

dr Roman Dolata

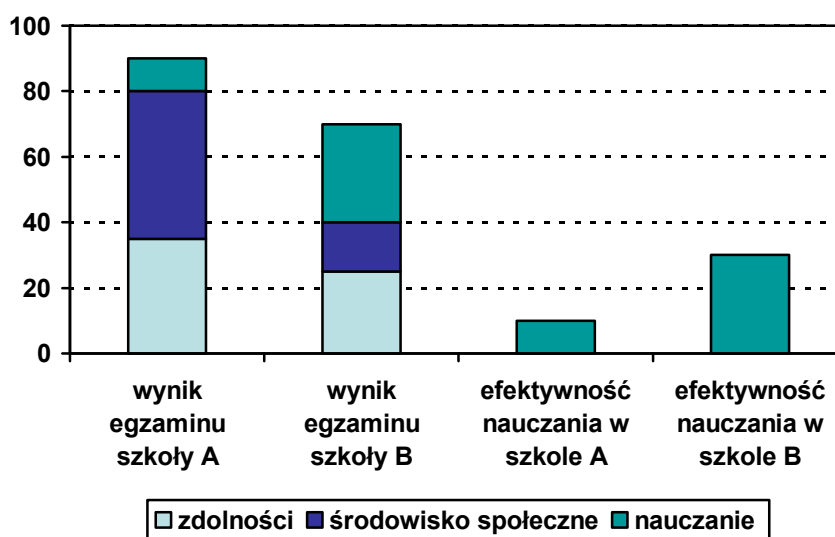
Edukacyjna wartość dodana, czyli jak wykorzystywać wyniki egzaminów zewnętrznych do oceny efektywności nauczania

Wprowadzony w 2002 roku system egzaminów zewnętrznych dostarcza obiektywnych danych o osiągnięciach szkolnych. Dane te mogą być wykorzystywane do oceny efektywności nauczania. Jednak by oceny efektywności nauczania formułowane na podstawie wyników egzaminów były adekwatne do realnego wkładu pracy szkoły, należy uwzględnić fakt, że osiągnięcia szkole ucznia nie zależą tylko od czynników związanych z jakością nauczania.

Wynik sprawdzianu/egzaminu jest uwarunkowany trzema grupami czynników:

- indywidualnymi, takimi jak zdolności ucznia i uprzednie osiągnięcia szkolne,
- społecznymi, takimi jak kapitał kulturowy i społeczny rodziny czy wpływ grupy rówieśniczej,
- szkolnymi, takimi jak kwalifikacje i zaangażowanie nauczycieli, metody nauczania czy warunki nauczania.

Jeżeli chcemy przeciętny wynik egzaminu dla danej szkoły interpretować jako miarę efektywności nauczania, musimy wyeliminować wpływ czynników, których szkoła nie może skutecznie kształtować. Poniższy wykres pozwala prześledzić, na jaki błąd narażamy się porównując szkoły na podstawie nieprzetworzonych wyników egzaminu zewnętrznego.



Wykres 1. Czynniki wpływające na wyniki egzaminu a ocena efektywności nauczania

W hipotetycznym porównaniu na powyższym wykresie przewaga szkoły A nad szkołą B wynika z bardziej korzystnego układu czynników indywidualnych i środowiskowych. Jeżeli z wyniku wytrącimy to, co zależy od zdolności i społecznego środowiska, ocena musi ulec zmianie – to szkoła B lepiej uczy w zakresie sprawdzanym przez egzamin. Metodą, która

pozwala tak przetworzyć wyniki egzaminów, by znacząco ograniczyć wpływ czynników indywidualnych i społecznych, jest edukacyjna wartość dodana.

1. Pojęcie edukacyjnej wartości dodanej

Pojęcie EWD wywodzi się z ekonomii, gdzie - w pewnym uproszczeniu - wartość dodana oznacza przyrost wartości dóbr w wyniku danego procesu produkcyjnego. Analogicznie **edukacyjną wartość dodaną można zdefiniować jako przyrost wiedzy uczniów w wyniku danego procesu edukacyjnego**¹. Ogólna idea edukacyjnej wartości dodanej była konkretyzowana na wiele sposobów. Wydaje się, że można wyróżnić dwa podstawowe rozumienia.

Proste rozumienie edukacyjnej wartości dodanej (Dolata 2005), odnosi się do idei postępu uczniów w nauce, czyli uwzględniania w szacunkach efektywności nauczania w szkołach początkowego stanu osiągnięć szkolnych uczniów. Wyróżniamy dwa typy prostych wskaźników wartości dodanej. Pierwszy typ to wskaźniki bezwzględne. Wymagają one, by pomiar osiągnięć na różnych etapach kształcenia był wykonywany za pomocą tych samych skal pomiarowych. Jeżeli na przykład dysponowalibyśmy skalą do pomiaru rozumienia przez uczniów czytanych tekstów, to w dowolnym momencie procesu kształcenia moglibyśmy szacować wartość dodaną w tam zakresie. W Polsce nie dysponujemy wynikami takich pomiarów.

