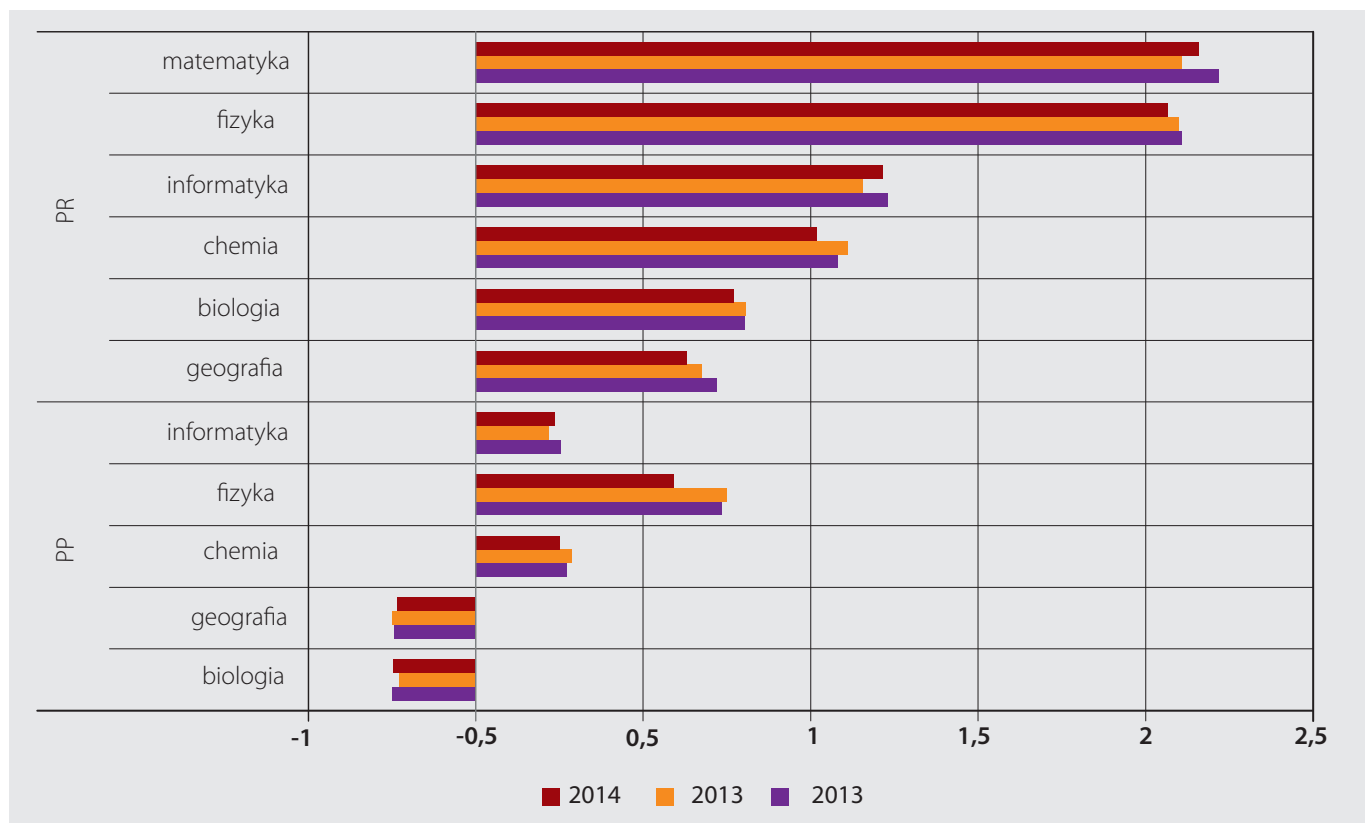
historia, fizyka). Warto zauważyć, że wartości parametrów selekcji dla fizyki na poziomie podstawowym były porównywalne z tymi dla rozszerzonej biologii czy geografii.

Warta odnotowania jest także duża stabilność wartości parametrów selekcji pomiędzy latami. Wskazuje to, że strategie wyboru przedmiotów przez zdających są – przynajmniej w zakresie, w jakim rolę odgrywa tu poziom umiejętności – dosyć stałe, pomimo sporej zmienności w wymaganiach rekrutacyjnych uczelni oraz czasami zróżnicowanego poziomu trudności danej części egzaminu między latami.

Rysunek 4.42. Wartości parametrów selekcji pokazujących siłę związku pomiędzy poziomem umiejętności humanistycznych a skłonnością do zdawania poszczególnych przedmiotów na poszczególnych poziomach



Rysunek 4.43. Wartości parametrów selekcji pokazujących siłę związku pomiędzy poziomem umiejętności matematyczno-przyrodniczych a skłonnością do zdawania poszczególnych przedmiotów na poszczególnych poziomach



Trafność wskaźników EWD dla liceów ogólnokształcących i techników

Analizowaniu trafności wskaźników EWD dla liceów ogólnokształcących i techników poświęcone zostało badanie *Ścieżki rozwoju edukacyjnego młodzieży – szkoły gimnazjalne*, zrealizowane w latach 2009–2013 przez IFiS PAN (Karwowski, 2013). W jego ramach analizowano między innymi kwestię, na ile nieuwzględnienie w polskich modelach EWD dodatkowych zmiennych kontrolnych, przede wszystkim opisujących status społeczno-ekonomiczny rodziny ucznia, może prowadzić do niekształcenia informacji na temat efektywności pracy szkoły. Badano też wpływ na osiągnięcia inteligencji ogólnej oraz, choć w bardzo ograniczonym zakresie, udziału w korepetycjach i zajęciach pozalekcyjnych.

Ramka 4.3. Opis podłużnego badania uwarunkowań wyników nauczania w szkołach podstawowych

Badanie *Ścieżki rozwoju edukacyjnego młodzieży – szkoły gimnazjalne* prowadzone było w latach 2009–2013 i objęło reprezentatywną, ogólnopolską próbę losową łącznie 200 oddziałów klas pierwszych ze 100 liceów ogólnokształcących (ok. 2,4 tys. uczniów), 60 techników i liceów profilowanych (ok. 1,2 tys. uczniów) oraz 40 zasadniczych szkół zawodowych (ok. 700 uczniów). Badanie realizowane było w ramach projektu *Badania dotyczące rozwoju metodologii szacowania wskaźnika edukacyjnej wartości dodanej (EWD)*, a jego realizację w wyniku otwartego przetargu zlecono Instytutowi Filozofii i Socjologii Polskiej Akademii Nauk. Badanie objęło pełen cykl kształcenia na etapie ponadgimnazjalnym (śledzone są losy tych samych uczniów). Uczestniczyli w nim uczniowie, ich rodzice, nauczyciele oraz dyrektorzy szkół.

Badanie miało na celu lepsze poznanie indywidualnych, rodzinnych i szkolnych czynników odpowiedzialnych za osiągnięcia szkolne uczniów. Było to punktem wyjścia do oceny, na ile stosowane w Polsce modele EWD dla szkół ponadgimnazjalnych pozwalają w odpowiedni sposób kontrolować czynniki niezależne od działań szkoły. Analizowano także, czym charakteryzują się szkoły osiągające wysokie, a czym osiągające niskie wartości wskaźników EWD. Podczas pierwszego etapu badania, który odbył się, kiedy badani uczniowie uczęszczali do pierwszych klas, zebrano dane z wykorzystaniem narzędzi z badania PISA 2009. Dotyczyło to zarówno pomiaru umiejętności w zakresie czytania i interpretacji, matematyki i rozumowania w naukach przyrodniczych, jak i kwestionariuszy dla uczniów, nauczycieli i dyrektorów szkół. Podczas kolejnego etapu badania, zrealizowanego na początku drugiej klasy, wykorzystano trzy testy psychologiczne: test matryc Ravena (inteligencja), test kompetencji społecznych i test nadziei na sukces. W rok po pierwszym etapie badania zrealizowano etap trzeci, w którym powtórnie zmierzono umiejętności uczniów z wykorzystaniem testów PISA oraz zastosowano testy psychologiczne do oceny samooceny i lęku. Zebrano też wtedy informacje o funkcjonowaniu uczniów w środowisku rówieśniczym (Karwowski, 2013). Więcej informacji na temat badania można znaleźć na stronie: http://www.ifispan.waw.pl/index.php?lang=pl&m=page&pg_id=215

Wyniki analiz wskazują, że na etapie edukacji ponadgimnazjalnej wykorzystanie w modelach EWD dla liceów ogólnokształcących i techników informacji o wynikach uczniów na egzaminie gimnazjalnym pozwala niemal całkowicie kontrolować wpływ czynników statusowych. W dużym stopniu, choć nie tak skutecznie jak w wypadku SES, pozwala też kontrolować wpływ inteligencji, rozumianej jako potencjał poznawczy wynikający z czynników biologicznych. Pewną poprawę w tym zakresie można by zapewne uzyskać, zwiększając rzetelność egzaminu gimnazjalnego.

Analizowano również, jakie cechy szkoły stanowią korelaty wysokich wartości wskaźników EWD. Zarówno w LO, jak i w technikach wysokim wartościom EWD sprzyjało efektywne wykorzystanie czasu na lekcji (czas poświęcony na naukę w stosunku do czasu poświęcanego na utrzymanie porządku w klasie i na czynności administracyjne) oraz proszkolny klimat panujący wśród uczniów (zarówno gdy był on oceniany przez nauczycieli, jak i gdy uczniowie oceniali nastawienie swoich koleżanek i kolegów). Co ciekawe, częste stosowanie w nauczaniu metod podających (mierzone poprzez deklaracje nauczycieli nt. stosowanych technik nauczania) negatywnie korelowało z EWD w liceach ogólnokształcących, ale w technikach albo było to bez znaczenia, albo, w przypadku języka polskiego, sprzyjało wyższym wartościom EWD. W liceach ogólnokształcących wysokie EWD szkoły było powiązane z pozytywnymi postawami wobec czytania (deklaracje, że czyta się często i chętnie), nasileniem wśród uczniów danej szkoły relacji koleżeńskich i relacji przyjaźni oraz bogatą ofertą zajęć dodatkowych organizowanych przez szkołę. Wyniki te są ogólnie rzecz biorąc zgodne z oczekiwaniami co do tego, jakimi cechami powinny charakteryzować się „dobre” szkoły i choć nie może to być decydującym argumentem, jednak przemawia na korzyść twierdzenia o trafności wskaźników EWD.