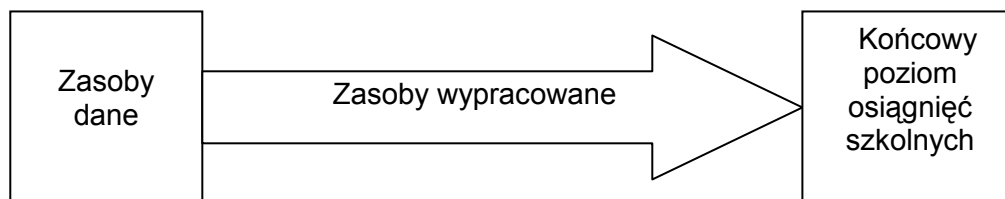
Drugi typ to wskaźniki względne. Przyjmujemy, że wynik na egzaminie na niższym szczeblu jest ogólną miarą potencjału edukacyjnego ucznia i na jej podstawie prognozujemy wynik na egzaminie kolejnego szczebla. Prognoza polega na określeniu, przy użyciu metod statystycznych, jaki przeciętnie wynik na egzaminie gimnazjalnym uzyska uczeń o konkretnym wyniku ze sprawdzianu. Wynik faktycznie uzyskany przez ucznia na egzaminie gimnazjalnym odnosimy do wartości oczekiwanej (prognozowanej) i w ten sposób otrzymujemy oszacowanie wartości dodanej na danym etapie kształcenia. Średnia tak obliczonych wskaźników dla danej szkoły jest wskaźnikiem efektywności nauczania w zakresie sprawdzanym przez egzamin.

Choć w prostych wskaźnikach EWD bezpośrednio bierze się pod uwagę tylko uprzednie osiągnięcia szkolne, to pośrednio model ten uwzględnia też zdolności i czynniki środowiskowe. Dzieje się tak, ponieważ uprzednie osiągnięcia szkolne są uwarunkowane tymi samymi czynnikami, co wynik końcowy.

Przejdźmy do rozbudowanego rozumienia edukacyjnej wartości dodanej², w którym horyzont analizy zostaje poszerzony przez wprowadzenie pojęcia zasobów danych szkoły. Schemat rozumowania przedstawia poniższy rysunek.

¹ Prawdopodobnie po raz pierwszy pojęcie edukacyjnej wartości dodanej pojawiło się w połowie lat 70. jako krytyczna kontynuacja idei rozliczalności szkół (*school accountability*). Z perspektywy czasu niektórzy badacze uważają pojęcie edukacyjnej wartości dodanej za najważniejsze narzędzie analityczne, jakie w naukach pedagogicznych pojawiło się w ostatnich 20 latach (Schagen, Hutchison 2003).

² Rozbudowane rozumienie edukacyjnej wartości dodanej odwołuje się do tradycji badawczej znanej pod nazwą *education production functions* a łączonej z nazwiskiem E. Hanusheka (1986).



Rysunek 1. Rozbudowane rozumienie edukacyjnej wartości dodanej

Lista składników zasobów danych może wyglądać następująco:

- zasoby indywidualne uczniów: dotychczasowe osiągnięcia szkolne, kapitał kulturowy, ekonomiczny i społeczny rodziny ucznia.
- zasoby grupowe: osiągnięcia szkolne rówieśników, ich kapitał kulturowy, ekonomiczny i społeczny, cechy kultury szkoły determinowane jej składem społecznym.
- zasoby instytucjonalne: nakłady finansowe, infrastruktura.

Zasoby dane rzutują na końcowy poziom osiągnięć szkolnych, ale znajdują się poza kontrolą szkoły (środki finansowe, infrastruktura) lub nie chcemy, by szkoła je kontrolowała (kontrola zasobów indywidualnych uczniów przez ich selekcję na wejściu).

Zasoby wypracowane powstają w wyniku dydaktycznej i wychowawczej pracy szkoły, są rezultatem całości praktyk edukacyjnych tworzących kulturę pedagogiczną szkoły. Oceniając efektywność nauczania w szkole należy wyeliminować wpływ zasobów danych, by końcowy poziom osiągnięć szkolnych mógł być przypisany tylko zasobom wypracowanym. Wskaźnik efektywności nauczania w danej szkole powinien być tak skonstruowany, by jego wartość dała się interpretować jako osiągnięcia szkolne w sytuacji całkowitej równości między szkołami po stronie zasobów danych.

W polskich realiach jesteśmy w stanie szacować EWD za pomocą prostych modeli uwzględniających przede wszystkim uprzednie osiągnięcia ucznia. Tworzenie rozbudowanych modeli szacowania EWD jest – przynajmniej na razie – niemożliwe z powodu braku potrzebnych, wysokiej jakości danych.

2. Po co nam edukacyjna wartość dodana?

Metoda edukacyjnej wartości dodanej może być rozpatrywana w ogólnym kontekście polityki poprawiania jakości oświaty przez mechanizmy rozliczania szkół na podstawie wymiernych wskaźników, a także na tle problemów związanych z rynkiem usług edukacyjnych i konkurencją szkół.

Zastosowanie EWD w oświacie służyć może osiągnięciu następujących celów:

- zmniejszeniu roli surowych wyników sprawdzianu/egzaminów jako jedynej miary jakości nauczania,
- łagodzeniu negatywnych skutków konkurencji między szkołami, czyli przeciwdziałaniu selekcji na progu szkoły i efektowi zwanemu „spijaniem śmietanki”,

- zmniejszeniu poczucia ciągłej frustracji nauczycieli pracujących z zaangażowaniem w środowiskach o niekorzystnych cechach społeczno - ekonomicznych oraz pozbawieniu wielu nauczycieli argumentu typu „ale z kim ja muszę pracować” usprawiedliwiającego w najprostszy, obronny sposób niepowodzenia uczniów.

Cele te są z sobą powiązane, ale dla uporządkowania wyводу omówione zostaną oddzielnie.

2.1. Zmniejszenie roli surowych wyników sprawdzianu/egzaminów jako jedynej miary jakości nauczania

Każda dojrzała polityka doskonalenia oświaty musi wykorzystywać zestaw dobrze zdefiniowanych wskaźników jakości nauczania. Niektóre z nich mogą mieć charakter zewnętrzny oraz ilościowy i wykorzystywać wyniki obiektywnych pomiarów, inne - charakter wewnętrzny oraz jakościowy i bazować na całościowych metodach diagnozy, popartych głębszym wglądem w praktyki edukacyjne danej szkoły.

System egzaminacyjny dostarcza obiektywnych danych ilościowych o poziomie osiągnięć uczniów. Waga tych informacji rośnie z perspektywy mechanizmów ewaluacji zewnętrznej szkoły. Coraz częściej średni wynik szkoły uzyskany na egzaminie staje się kryterium jej oceny przez otoczenie społeczne.

Jednakże, jak wskazano we wstępnej części artykułu, traktowanie średniego wyniku szkoły jako miary efektywności nauczania jest błędem, ponieważ zależy on także od czynników, na które szkoła nie ma wpływu. Zatem używanie tego rodzaju nieprzetworzonych danych jako miary efektywności musi prowadzić do nieadekwatnych ocen. System oświaty musi zaproponować lepszą alternatywę. Nie wystarczy komentarz do wyników egzaminów przestrzegający przed pochopnym ich wykorzystywaniem do oceny szkoły bez rozpoznania kontekstów kształcenia.

Edukacyjna wartość dodana jest rozwiązaniem konkurencyjnym, ciągle niedoskonałą, ale zdecydowanie bardziej wartościową metodą oceny efektywności nauczania w szkole/szkołach. Wprowadzenie do obiegu informacyjnego kolejnego miernika jakości pozwoli formułować na każdym szczeblu zarządzania bardziej dojrzałe oceny i podejmować bardziej racjonalne decyzje oświatowe.

EWD realnie zacznie wpływać na oświatę, gdy będzie kompetentnie wykorzystywane przez nauczycieli i dyrektorów, nadzór pedagogiczny, organy prowadzące, media oraz rodziców.

2.2. Łagodzenie negatywnych skutków konkurencji między szkołami

Badania nad funkcjonowaniem systemów oświatowych, w których szkoły rywalizują o uczniów i są finansowane zależnie od ich liczby (tzw. quasi-rynkowe mechanizmy w oświacie) wskazują, że poza ewentualnymi korzyściami płynącymi z tych rozwiązań, pojawiają się zagrożenia.

Najczęściej wskazuje się na silne różnicowanie się szkół. Najslabsze szkoły pod względem wyników uczniów na egzaminach końcowych nie są eliminowane, ale z różnych powodów z coraz mniejszą liczbą uczniów trwają na rynku. Najlepsze - wykorzystując mechanizm zwany przez Anglosasów „spijaniem śmietanki”, umacniają swoje pozycje. Różnicowanie to owocuje nasileniem segregacji społecznych w oświacie i zagraża ważnemu celowi polityki oświatowej jakim jest równość szans edukacyjnych (por. Ball 1993, Gorard, Fitz 2000, Dolata 2002).

Również w polskiej oświacie takie zjawisko jest obserwowane. Niż demograficzny oraz liberalizacja zasady rejonizacji, szczególnie w dużych miastach, doprowadziły do uruchomienia konkurencji między szkołami. W dużych miastach gimnazja coraz bardziej się polaryzują, rosłą różnice w wynikach egzaminów między szkołami.

Opisywany proces wymaga reakcji. Jeżeli zrezygnujemy ze stosowania narzędzi polityki nakazowej - a wydaje się, że przynosi ona dostatecznie dużo skutków ubocznych, by ją porzucić - parametryczne zarządzanie oświatą z użyciem wskaźników EWD wydaje się potencjalnie skutecznym instrumentem osłabiania procesów różnicowania w oświacie i realizacji postulatu działań na rzecz spójności społecznej.

Ocena szkoły na podstawie edukacyjnej wartości dodanej powinna zmniejszyć nacisk na selekcję uczniów na wejściu do szkoły, zaś ocena nauczycieli według tej miary powinna podnieść atrakcyjność pracy w szkołach funkcjonujących w mniej korzystnych warunkach społecznych. EWD z pewnością nie zahamuje w krótkiej perspektywie czasowej segregacji społecznych i ekonomicznych w skali makrospołecznej, ale w dalszej - może przyczynić się do rewitalizacji wielu szkół dotąd nisko plasujących się w rankingach tworzonych na podstawie uśrednionych wyników egzaminów zewnętrznych.

2.3. Zmniejszenie poczucia frustracji nauczycieli pracujących w środowiskach o niekorzystnych cechach społeczno – ekonomicznych

Znaczne zróżnicowanie cywilizacyjne Polski sprawia, że duże grupy nauczycieli pracują w trudnych środowiskach społecznych. Wprowadzenie egzaminów zewnętrznych unaocznilo różnice w wynikach końcowych szkół. Istotnym problemem psychologicznym jest włączenie tych wyników w swoisty rachunek zawodowych sukcesów i porażek decydujący o tym, czy nauczyciel spostrzega swoją pracę jako satysfakcjonującą. Jeżeli nauczyciel z zaangażowaniem pracujący w trudnym środowisku przez kolejne lata dostawał informację zwrotną, że jego uczniowie marnie wypadali na egzaminach, to z pewnością poczucie satysfakcji – jeden z ważnych czynników zawodowej efektywności – musiało na tym cierpieć.