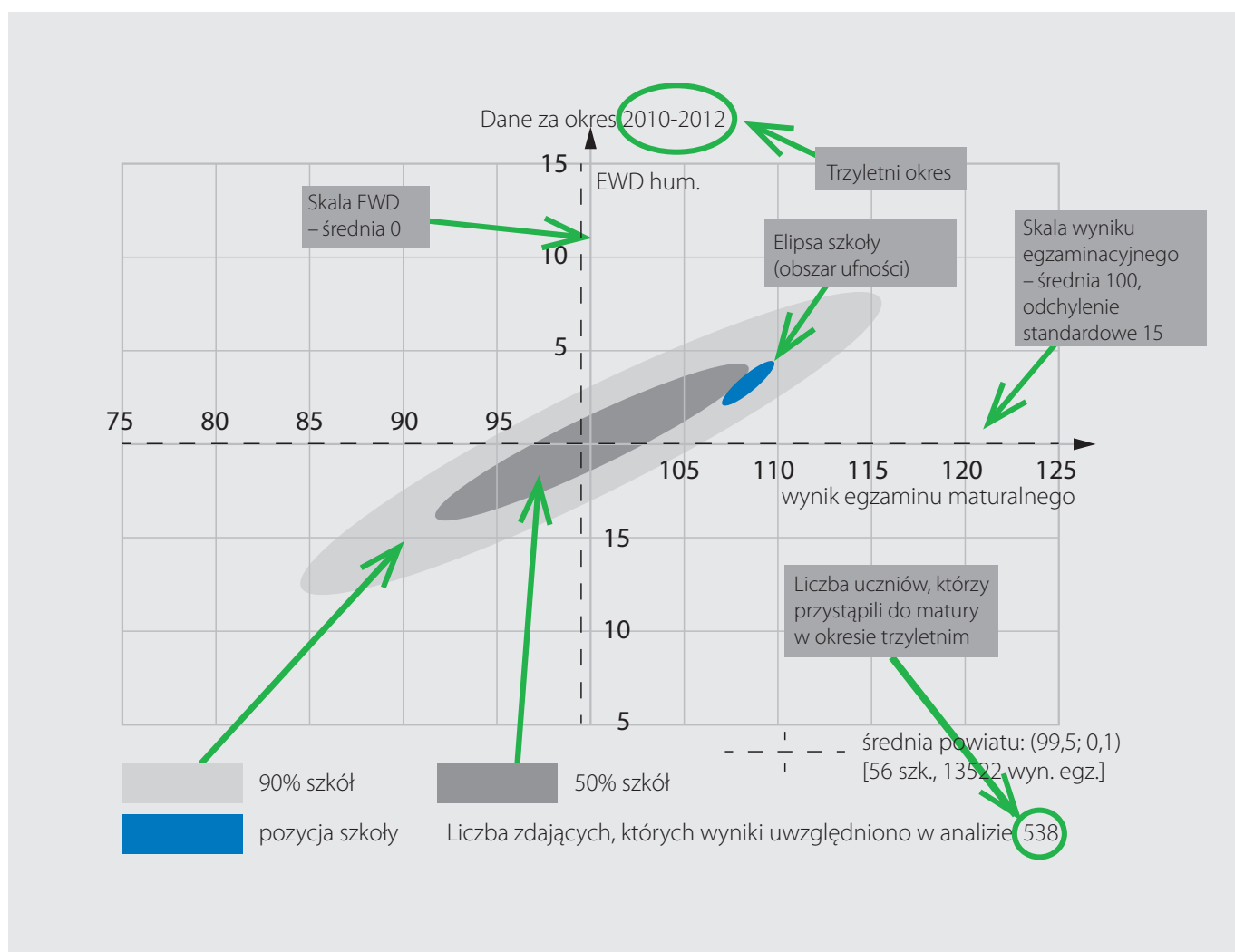
Sposób prezentacji wskaźników EWD dla liceów ogólnokształcących i techników

Dla prezentacji trzyletnich wskaźników maturalnych EWD, podobnie jak dla gimnazjów, przyjęto formę graficzną: na jednym wykresie pokazano jednocześnie wyniki egzaminacyjne oraz EWD. Szkołę reprezentuje elipsa (obszar ufności), wielkość której zależy od liczby uczniów w szkole. Elipsę szkoły należy traktować jako obszar ufności, czyli obszar, w którym z 95% prawdopodobieństwem znajduje się prawdziwy wynik szkoły. Im więcej uczniów przystąpiło w szkole do egzaminu maturalnego, tym dokładniej szacowane są wyniki i EWD, a więc, tym mniejsza jest elipsa.

Na podstawie wykresu możemy też wyznaczyć jednowymiarowe przedziały ufności, oddzielnie dla średnich wyników szkoły i oddzielnie dla EWD. Uzyskamy to, rzutując elipsę odpowiednio na oś wyników matury lub na oś EWD. Należy jednak pamiętać, że po dokonaniu takich rzutowań tracimy informację o współzależności między tymi dwoma cechami szkoły, o której informuje nas kształt całej elipsy.

Na rysunku 4.44. pokazano wszystkie najważniejsze elementy graficznej prezentacji trzyletnich wskaźników maturalnych. Dwie środkowe (szare) elipsy, wskazują obszary, w których koncentruje się odpowiednio 50% i 90% średnich wyników szkół. Innymi słowy, szare elipsy określają obszary najbardziej typowych wyników.

Rysunek 4.44. Podstawowe elementy graficznej prezentacji trzyletnich wskaźników maturalnych



Uzupełnieniem graficznej prezentacji wyniku egzaminacyjnego i EWD są tabele, w których podano informację o liczbie osób, które przystąpiły w szkole do egzaminu maturalnego w analizowanym okresie, wraz ze strukturą wyboru przedmiotów i poziomów egzaminów.

Kalkulator EWD 100 daje możliwość przeprowadzenia wielu różnych typów analiz wewnątrzszkolnych, odnoszących się zarówno do wyników matury z matematyki, EWD, jak i do wyników egzaminu gimnazjalnego. Funkcjonalności Kalkulatora zostaną przedstawione w części opisującej praktyczne wykorzystanie wskaźników EWD. Warto jedynie zaznaczyć, że informacje o EWD i/lub średnich wynikach matury prezentowane są w nim wyłącznie w formie jednowymiarowych przedziałów ufności (reprezentowanych graficznie w postaci odcinków). Nie ma możliwości wyrysowania w Kalkulatorze EWD obszaru ufności (elipsy), który pokazywałby łączną informację zarówno o wynikach matury, jak i o EWD.

Zarówno w przypadku wskaźników jednorocznych, jak i trzyletnich należy pamiętać, że wskaźniki EWD mają charakter względny. W skali kraju wskaźnik EWD ma z definicji wartość równą zero. Wskaźnik EWD dla liceum ogólnokształcącego mówi o tym, na ile wysokie/niskie wyniki matury uzyskali jego absolwenci w porównaniu do uczniów o analogicznych wynikach na egzaminie gimnazjalnym. Podobnie wskaźnik EWD dla technikum pokazuje, na ile wysokie/niskie wyniki matury uzyskali jego absolwenci w porównaniu do uczniów techników w Polsce przy statystycznej kontroli wyników na egzaminie gimnazjalnym. Wartość dodatnia EWD wskazuje na ponadprzeciętną w skali kraju efektywność nauczania, wartość ujemna na niższą niż przeciętna efektywność.

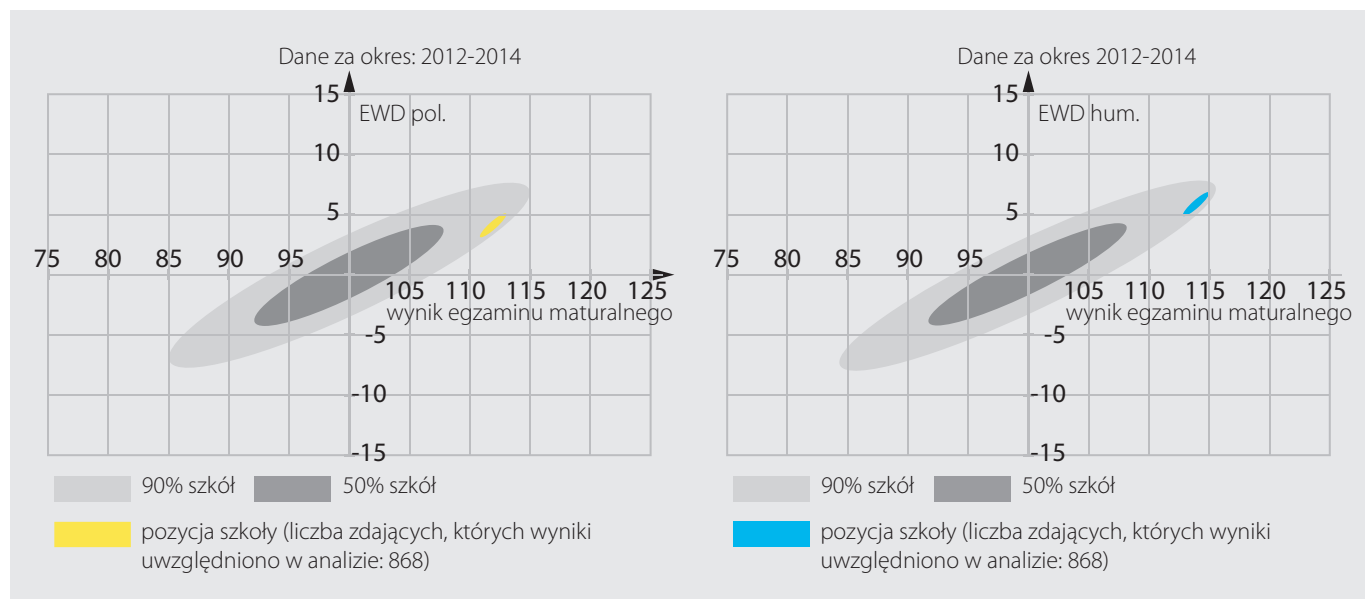
Możliwości wykorzystania metody EWD przez licea ogólnokształcące i technika

W tym rozdziale przedstawiono dwa przykłady wykorzystania metody EWD do analizy wyników egzaminacyjnych w liceach i technikach. Zaprezentowane wykresy pokazują potencjał analityczny trzyletnich wskaźników EWD i Kalkulatora EWD 100.

Liceum Ogólnokształcące A. To bardzo duża wielkomijska szkoła, ze stuletnią tradycją. Po jej skończeniu większość młodzieży dostaje się na wybrany kierunek studiów.

Zobaczmy, jaki obraz szkoły rysuje się na podstawie trzyletnich wskaźników EWD za okres 2012–2014.