Edukacyjna wartość dodana może zmienić ten stan rzeczy. Być może wielu nauczycieli po raz pierwszy, od kiedy zaistniały egzaminy zewnętrzne, otrzyma jednoznaczny komunikat: *dobrze pracuję, osiągnięcia moich uczniów są na miarę ich możliwości, wyniki jakie uzyskują są lepsze, niż innych nauczycieli pracujących w podobnych warunkach.*

Z drugiej strony nauczyciel, którego uczniowie uzyskują wyniki poniżej swoich możliwości (czyli EWD jest niska), nie będą mogli sięgać po proste wytłumaczenie: *no tak, moi uczniowie są niezdolni i słabi, nikt by ich więcej nie nauczył, bo to nie jest możliwe.* Odcięcie od takiego obronnego tłumaczenia niskich wyników uczniów może uruchomić zdrową konkurencję i próbę poprawy choćby poprzez zmianę metod nauczania, czy motywowania uczniów do aktywnego udziału w procesie edukacji. Równocześnie EWD da mocny argument tym, którzy nauczycieli oceniają.

3. Czy edukacyjna wartość dodana jest niezawodnym panaceum?

Krytyczne argumenty formułowane wobec EWD można poznać śledząc dyskusję prowadzoną w Stanach Zjednoczonych w związku z jednym z najbardziej rozbudowanych systemów wartości dodanej a mianowicie TVAAS³. Za głównego twórcę systemu uważany

³ Tennessee Value-Added Assessment System

jest W. Sanders. Przyjął on założenie, że można oddzielić wpływ czynników pozaszkolnych od wpływu jakości nauczania śledząc indywidualne wzory wykonania przez uczniów standaryzowanych testów osiągnięć szkolnych⁴. Oczywiście taki system szacowania EWD może istnieć tylko wtedy, gdy uczniowie dostatecznie często są oceniani przy pomocy standaryzowanych testów – ten warunek jest spełniony w Stanach Zjednoczonych.

Na podstawie dokonanych analiz Sanders i współpracownicy doszli do wniosku, że problem nierówności edukacyjnych da się rozwiązać metodami czysto edukacyjnymi. Rozumowanie ich było następujące: jeżeli różnica w przeciętnych osiągnięciach szkolnych grup uczniów z niższej i średniej klasy społecznej, wynosi 25 punktów testowych, a różnica między efektywnością nauczycieli o przeciętnej (50.centyli) i wysokiej (np.: 90.centyli) efektywności wynosi również 25 punktów testowych, to należy efektywność nauczycieli pracujących z uczniami z niższej klasy społecznej podnieść do poziomu efektywności charakterystycznej dla 90 centyla. Ten skok efektywności przełoży się na zniesienie obserwowanej nierówności edukacyjnej.

Wyobraźmy sobie, że informacje o efektywności poszczególnych nauczycieli i przeciętnej efektywności danej szkoły są upubliczniane a rodzice są zachęceni do ich wykorzystywania przy wyborze szkoły. W takiej sytuacji należałoby się spodziewać odpływu uczniów z klas i szkół o niższej efektywności do nauczycieli i placówek o wyższej efektywności, co przełoży się na wzrost osiągnięć całej populacji. Jeżeli odpływ ten dodatkowo uruchomiłby rywalizację, a więc doprowadziłby do wzrostu zaangażowania i podwyższenia kwalifikacji przez słabszych nauczycieli, to - w połączeniu z *bonusem* mniej licznych klas (efekt odpływu) – mogłoby to zaowocować wzrostem efektywności i odzyskiwaniem rynku przez szkoły, w których notowano odpływ uczniów.

Przeciw trafności tego typu argumentacji można sformułować wiele zastrzeżeń. R. Rothstein (2004) zwraca uwagę na kilka istotnych problemów.

Po pierwsze, TVAAS posługując się pojęciem efektywności zbudowanym na metodzie wartości dodanej nie dostarcza informacji, jakie praktyki edukacyjne za tę efektywność odpowiadają. Efektywność to wysokie wyniki w testach, wysokie wyniki w testach to efektywność. Co zmienić w swoim działaniu, by podnieść swą efektywność? To pytanie pozostaje bez odpowiedzi.

Cytowany R. Rothstein twierdzi, że wiara w możliwość dokonania w krótkim czasie znaczącego rozwoju umiejętności nauczycieli jest nierealistyczna. Naukom pedagogicznym bardzo daleko do dokonania przełomowych odkryć w tym zakresie i nie ma żadnych podstaw by wierzyć, że można uzyskiwać spektakularne zmiany w krótkim czasie.

Lepsze szkoły mają skończone zasoby i ich możliwości absorpcji uczniów z gorszych szkół szybko się wyczerpią. A jaki mechanizm zapewni, że nowe szkoły, nowi nauczyciele będą znacząco lepsi, od tych których mechanizm rywalizacji wyrzuci z rynku?

Ponadto, i ten argument wydaje się najważniejszy, TVAAS - jak i inne metody szacowania wartości dodanej - opiera się na założeniu, że nie ma interakcji między kapitałem kulturowym ucznia a skutecznością różnych sposobów nauczania. Jeżeli coś jest dobre dla ucznia X, dobre będzie dla wszystkich uczniów, jeżeli coś przynosi owoce w klasie IIIa, to będzie efektywne dla wszystkich klas. Założenie to jest w najwyższym stopniu wątpliwe.

⁴ Na przykład, jeżeli dany uczeń przez kilka kolejnych lat otrzymywał na testach raczej słabe wyniki w okolicy 30 centyla, (skala centylowa informuje nas, jaki procent uczniów uzyskało wynik słabszy od danego; w tym wypadku 30% uczniów wypadło słabiej, czyli 70% lepiej), by nagle osiągnąć 60 centyl, i ta gwałtowna zmiana pokrywa się ze zmianą nauczyciela, to przyrost ten można przypisać efektywności nauczania.