Rysunek 4.45. Liceum A. Przedmioty humanistyczne

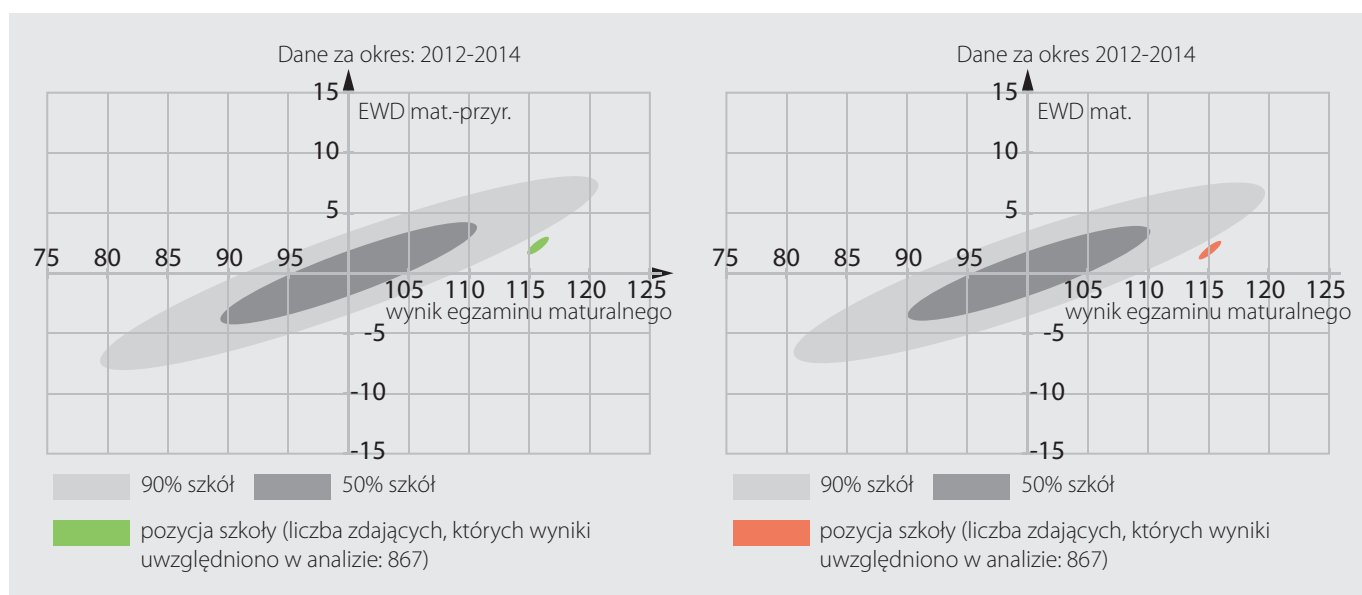


Przedmiot	Łącznie	Poziom rozszerzony
j. polski	868	272
historia	133	133
WOS	168	167

W liceum prawie 1/3 uczniów zdaje język polski na poziomie rozszerzonym, jednak można zauważyć, że wyższą efektywność i wyniki egzaminacyjne osiągają zdający, gdy wskaźniki liczone są łącznie dla wszystkich przedmiotów humanistycznych, niż dla samego języka polskiego. Elipsa wyników szkoły (EWD oraz wynik egzaminacyjny) znajduje się wtedy nad przerywaną zieloną kreską, natomiast dla samego języka polskiego efektywność liczona wskaźnikiem EWD jest niższa (pod przerywaną kreską). Analogiczna zależność występuje w odniesieniu do wyników matury, tj. elipsa opisująca wyniki w zakresie samego języka polskiego przesunięta jest w lewo względem tej, która uwzględnia wszystkie przedmioty humanistyczne. Wyniki z historii i wiedzy o społeczeństwie przyczyniają się do osiągania przez szkołę wyższych wyników egzaminacyjnych i wyższej efektywności.

W części matematyczno-przyrodniczej nie występują różnice w wartościach wskaźników z przedmiotów matematyczno-przyrodniczych i dla samej matematyki.

Rysunek 4.46. Liceum A. Przedmioty matematyczno-przyrodnicze

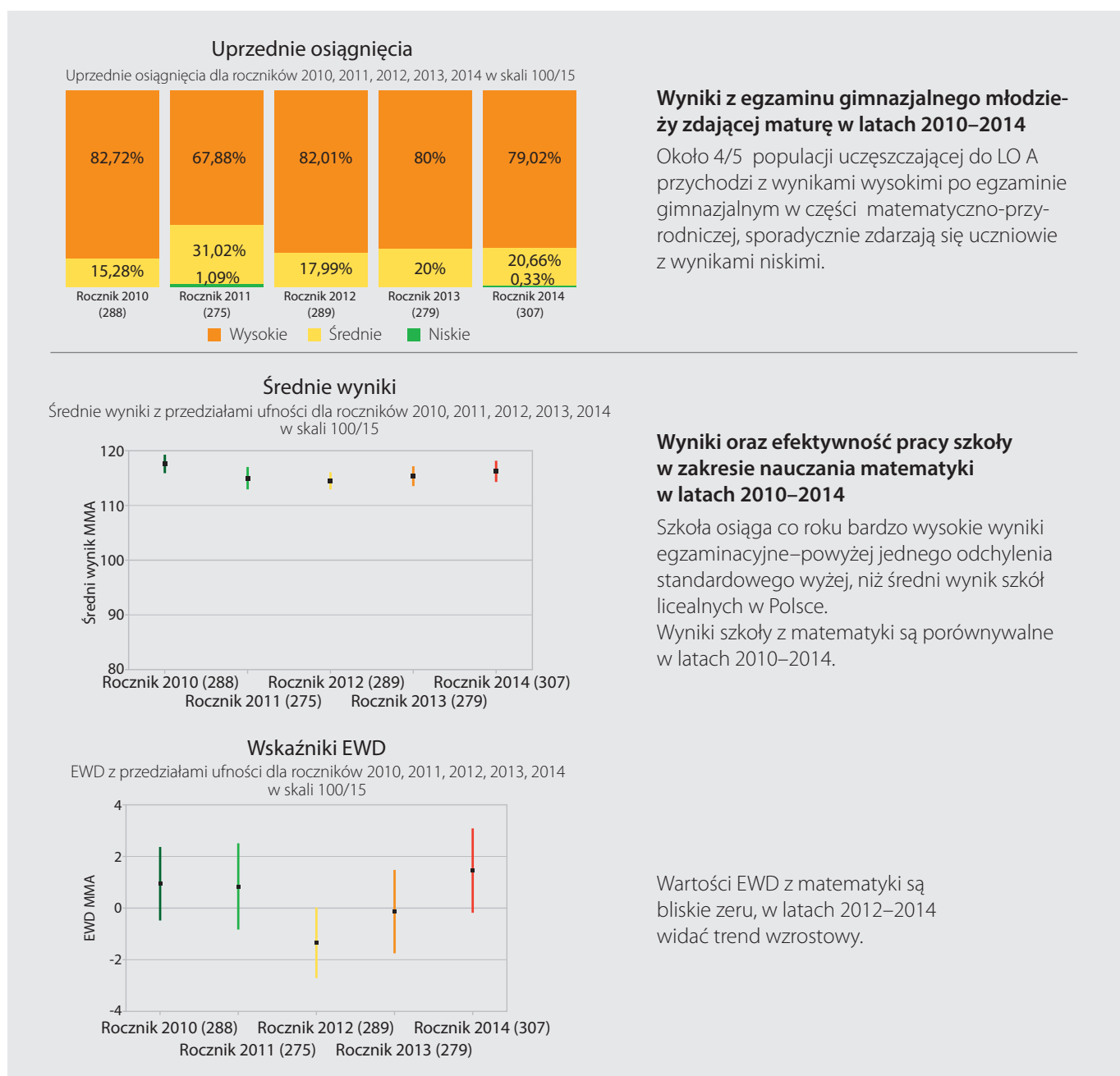


Przedmiot	Łącznie	Poziom rozszerzony
matematyka	868	438
biologia	203	199
chemia	229	226
fizyka	38	33
geografia	135	133
informatyka	20	19

Wyniki z przedmiotów matematyczno-przyrodniczych są o jedno odchylenie standardowe wyższe od średnich wyników dla szkół licealnych, efektywność pracy szkoły jest ponadprzeciętna, ale nie jest bardzo wysoka.

Dzięki Kalkulatorowi EWD 100 można dokładniej przyjrzeć się efektywności nauczania w liceum A. Analizy wyników szkoły zostały wykonane dla danych egzaminacyjnych z matematyki za lata 2010–2014. Kalkulator EWD 100 dla szkół maturalnych umożliwia wykonywanie obliczeń statystycznych dla połączonych wyników egzaminu gimnazjalnego z części matematyczno-przyrodniczej i **matematyki** zdawanej na maturze (poziom podstawowy i rozszerzony).

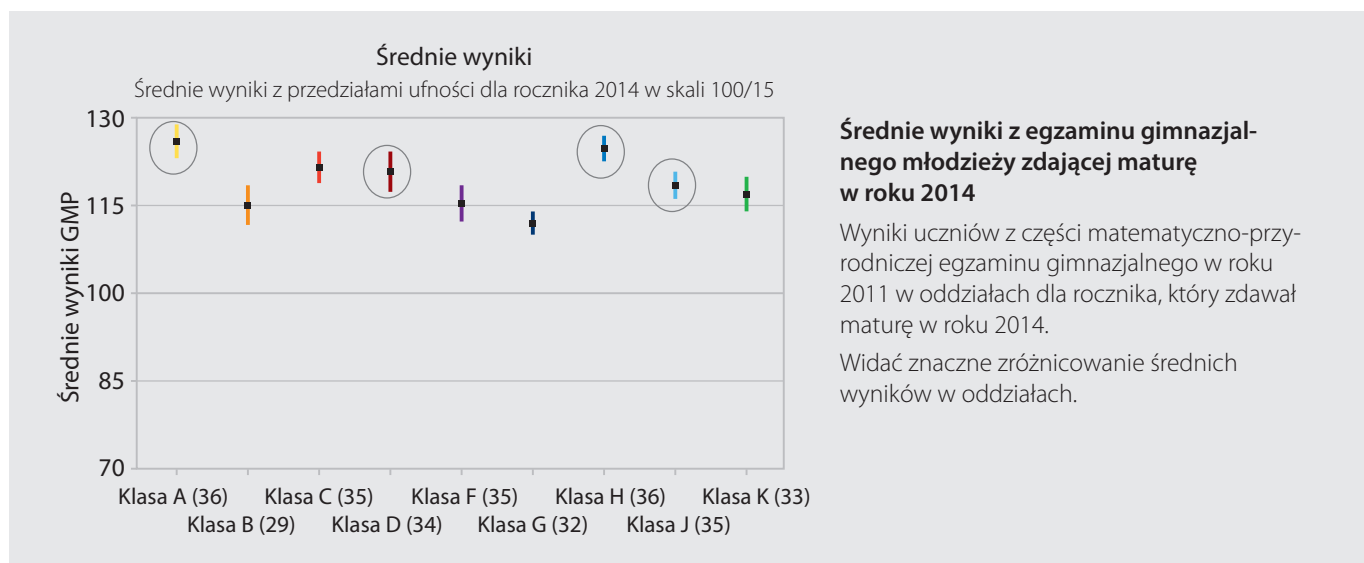
W LO A uczy się młodzież o bardzo wysokich osiągnięciach na egzaminie gimnazjalnym. Pokazuje nam to wykres *Uprzednie osiągnięcia*, jedna z analiz Kalkulatora EWD 100.