Istnieje wiele dowodów na rzecz hipotezy zgodności kulturowej, która mówi, że nauczanie szkolne jest najbardziej efektywne wtedy, gdy metody oddziaływania stosowane przez nauczyciela są zgodne z wzorcami uczenia się znanymi dziecku z doświadczeń kulturowych (Vasta, Haith, Miller 1995). Jeżeli hipoteza zgodności kulturowej jest trafna, to zmiana szkoły przez dziecko z rodziny o innych wzorcach kulturowych niż preferowane w nowej, „bardziej efektywnej”, wcale nie musi przynieść podwyższenia jego osiągnięć.

Z powyższych rozważań wyraźnie widać, że EWD nie jest żadnym niezawodnym panaceum na problemy współczesnej oświaty. Racjonalnie budowane przez nadzór pedagogiczny wskaźniki jakości pracy szkoły mogą uruchomić procesy poprawy efektywności nauczania. Jednak bez wiedzy nauczycieli, jak skutecznie nauczać, może okazać się to niemożliwe.

Ocena szkół ze względu na EWD może osłabić negatywne skutki rywalizacji między szkołami, ale może się okazać narzędziem zbyt słabym.

4. Zastosowanie EWD do oceny efektywności nauczania w gimnazjum

W 2005 r. przy Centralnej Komisji Egzaminacyjnej zapoczątkowano prace nad rozwojem metody edukacyjnej wartości dodanej w realiach polskiego systemu oświaty⁵.

W tym tekście skupimy się na najbardziej zawansowanych pracach nad EWD dla gimnazjów. Ogólną logikę szacowania EWD dla gimnazjów pokazuje poniższy schemat postępowania:

- na podstawie wyniku ucznia na sprawdzianie szacujemy wynik przewidywany na egzaminie gimnazjalnym,
- obliczamy różnicę – tzw. resztę - między faktycznym wynikiem ucznia na egzaminie a jego wynikiem przewidywanym,
- obliczamy średnią reszt, czyli EWD dla szkoły,
- szacujemy przedział ufności dla EWD.

Opiszę te kroki, a następnie przedstawię najważniejsze problemy związane z szacowaniem EWD.

4.1. Wynik przewidywany

W uproszczeniu wynik przewidywany można rozumieć jako uśredniony wynik na egzaminie gimnazjalnym grupy uczniów o tym samym wyniku na sprawdzianie *po szkole podstawowej*.

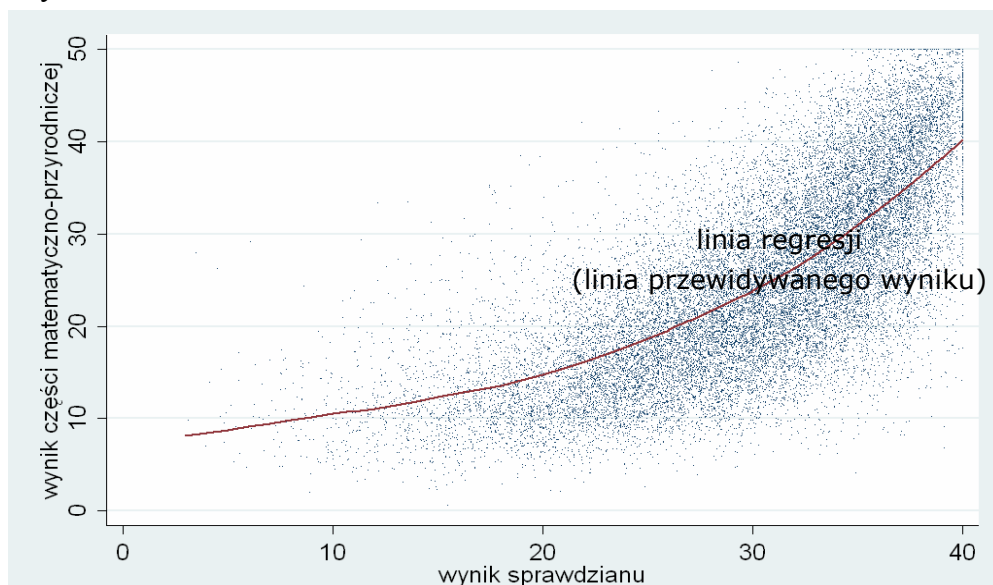
W rzeczywistości model jest bardziej złożony, bo wynik przewidywany określamy za pomocą zależności funkcyjnej, najlepiej odpowiadającej rzeczywistym danym *uzyskanym na poziomie kraju*, a otrzymaną wartość korygujemy o wpływ takich czynników jak płeć i dysleksja.

Wynik przewidywany jest szacowany na podstawie zależności między wynikami sprawdzianu a egzaminu gimnazjalnego w danym roku dla wszystkich trzecioklasistów w kraju.

⁵ Badania dotyczące edukacyjnej wartości dodanej są prowadzone w ramach projektu "Badania dotyczące wyników egzaminów zewnętrznych" realizowanego przez Centralną Komisję Egzaminacyjną współfinansowanego przez Europejski Fundusz Społeczny.

Wynik przewidywany na egzaminie gimnazjalnym uczniów na podstawie ich wyników ze sprawdzianu szacujemy metodą regresji. Jest to metoda stosowana w statystyce do oceny zależności między zmiennymi (w tym przypadku wynikami egzaminu gimnazjalnego oraz sprawdzianu), którą można ująć w postaci funkcji.