Rysunek 4.47. Liceum A. Analizy wykonane za pomocą Kalkulatora EWD 100

Wyniki analizy wskazują, że wysokie wyniki egzaminu maturalnego z matematyki są głównie efektem tego, że do liceum A przychodzi młodzież z bardzo wysokimi wynikami z matematyki na egzaminie gimnazjalnym, natomiast wpływ edukacyjny szkoły jest przeciętny (od roku 2012 zarysowuje się trend wzrostowy EWD).

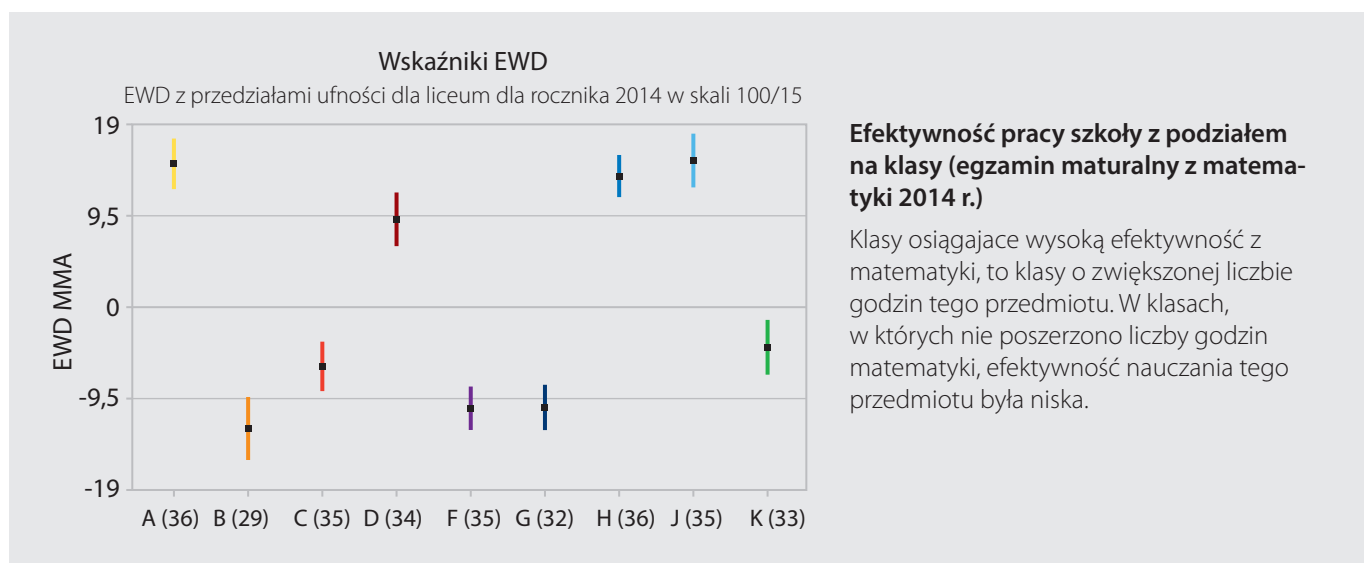
Przejdźmy do pogłębionych analiz wewnętrznych. Zaczniemy od analizy uprzednich osiągnięć uczniów w poszczególnych oddziałach liceum A.

Rysunek 4.48. Średni wynik uczniów na egzaminie gimnazjalnym z podziałem na oddziały



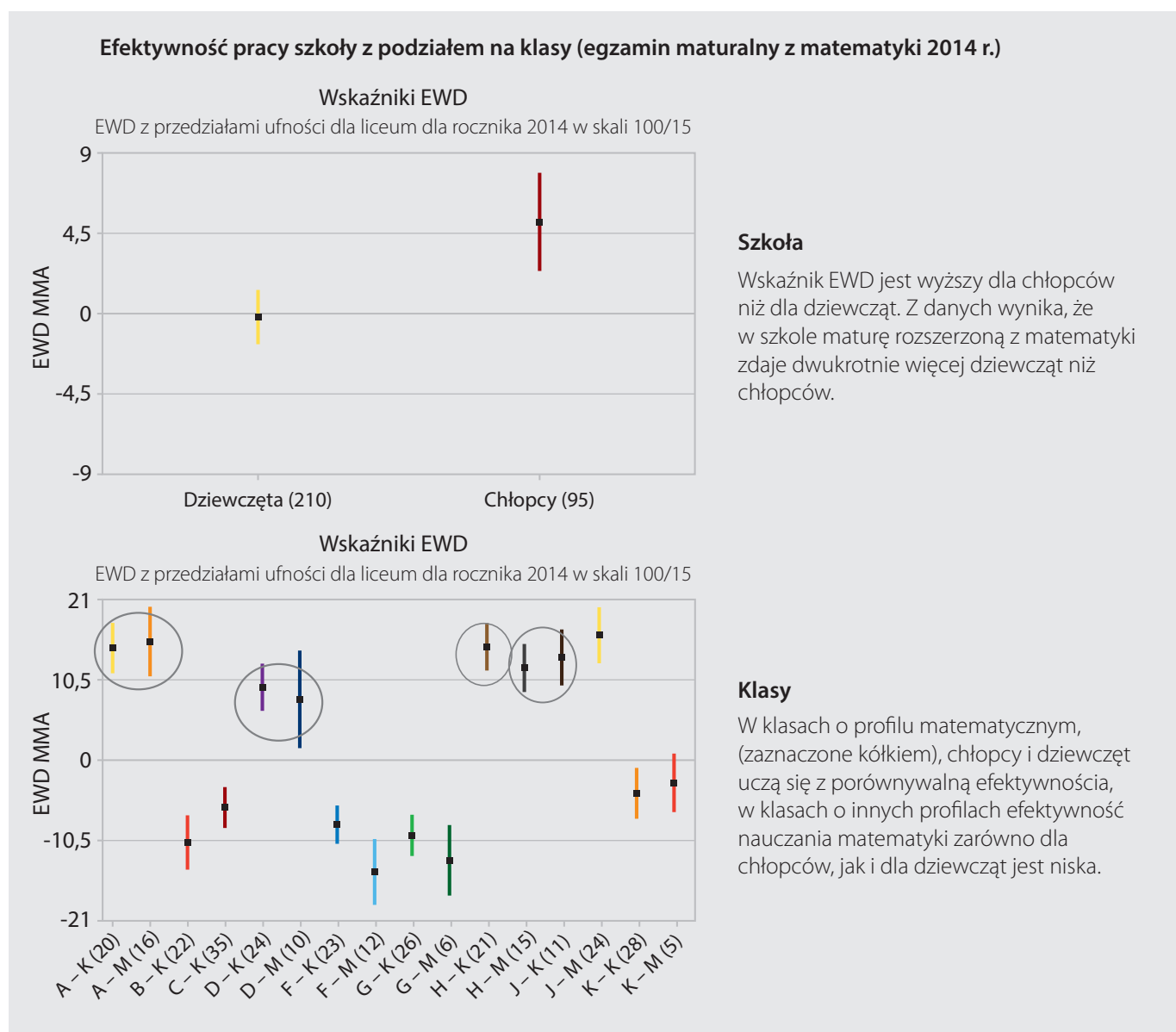
Oddziały zaznaczone kółkiem (A, D, H i J) realizowały program z dodatkową liczbą godzin z matematyki. Do klas A i H została wybrana młodzież o bardzo wysokich wynikach z egzaminu gimnazjalnego (część matematyczno-przyrodnicza): 1,5 odchylenia standardowego powyżej średniej w Polsce. Przyjrzyjmy się zatem EWD w poszczególnych oddziałach

Rysunek 4.49. EWD w poszczególnych oddziałach



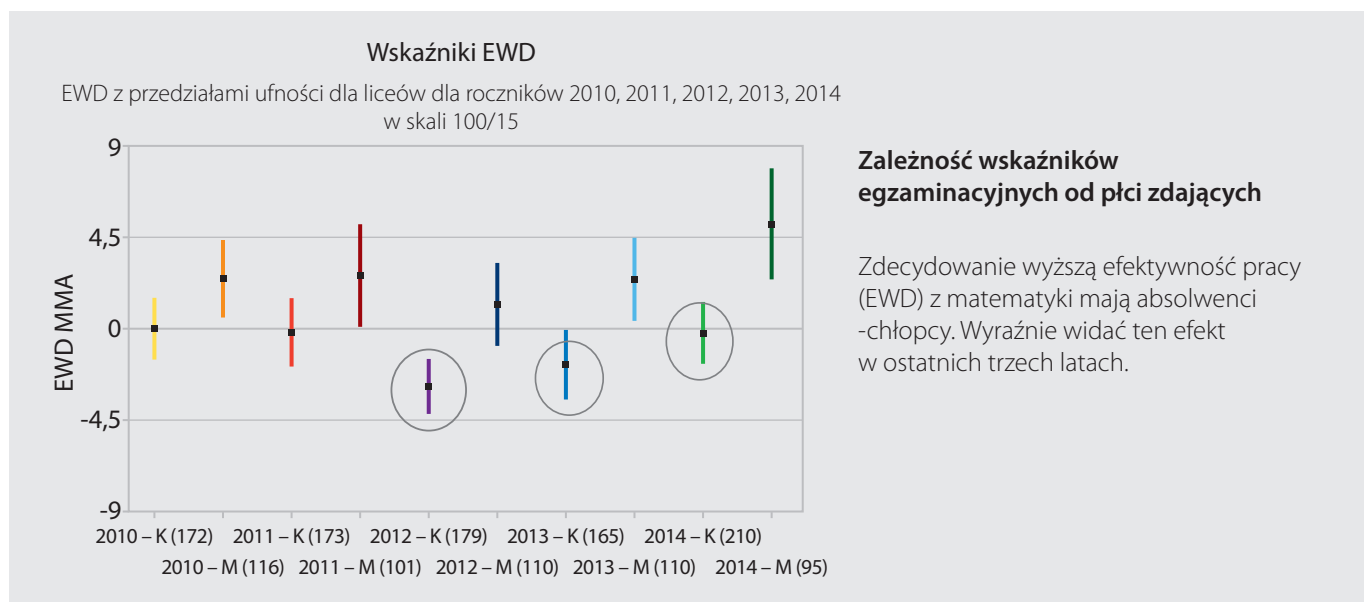
Ponadprzeciętną efektywność nauczania matematyki osiągnięto w oddziałach, w których matematyka była realizowana w zwiększonej liczbie godzin. W pozostałych klasach, ze standardową liczbą godzin, efektywność mierzona wskaźnikiem EWD była poniżej przeciętnej. W następnym etapie analiz sprawdzimy, czy w liceum A obserwujemy zróżnicowanie efektywności nauczania matematyki ze względu na płeć uczniów.