Sposób wyznaczania wartości przewidywanych prześledźmy na przykładzie danych ze sprawdzianu w 2002 r. i egzaminu gimnazjalnego w części matematyczno-przyrodniczej w roku 2005 (wykres 2). Analiza taka jest możliwa, gdy dla każdego ucznia mamy informacje zarówno o wyniku sprawdzianu jak i wyniku egzaminu. Przy aktualnym stanie egzaminacyjnego systemu informatycznego łączenie danych z różnych egzaminów dla uczniów jest trudne. Pokonanie tej przeszkody technicznej jest jednym z warunków rozwoju metody EWD.



Wykres 2. Zależność między wynikami na sprawdzianie w 2002r. a wynikami tych samych uczniów 3 lata później na egzaminie gimnazjalnym w części matematyczno-przyrodniczej

Na wykresie widzimy zależność między analizowanymi danymi. Małe punkty reprezentują uczniów. Położenie punktu na wykresie zależy od wyniku na sprawdzianie (oś pozioma) i od wyniku na egzaminie gimnazjalnym (oś pionowa). Zaznaczona linia regresji, najlepiej odzwierciedla kształt związku między wynikami. To, jaki kształt będzie miała linia, zależy od tego, z jakiego rodzaju zależnością między wynikami sprawdzianu a egzaminu gimnazjalnego będziemy mieli do czynienia w danym roku. W kolejnych latach kształt zależności może być nieco inny.

Można powiedzieć, że linia regresji – pozwala w sensie statystycznym - przewidywać wyniki egzaminu gimnazjalnego w zależności od wyniku ucznia na sprawdzianie. Pojedyncze punkty prezentujące wyniki poszczególnych uczniów w większości leżą poza nią, a więc różnią się od przewidywanych wartości. Jest to naturalne i wynika zarówno z różnej efektywności nauczania – co oczywiście nas tu najbardziej interesuje - jak i innych czynników odpowiedzialnych za wynik testowania, w tym czynnika losowego związanego z błędem pomiaru.

Postępując się linią regresji wyznaczamy wynik przewidywany egzaminu gimnazjalnego dla uczniów o danym wyniku na sprawdzianie. Na przykład, jeżeli interesuje nas wynik przewidywany dla uczniów, którzy na sprawdzianie uzyskali 20 pkt, z wykresu

odczytujemy, że wynikowi temu odpowiada 15 pkt na egzaminie gimnazjalnym w części matematyczno-przyrodniczej.

W praktyce wynik ten odczytywać będziemy ze specjalnych tabel, jednak wykres w dobry sposób pokazuje skąd bierzemy te wartości.

W szacowaniu wyniku przewidywanego dodatkowo brane są pod uwagę tzw. **zmiennie kontrolne** powiązane z wynikami egzaminów. W modelu przyjętym dla gimnazjum wykorzystywane są informacje o dysleksji i płci. Zmienne te są dostępne w bazie danych egzaminacyjnych.

Potrzebę użycia zmiennych kontrolnych rozważmy na następującym przykładzie. W części humanistycznej w skali kraju lepiej wypadają dziewczęta. Gdybyśmy nie wzięli tego pod uwagę w modelu szacowania EWD, to szkoła z silną przewagą liczby chłopców byłaby tak samo traktowana w obliczeniach, jak szkoła z przewagą dziewcząt. Wprowadzenie do modelu zmiennej płci pozwala nam porównywać szkoły, biorąc pod uwagę tę cechę. W ten sposób np. szkoły żeńskie będą porównywane z innymi szkołami żeńskimi. Mówiąc w skrócie, zmiennie kontrolne pozwalają bardziej adekwatnie wyznaczać EWD.

4.2. Reszty

Kolejnym ważnym pojęciem, a równocześnie etapem w obliczaniu EWD jest wspomniana już reszta z równania regresji. Reszta w języku statystyki oznacza różnicę między wartością faktyczną a wartością przewidywaną na podstawie równania regresji. **W metodzie EWD reszta to różnica między realnym wynikiem ucznia na egzaminie gimnazjalnym a wartością oczekiwaną obliczoną na podstawie wyniku na sprawdzianie i wartości zmiennych kontrolnych – płci i dysleksji.**

Czym jest reszta, łatwo zrozumiemy, gdy wrócimy do wykresu 2, który ilustrował obliczanie wyniku przewidywanego. Przypomnijmy, punkty reprezentują uczniów. Jeżeli jakiś punkt znajduje się ponad linią wyniku przewidywanego, reszta jest dodatnia. Oznacza to, że wynik ucznia na egzaminie gimnazjalnym jest wyższy niż oszacowany, przeciętny wynik uzyskany w kraju przez uczniów o tym samym wyniku na sprawdzianie oraz wartości zmiennych kontrolnych. Jeżeli reszta dla ucznia A wynosi 10 pkt. oznacza to, że na egzaminie gimnazjalnym uzyskał o tyle punktów więcej, niż średnio w kraju uczniowie o tych samych cechach: wynik na sprawdzianie, dysleksja, płeć. Jeżeli jakiś punkt znajduje się poniżej linii wyniku przewidywanego, reszta jest ujemna. Oznacza to, że wynik ucznia na egzaminie gimnazjalnym jest niższy niż przeciętny wynik uzyskany przez uczniów o tych samych cechach. Jeżeli reszta dla ucznia B wynosi -10 pkt. oznacza to, że na egzaminie gimnazjalnym uzyskał o tyle punktów mniej, niż średnio rówieśnicy w kraju o tym samym wyniku na sprawdzianie, dysleksji, płci.

Jak interpretować wartość reszty? Na jej wartość wpływają różne czynniki, np.:

- czynniki motywacyjne – wzrost motywacji w porównaniu z końcem szkoły podstawowej może dać relatywnie wyższy wynik,
- czynniki środowiskowe – poprawa sytuacji materialnej rodziny pozwala na większe inwestycje w edukację dziecka,
- czynniki biologiczne – zmienne tempo dojrzewania i rozwoju umysłowego,
- czynniki sytuacyjne czy losowe - lepsze samopoczucie w trakcie pisania testu, więcej szczęścia w wskazywaniu poprawnych odpowiedzi w zadaniach zamkniętych.

Dla nas jednak najważniejszy jest czynnik związany z efektywnością nauczania – za część wartości wskaźnika resztowego odpowiada szkoła i efektywność prowadzonego procesu dydaktycznego.

Ponieważ reszta jest wartością nieprecyzyjną, nie należy jej wykorzystywać jako indywidualnej miary edukacyjnych postępów danego ucznia w gimnazjum. **Reszty należy używać jedynie do analizowania EWD szkoły lub innej badanej grupy.**

4.3. Wskaźnik edukacyjnej wartości dodanej

EWD obliczamy uśredniając reszty dla danej szkoły (lub innej analizowanej grupy uczniów np. klasy, grupy uczniów dojeżdżających i niedojeżdżających itp.). Średnią tę interpretujemy jako miarę efektywności nauczania w danej szkole lub grupie uczniów.

Warto jeszcze raz podkreślić, że *EWD szkoły* szacujemy na podstawie reszt wszystkich uczniów danej szkoły. Niektórzy z nich mogą mieć reszty ujemne, jednak jeśli większość ma reszty dodatnie, to EWD szkoły będzie także dodatnia.

Jeśli wyniki większości uczniów będą leżeć nad linią regresji, to szkoła ma dodatnią EWD (większość uczniów uzyskuje wyniki wyższe niż oczekiwane dla przeciętnego ucznia w Polsce o podobnym wyniku ze sprawdzianu). Im dalej nad krzywą leży większość wyników uczniów szkoły, tym wyższa jej EWD. Jeśli większość leży pod krzywą, to EWD będzie ujemna (większość uczniów uzyskała wyniki poniżej oczekiwanego).

Trzeba pamiętać, że dla wszystkich uczniów w kraju średnia reszt wynosi zero. EWD jest więc miarą względną – część szkół będzie miała dodatnią EWD i podobna liczba - ujemną EWD. Miara ta pozwala na porównanie szkół między sobą w skali kraju.

EWD dla szkoły wyrażona jest w jednostkach, w których odbywa się pomiar na egzaminie gimnazjalnym, czyli w punktach egzaminacyjnych.

Jeżeli w szkole EWD wynosi +3,5 pkt oznacza to, że uczniowie w tej placówce uzyskali średnio o tyle punktów więcej na egzaminie gimnazjalnym w porównaniu ze szkołami w całym kraju o populacji uczniów takiej samej ze względu na wyniki na sprawdzianie, płeć i dysleksję. Natomiast jeżeli w szkole EWD wynosi np.: -2,3 pkt oznacza to, że średnio uczniowie uzyskali o tyle punktów mniej na egzaminie gimnazjalnym w porównaniu ze szkołami o analogicznym składzie uczniowskim.

EWD można wyrażać w skali staninowej i centylowej.

Przypomnijmy, **skala staninowa** jest skalą 9. stopniową:

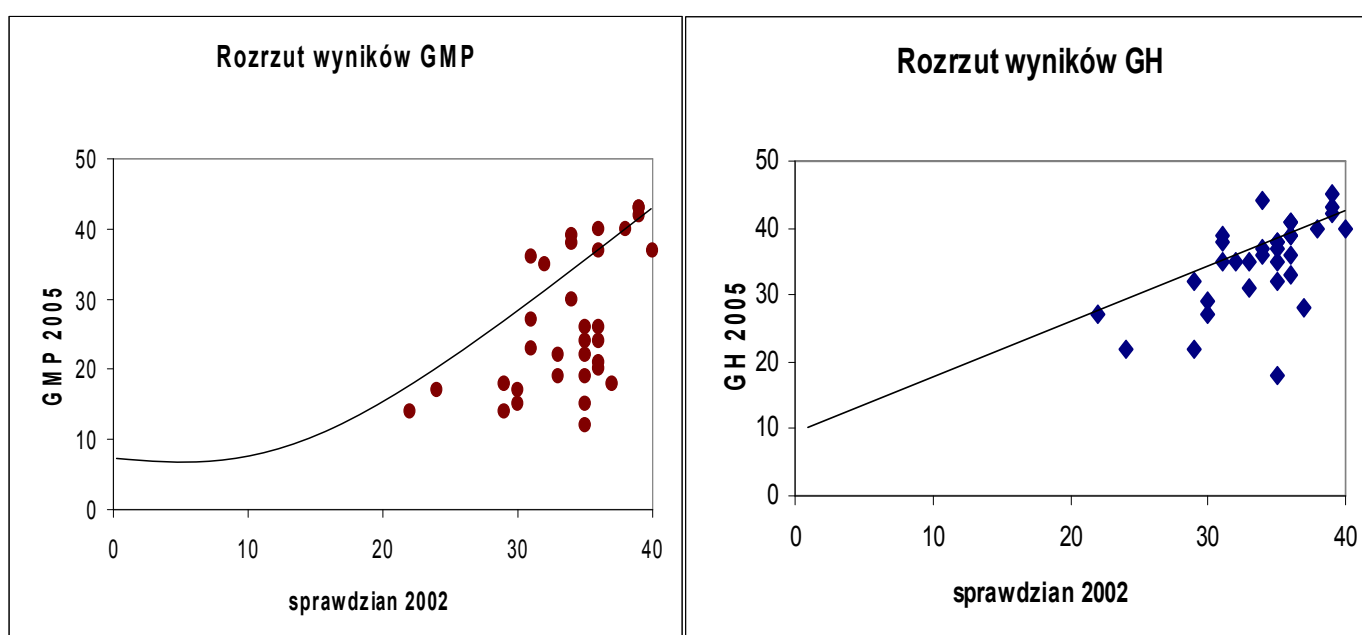
- pierwszy stanin to przedział obejmujący 4% najniższych wyników,
- drugi stanin to przedział obejmujący 7% kolejnych wyników,
- trzeci stanin to przedział obejmujący 12% kolejnych wyników,
- czwarty stanin to przedział obejmujący 17% kolejnych wyników,
- piąty stanin to przedział obejmujący 20% kolejnych wyników,
- szósty stanin to przedział obejmujący 17% kolejnych wyników,