Rysunek 4.50. Płeć a efektywność nauczania matematyki



Wskaźnik EWD jest wyższy dla chłopców niż dla dziewcząt a z danych wynika, że w szkole maturę rozszerzoną z matematyki zdaje dwukrotnie więcej dziewcząt niż chłopców. Jednak analiza, w której uwzględniono zarówno podział na oddziały, jak i płeć maturzystów, pokazuje, że efektywność nauczania matematyki w szkole zdecydowanie silniej zależy od profilu klasy, do której uczęszcza maturzysta, niż od płci ucznia. Uczniowie wybierający profile niematematyczne osiągają bardzo niski wskaźnik EWD z matematyki. Analogiczną zależność można obserwować od roku 2010, równocześnie do liceum uczęszcza coraz więcej dziewcząt oraz maleje liczba chłopców.

Rysunek 4.51. Płeć a efektywność nauczania matematyki w latach 2010 – 2014



Od roku 2010 dziewczęta nie miały ani razu wskaźnika EWD powyżej zera, a w dwóch rocznikach (2012, 2013) osiągały poziom poniżej przeciętnej.

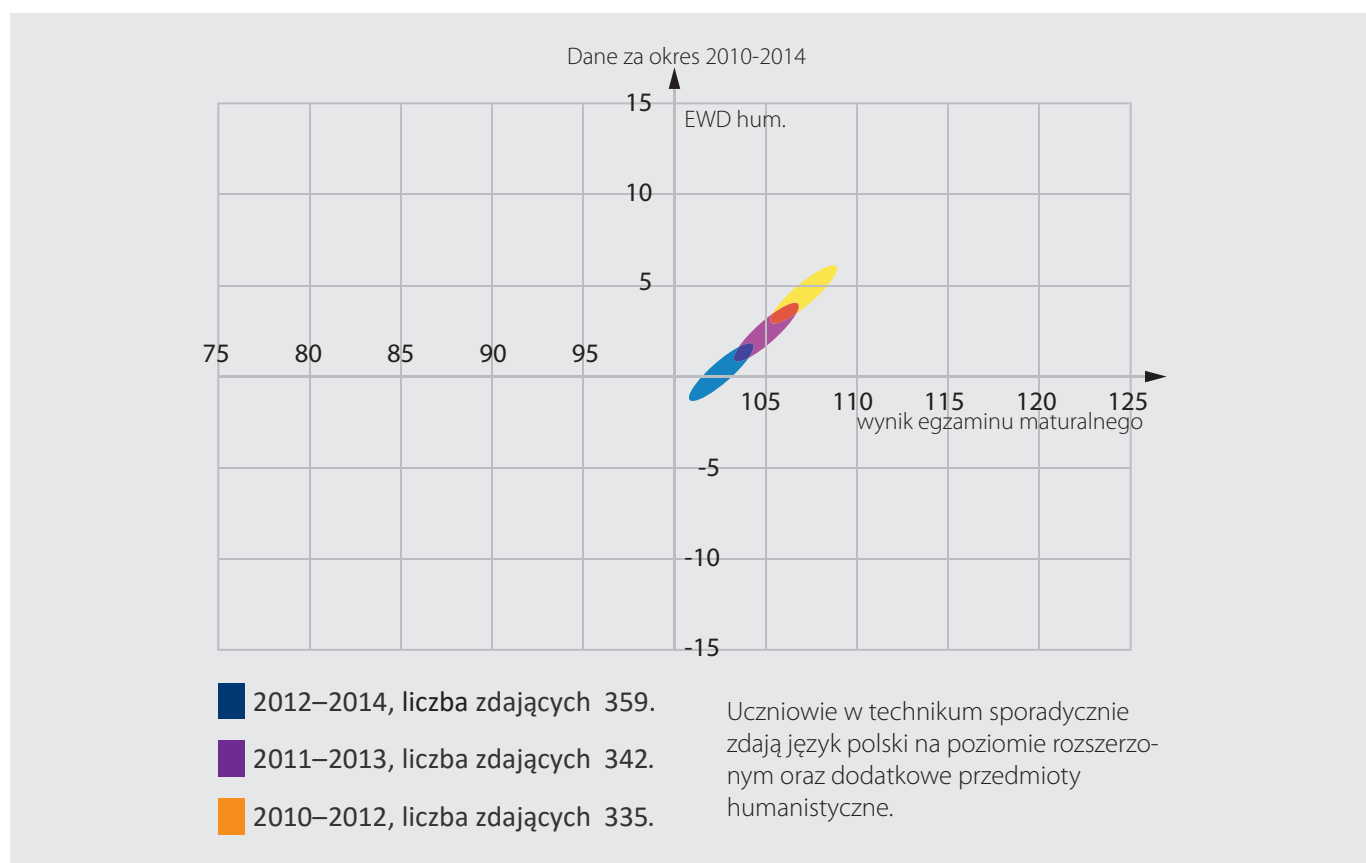
Uzyskane w liceum A wyniki analiz można podsumować w następujący sposób:

- Wyższe wyniki egzaminacyjne i efektywność (EWD) z historii i WOS-u na maturze niż z języka polskiego.
- Wysokie wartości wskaźników EWD z matematyki w klasach z dodatkowymi godzinami nauczania przedmiotu oraz znacząco niższa efektywność nauczania matematyki w oddziałach o profilach innych niż matematyczne.

Uzyskane wyniki to oczywiście początek, a nie koniec działań ewaluacyjnych.

Technikum LR. Szkoła znajduje się w mieście wojewódzkim. Była to szkoła o profilu energetyczno-elektrycznym, która w związku ze zmianami na rynku pracy modyfikuje swój profil na informatyczno-elektroniczny. Szkoła przygotowuje do pracy w zawodach technik elektryk, technik elektronik oraz technik informatyk.

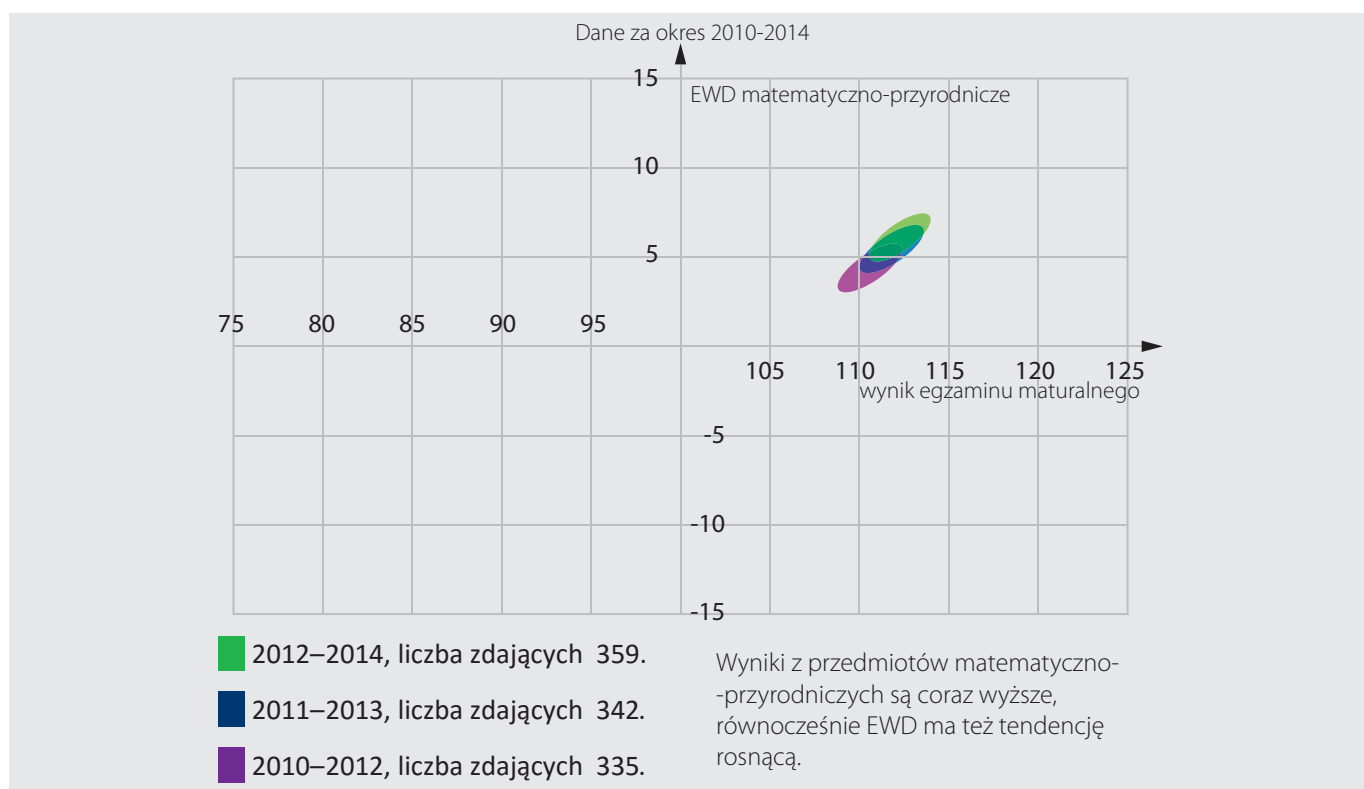
Analizę zaczniemy od przedmiotów humanistycznych.

Rysunek 4.52. Technikum LR. EWD w obszarze humanistycznym

Liczba osób zdających poszczególne przedmioty w latach 2012–2014

przedmiot	łącznie	poziom rozszerzony
j. polski	359	5
historia	2	0
WOS	2	0

W zakresie przedmiotów humanistycznych w technikum LR obserwujemy spadek efektywności nauczania i maturzyści osiągają coraz niższe wyniki egzaminacyjne.

Rysunek 4.53. Technikum LR. EWD w obszarze matematyczno-przyrodniczym



Z przedmiotów matematyczno-przyrodniczych technikum LR osiąga wysokie wyniki egzaminacyjne oraz pracuje z wysoką efektywnością. Dla okresu 2012–2014 przedział ufności dla technikum MP w obszarze osiągnięć matematyczno-przyrodniczych rozciąga się od około 112 do 114 punktów, czyli o mniej więcej $\frac{2}{3}$ odchylenia standardowego powyżej średniej krajowej dla techników. Wskaźnik EWD również jest wysoki i mieści się w przedziale (5; 8). Z danych uzyskanych od szkoły wiemy, że prawie wszyscy uczniowie podchodzą do egzaminów maturalnych i zawodowych i zdecydowana większość z sukcesem.

Jak widać w poniższej tabeli, przedmiotem najczęściej zdawanym w technikum LR jako przedmiot dodatkowy jest **informatyka**. W latach 2010–2012 zdawały go 44 osoby, a w okresie 2012–2014 już 82 osoby. Jest to zgodne z obecnym profilem technikum (informatyczno-elektroniczne). W tej szkole matematyka na poziomie rozszerzonym jest wybierana każdego roku przez ponad jedną czwartą maturzystów. Maturzyści z klasy elektronicznej stosunkowo często, jak na technika, zdają na maturze fizykę jako przedmiot dodatkowy.

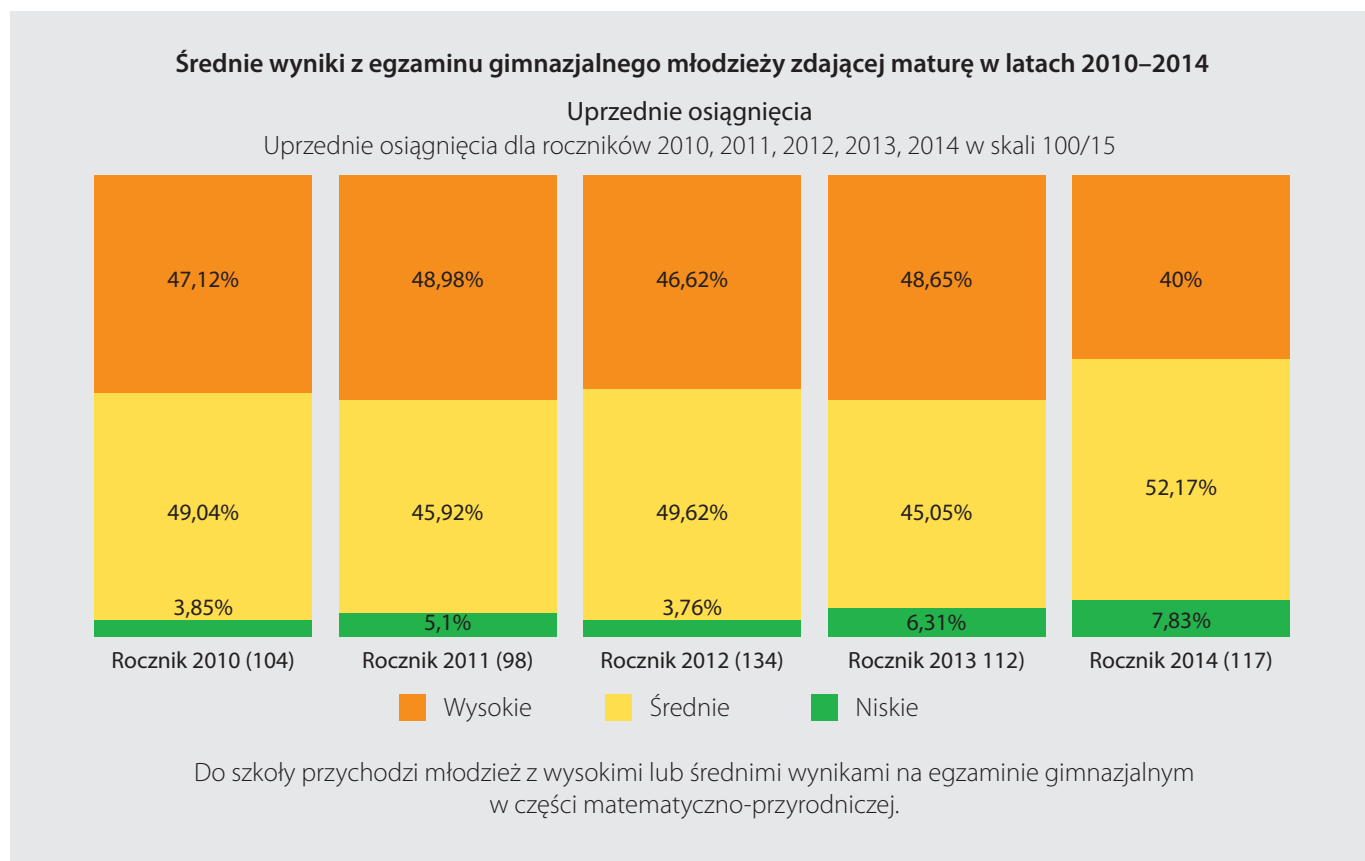
Tabela 4.8. Liczba osób zdających przedmioty matematyczno-przyrodnicze w technikum LR:

przedmiot	2010–2012		2011–2013		2012–2014	
	łącznie	poziom rozszerzony	łącznie	poziom rozszerzony	łącznie	poziom rozszerzony
matematyka	335	88	342	92	359	93
biologia	0	0	0	0	0	0
chemia	0	0	1	0	1	0
fizyka	18	8	23	9	21	7
geografia	35	8	24	7	17	3
informatyka	44	31	65	53	82	80

Dokładniejszą informację o efektywności nauczania w technikum LR dostarczą nam analizy wykonane przy użyciu programu Kalkulator EWD 100. Do analiz wykorzystano dane egzaminacyjne z lat 2010–2014.

Zacznijmy od analizy naboru do szkoły.

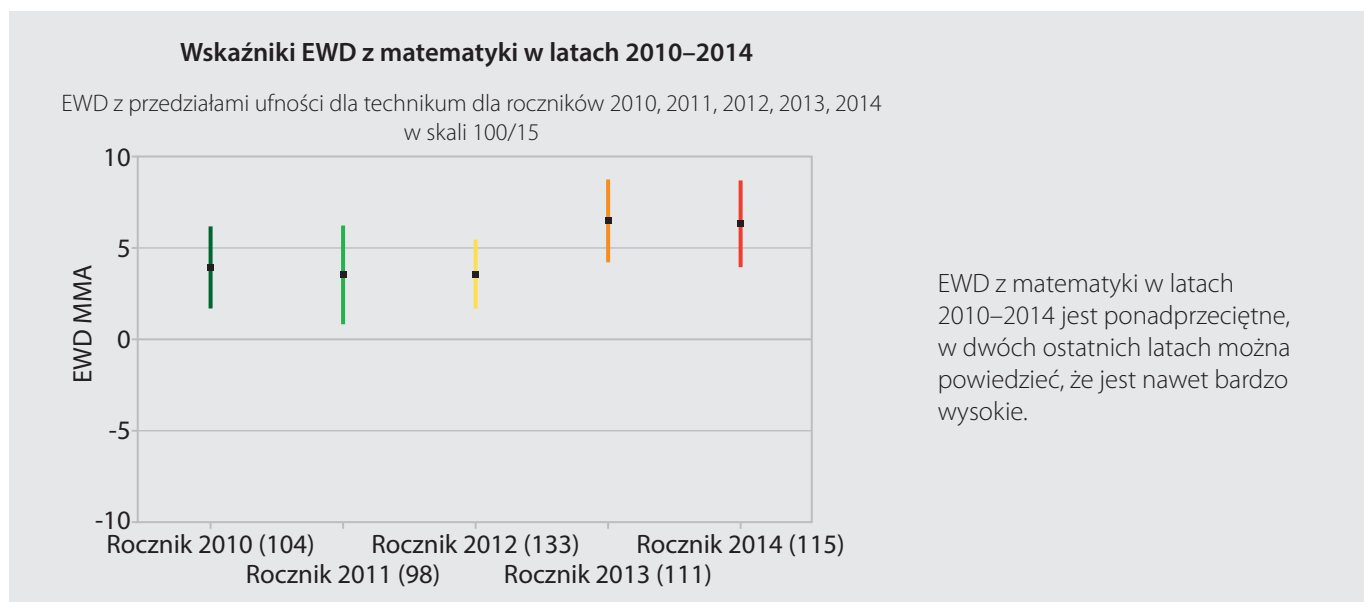
Rysunek 4.54. Technikum LR. Wyniki egzaminu gimnazjalnego w części matematyczno-przyrodniczej roczników przystępujących do matury w latach 2010–2014



Technikum LR jest szkołą o stabilnym naborze uczniów, mniej więcej połowa uczniów przychodzi z wynikami średnimi z egzaminu gimnazjalnego, druga połowa ma wyniki wysokie, nieliczni to absolwenci gimnazjów z niskim wynikiem z przedmiotów matematyczno-przyrodniczych.

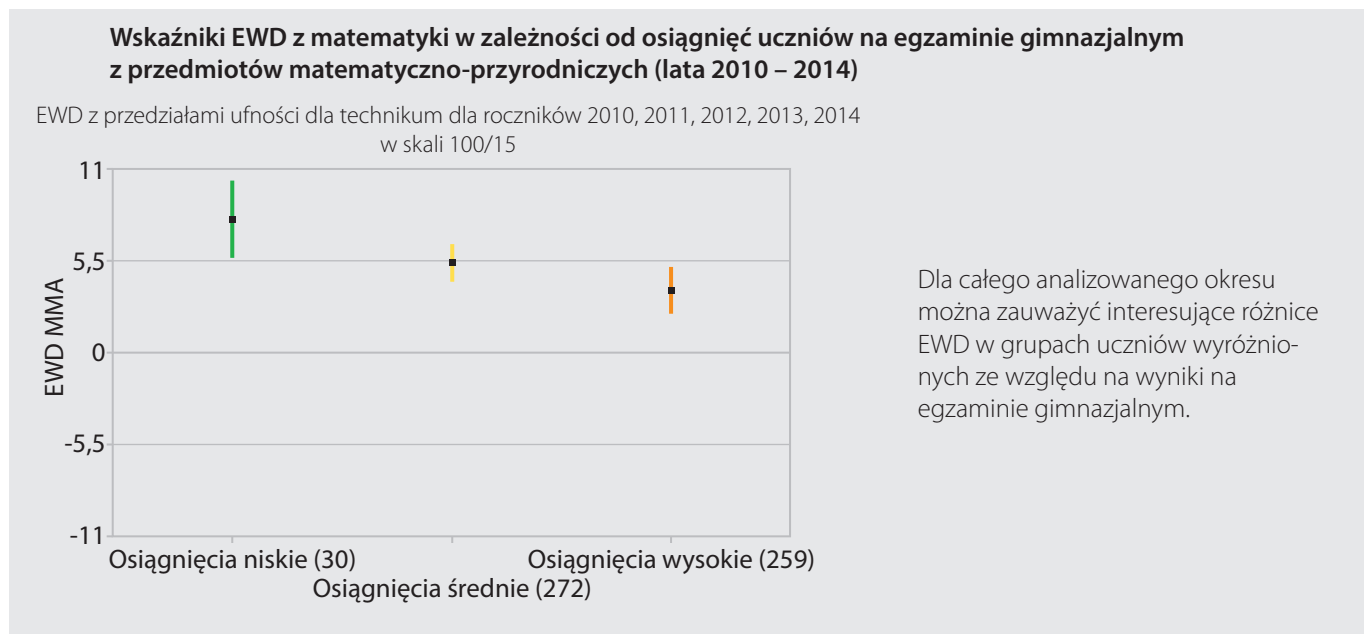
Wykorzystując Kalkulator EWD 100, prześledźmy bardziej wnikliwie efektywność nauczania matematyki.

Rysunek 4.55. Technikum LR. Jednoroczne wskaźniki EWD z matematyki w latach 2010–2014



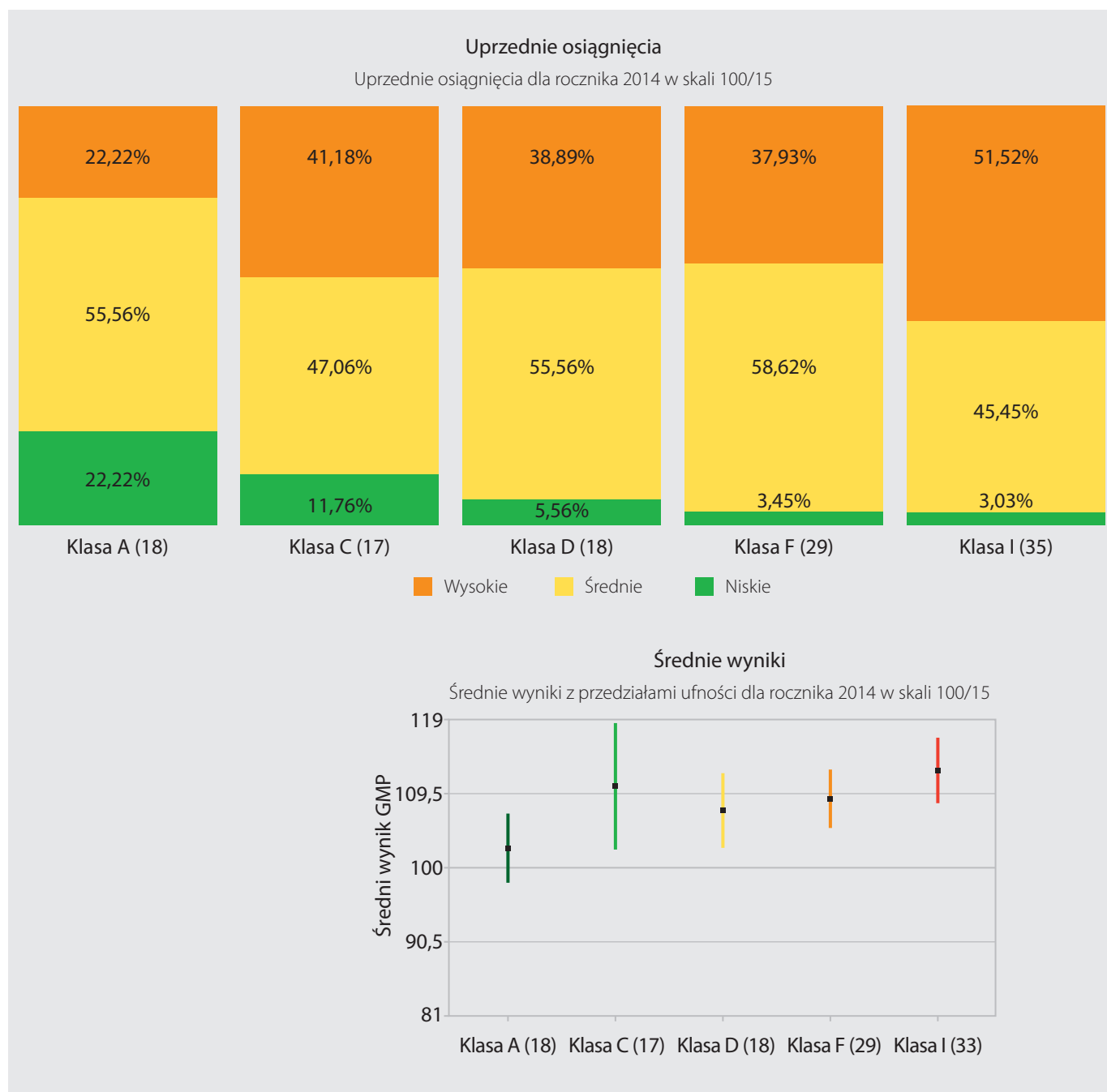
Wskaźnik EWD z matematyki w analizowanym okresie waha się od plus 3 do plus 7 punktów. Oznacza to ponadprzeciętną efektywność nauczania matematyki na tle wszystkich techników w kraju. Przyjrzyjmy się efektywności nauczania matematyki w podziale na trzy poziomy osiągnięć na egzaminie gimnazjalnym. Wyniki odpowiedniej analizy wykonanej w Kalkulatorze EWD pokazuje poniższy rysunek.

Rysunek 4.56. Technikum LR. Łączny wskaźnik EWD z matematyki w latach 2010–2014 dla uczniów o trzech poziomach uprzednich osiągnięć



W latach 2010–2014 wtechnikum LR zdawało maturę 30 absolwentów z niskimi wynikami na egzaminie gimnazjalnym i z nimi szkoła pracowała najefektywniej. Zajmijmy się teraz analizami różnic między oddziałami. Jak przydziela się uczniów do oddziałów ze względu na uprzednie osiągnięcia?

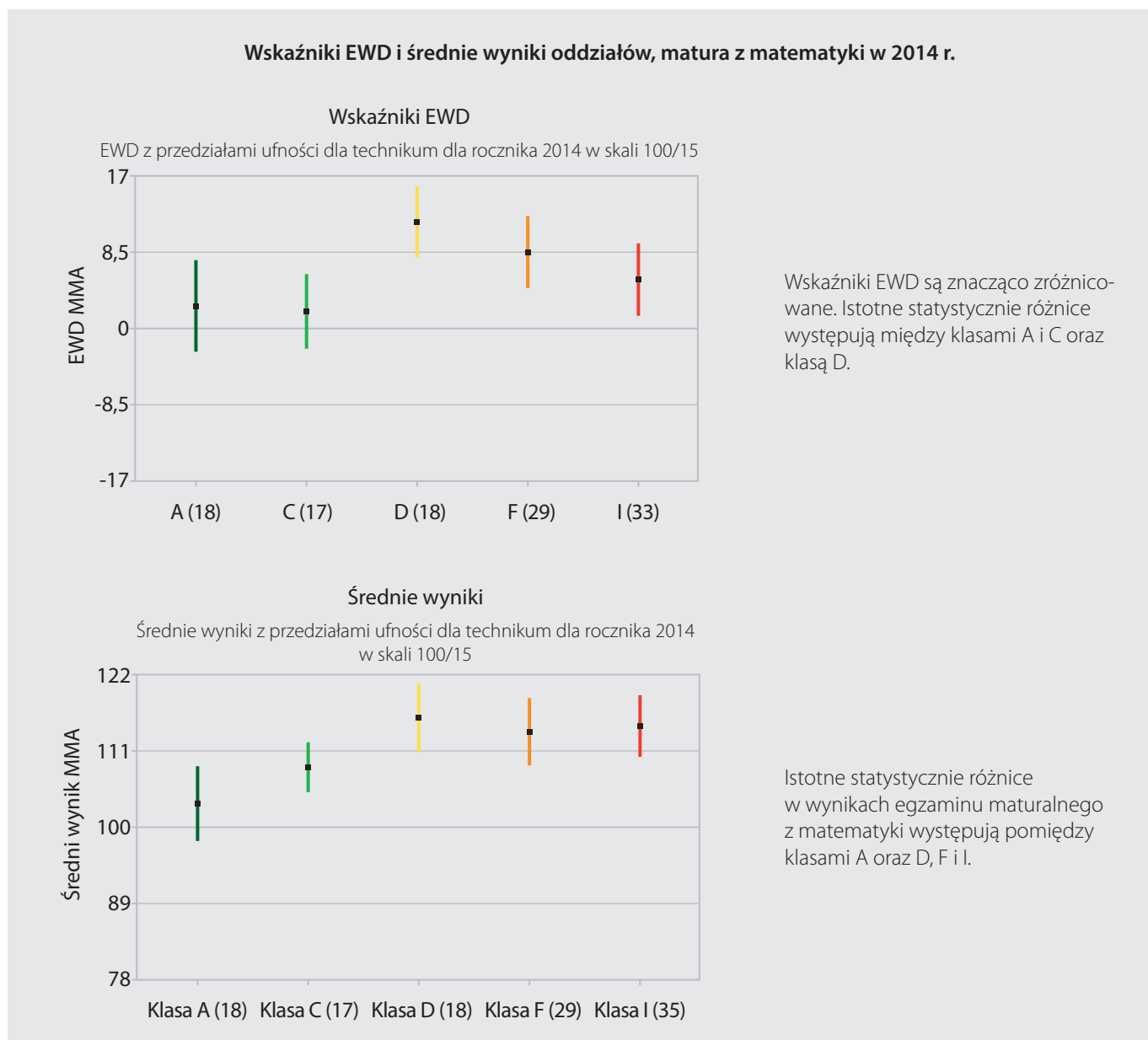
Rysunek 4.57. Technikum LR. Uprzednie osiągnięcia maturzystów w podziale na oddziały. Dwa sposoby analizy: rozkład procentowy trzech grup uczniów wyróżnionych ze względu na wyniki na egzaminie gimnazjalnym i średnie wyników egzaminu gimnazjalnego wraz z przedziałami ufności



Wykres *Upřednie osiągnięcia* dostarcza informacji o strategii przyjmowania młodzieży do klas I w technikum LR. Najwięcej uczniów o niskich wynikach przyjęto do klasy A (elektryk), w klasie tej rozkład uczniów jest porównywalny z rozkładem ogólnopolskim, w pozostałych klasach osiągnięcia uczniów z egzaminu gimnazjalnego są wyższe: w klasach D, F, I wyniki egzaminacyjne są wysokie i średnie. Porównywanie średnich wyników po egzaminie gimnazjalnym (matematyczno-przyrodniczym) pokazuje, że statystycznie różnią się one tylko pomiędzy klasami A oraz I, ponieważ pozostałe przedziały ufności dla klas nie są rozłączne.

Na koniec prezentacji przykładowych analiz EWD dla techników przyjrzyjmy się międzyoddziałowemu zróżnicowaniu EWD i wyników matury z matematyki.

Rysunek 4.58. Technikum LR. Wskaźniki EWD i średnie wyniki z matury z matematyki w podziale na oddziały



Uzyskane w technikum LR wyniki analiz można podsumować w następujący sposób:

- Coraz niższe wyniki egzaminacyjne i EWD z przedmiotów humanistycznych.
- Stabilne, ponadprzeciętne wartości wskaźnika EWD w obszarze matematyczno-przyrodniczym.
- Znaczące różnice w efektywności nauczania matematyki w dwóch klasach o profilu elektronicznym (klasy C i D), przy porównywalnych osiągnięciach młodzieży z egzaminu gimnazjalnego.

Ograniczenia i rozwój metody EWD dla liceów ogólnokształcących i techników

Wskaźniki EWD dla liceów ogólnokształcących i techników wyliczane są obecnie przy użyciu bardzo złożonego instrumentarium statystycznego, przede wszystkim w zakresie skalowania wyników matury. Choć generalnie można uznać je za trafnie obrazujące efektywność pracy szkół w zakresie kształcenia ogólnego, to jednak mają one pewne istotne ograniczenia. Niestety przewyższenie większości z nich wydaje się obecnie bardzo trudne, a niektórych po prostu niemożliwe.

Pierwsze istotne ograniczenie odnosi się specyficznie do techników i polega na nieuwzględnieniu efektów pracy szkoły w zakresie kształcenia zawodowego, co znacznie zmniejsza zainteresowanie

wskaźnikami EWD ze strony szkół tego rodzaju. Niestety kwestia ta wydaje się obecnie niemożliwa do rozwiązania.

Drugi problem to ograniczenie zestawu wskaźników jednorocznych, umożliwiających prowadzenie analiz wewnątrzszkolnych, jedynie do wskaźnika EWD w zakresie matematyki. Stworzenie wskaźników jednorocznych obejmujących kilka przedmiotów napotyka na problem przeliczania wyników poszczególnych części egzaminu na wartości ogólnych, kompozycyjnych miar osiągnięć. Schematy służące do takiego przeliczania byłyby bowiem bardzo złożone i trudne do zaimplementowania w Kalkulatorze EWD. Niemniej wielu użytkowników Kalkulatora EWD zgłasza zapotrzebowanie na przygotowanie wskaźników w zakresie języka polskiego. Wydaje się to postulat możliwy do spełnienia, przy wykorzystaniu podobnych narzędzi jak wykorzystywane obecnie do wyliczania wskaźników w zakresie matematyki, choć dodatkową komplikację stanowi fakt, że przy wyliczaniu miar osiągnięć uczniów należałoby również uwzględnić wybór tematu wypracowania. Niestety, ze względu na niepożądane cechy psychometryczne arkuszy maturalnych w zakresie języka polskiego, przede wszystkim ich niezbyt wysoką rzetelność i, co za tym idzie, niezbyt silny związek z wynikami egzaminu gimnazjalnego, wskaźniki te mogą cechować się znaczną niedokładnością. Działania zmierzające do ich przygotowania są obecnie w toku, a o ich efektach będzie można się przekonać najprawdopodobniej jesienią tego roku.

Trzeci problem, o którym warto wspomnieć, to fakt, że w modelowaniu EWD możemy uwzględnić wyniki tylko tych absolwentów szkół maturalnych, którzy przystępują do egzaminu dojrzałości. O ile w liceach ogólnokształcących dla młodzieży jest to problem marginalny, to w pozostałych typach szkół jest on znaczący. Niestety brak w bazach danych egzaminacyjnych informacji o absolwentach, którzy nie podchodzą do matury, uniemożliwia jakiegokolwiek analizy tego zagadnienia.

Trudno powiedzieć, jaki wpływ na własności modeli i wskaźników EWD będą mieć zmiany w formule matury 2015. Nie wprowadzają one co prawda żadnych elementów, które wymagałyby dodatkowego komplikowania modeli, jednak jak na razie otwarte pozostaje pytanie o własności psychometryczne egzaminu w nowej formule. Warto też zauważyć, że w związku ze zmianami w egzaminie gimnazjalnym, jakie zaszły w 2012 r., konieczne będzie przeanalizowanie kwestii, czy w przypadku wskaźników w zakresie języka polskiego i matematyki nie byłoby lepiej zawęzić miary osiągnięć „na wejściu” tylko do wyników testu odpowiednio z języka polskiego i matematyki.

Bibliografia

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

Braun, H. i Wainer, H. (2007). Value-added modeling. w Rao, C.R. i Sinharay, S. (red.), *Handbook of Statistics* 26: Psychometrics, s. 475–501. Amsterdam, Elsevier.

Conijn, J. M., Emons, W. H. M., i Sijtsma, K. (2014). Statistic Iz-Based Person-Fit Methods for Noncognitive Multiscale Measures. *Applied Psychological Measurement*, 38(2), 122–136.

Dolata, R. (red.) (2007). *Edukacyjna wartość dodana jako metoda oceny efektywności nauczania*. Warszawa: Centralna Komisja Egzaminacyjna.

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

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

- Dolata, R. (2008). *Szkoła-segregacje-nierówności*. Warszawa, Wydawnictwa Uniwersytetu Warszawskiego.
- Dolata, R., Hawrot, A., Hummeny, G., Jasińska, A., Koniewski M., Majkut P. i Żółtak T. (2013). *Trafność metody edukacyjnej wartości dodanej dla gimnazjów*. Warszawa, Instytut Badań Edukacyjnych.
- Dolata, R., Jasińska, A. i Modzelewski, M. (2012). Wykorzystanie krajowych egzaminów jako instrumentu polityki oświatowej na przykładzie różnicowania się gimnazjów w dużych miastach. *Polityka Społeczna*, nr tematyczny 1, s. 41–47.
- Dolata, R. i Pokropek, A. (2012). Czy warto urodzić się w styczniu? Wiek biologiczny a wyniki egzaminacyjne. w: B. Niemierko i M. K. Szmigiel (red.), *Regionalne i lokalne diagnozy edukacyjne: XVIII Krajowa Konferencja Diagnostyki Edukacyjnej, Wrocław, 21–23 września 2012 r.* Kraków: Grupa Tomami.
- Dolata, R., Hawrot, A., Humenny, G., Jasińska-Maciążek, A., Koniewski, M. i Majkut, P. (2014). *Kontekstowy model oceny efektywności nauczania po pierwszym etapie edukacyjnym*. Warszawa: Instytut Badań Edukacyjnych.
- Hanushek, E. A. (2003). The Failure of Input-based Schooling Policies. *Economic Journal* 113, February, s. F64–F98.
- Jacob, B. A. i Levitt, S. D. (2003a). Catching cheating teachers: The results of an unusual experiment in implementing theory. *National Bureau of Economic Research*.
- Jacob, B. A. i Levitt, S. D. (2003b). Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating. *The Quarterly Journal of Economics*, 118(3), 843–877.
- Jasińska, A. i Modzelewski, M. (2014). Testy osiągnięć szkolnych TOS 3 jako przykład narzędzia skonstruowanego z wykorzystaniem modelu Rascha. *Edukacja*, 2(127), 85–107.
- Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education*, 16(4), 277–298.
- Karwowski, M. (Red.). (2013). *Ścieżki rozwoju edukacyjnego młodzieży – szkoły pogimnazjalne*. Warszawa: Wydawnictwo Instytutu Filozofii i Socjologii PAN.
- Korobko, O. B., Glas, C. A. W., Bosker, R. J. i Luyten, J. W. (2008). Comparing the difficulty of examination subjects with item response theory. *Journal of Educational Measurement*, 45(2), 139–157.
- Markman, J. M., Hanushek, E. A., Kain, J. F. i Rivkin S. G., (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, vol. 18(5), s. 527–544.
- McCafrey, D. F., Lockwood, J. R., Koretz, D. M. i Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, Kalifornia: RAND.
- McCafrey, D. F., Lockwood, J. R., Koretz, D. M., Louis, T. A. i Hamilton, L. S. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29, 67–101.
- Meyer, R. (1997). Value-Added Indicators of School Performance: A Primer. *Economics of Education Review*, Vol. 16, No.3, s. 283–301.

- OECD. (2008). *Measuring improvements in learning outcomes: best practices to assess the value-added of schools*. Paryż: OECD.
- Pokropek, A. (2011). Matura z języka polskiego. Wybrane problemy psychometryczne. W: B. Niemierko i M. K. Szmigiel (red.), *Ewaluacja w edukacji: koncepcje, metody, perspektywy*. Kraków: Polskie Towarzystwo Diagnostyki Edukacyjnej.
- Pokropek, A. (red.). *Modele cech ukrytych w badaniach edukacyjnych, psychologii i socjologii. Teoria i zastosowania*. Warszawa: Instytut Badań Edukacyjnych.
- Pokropek, A. i Żółtak, T. (2012a). *Trzyletni wskaźnik egzaminacyjny*. Dokumentacja techniczna. Wersja 2.0. <http://2013.ewd.edu.pl/downloads/Dokumentacja%20techniczna%20v2.0.pdf>
- Pokropek, A. i Żółtak, T. (2012b). Nowe modele jednorocznej EWD. W: K. Szmigiel i B. Niemierko (red.), *Regionalne i lokalne diagnozy edukacyjne: XVIII Krajowa Konferencja Diagnostyki Edukacyjnej, Wrocław, 21–23 września 2012 r.* Kraków: Grupa Tomami.
- Raudenbush, S. W. i Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, 20, 307–335.
- Rubin, D. B., Stuart E. A. i Zanutto, E. L. (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics* 29:103–116.
- Schagen, I. i Hutchinson, D. (2003). Adding value in educational research – the marriage of data and analytical power. *British Educational Research Journal*, vol. 29, no. 5.
- Stożek, E. (2008). *Analiza wyników egzaminów zewnętrznych z wykorzystaniem metody EWD w ewaluacji wewnętrznej*. http://2013.ewd.edu.pl/materiały-szkoleniowe/broszura_ewd.pdf
- Stożek, E. (2009). EWD w ręku dyrektora szkoły. *Dyrektor Szkoły* 12(192)/2009, s.I-VIII.
- Stożek, E. (2010). *Analiza wyników egzaminów zewnętrznych z wykorzystaniem metody EWD w ewaluacji wewnętrznej. Materiał pomocniczy dla dyrektorów i nauczycieli gimnazjów* http://2013.ewd.edu.pl/materiały-szkoleniowe/broszura_2010.pdf
- Żółtak, T. (2013a). *Statystyczne modelowanie wskaźników edukacyjnej wartości dodanej – podsumowanie polskich doświadczeń*. Analizy IBE/02/2013.
- Żółtak, T. (2013). Gimnazjalne wskaźniki EWD. w: R. Dolata (red.), *Trafność metody edukacyjnej wartości dodanej. Raport podsumowujący wyniki badania podłużnego w gimnazjach*. Warszawa, Instytut Badań Edukacyjnych w Warszawie, s.90-102