- siódmy stanin to przedział obejmujący 12% kolejnych wyników,
- ósmy stanin to przedział obejmujący 7% kolejnych wyników,
- dziewiąty stanin to przedział obejmujący 4% najwyższych wyników.

Wynik w staninie 8. (tzw. bardzo wysokim) oznacza, że 4% szkół uzyskało wynik wyższy, 7% - porównywalny, a 89% - wynik niższy.

Inną miarą pozycyjną jest 99. stopniowa **skala centylowa**. Pozycja centylowa mówi o tym, jaki procent szkół ma wynik niższy, niż uzyskany przez tę szkołę, np.: 45 centyl oznacza, że 45% szkół uzyskało wynik niższy.

Na przykładzie wybranej szkoły zobaczymy, jak może wyglądać relacja między wynikami egzaminacyjnymi a edukacyjną wartością dodaną.



GMP⁶=26,45, stanin - wysoki
EWD =-2,95, stanin - niski, 12 centyl

GH=34,97, stanin - wyżej średni
EWD =-2,59, stanin - niski, 14 centyl

Wykres 3. Wynik egzaminu a edukacyjna wartość dodana dla przykładowego gimnazjum

Powyższe wykresy przedstawiają wyniki w przykładowym gimnazjum. Na osi poziomej mamy wyniki na sprawdzianie, na osi pionowej wyniki na egzaminie gimnazjalnym. Na wykresie orientacyjnie wrysowano linie wyniku przewidywanego. Punkty na wykresie reprezentują uczniów tej szkoły. Analiza wykresów pozwala zrozumieć, dlaczego wyniki surowe (podane w punktach) egzaminu gimnazjalnego lokują tę szkołę wysoko w skali staninowej, a w EWD – nisko. Choć wyniki egzaminacyjne uczniów są dość wysokie, to jednak większość punktów znajduje się poniżej linii wyniku przewidywanego.

Dzieje się tak, ponieważ uczniowie legitymowali się w większości przypadków wysokimi wynikami na sprawdzianie *po szkole podstawowej*. Innymi słowy wysoki potencjał uczniów nie został przez szkołę wykorzystany.

⁶ GMP = egzamin gimnazjalny, część matematyczno – przyrodnicza; GH = egzamin gimnazjalny, część humanistyczna

Należy podkreślić, że w całej populacji gimnazjów dominować będą przypadki niewielkich lub umiarkowanych różnic między wynikiem surowym egzaminu a EWD. Jednak dla pewnej grupy gimnazjów notujemy duże lub bardzo duże rozbieżności między wynikiem surowym a EWD.

4.4. Precyzja szacowania edukacyjnej wartości dodanej

Wskaźnik EWD jak każdy wynik pomiaru nie jest doskonale precyzyjny. Jak oszacować błąd pomiaru? Służy temu statystyczne narzędzie zwane przedziałem ufności.

Przedział ufności jest to przedział, który z określonym prawdopodobieństwem zawiera prawdziwą wartość interesującego nas parametru. Mówi nam, że dany parametr (w naszym przypadku EWD) z określonym prawdopodobieństwem mieści się między wyznaczonymi wartościami.

W wypadku EWD szacowanie przedziału ufności jest konieczne, ponieważ określanie tej wartości obarczone jest dwojakim błędem: błędem pomiaru na sprawdzianie i egzaminie gimnazjalnym oraz błędem próbkowania – absolwenci danego rocznika są tylko próbą uczniów szkoły, na podstawie której wypowiadamy się o efektywności nauczania w danym gimnazjum, a ta jest cechą charakteryzującą szkołę w dłuższej perspektywie czasowej.

Zauważmy przy okazji, że ocena jakości pracy szkoły – w różnych jej aspektach, bo przecież EWD to tylko jeden ze wskaźników tej jakości – zawsze wymaga wieloletniej perspektywy. Formułowanie ocen na podstawie pojedynczego pomiaru, analizy jednego rocznika jest nieodpowiedzialne.

Wielkość przedziału ufności dla EWD danej szkoły zależy od trzech czynników:

- zróżnicowania wartości reszt, czyli indywidualnych wskaźników postępu uczniów - im bardziej są zróżnicowane, tym szerszy jest przedział ufności. To intuicyjne zrozumiałe. Reszty to informacje, na podstawie których wnioskujemy o efektywności nauczania. Jeżeli ich wartości są bardzo różne, trudno o pewny wniosek co do „wkładu” pracy szkoły w wynik egzaminacyjny ucznia.
- liczebności populacji uczniów - im mniej uczniów przystępuje do egzaminu, tym szerszy przedział ufności. Ta zasada jest też łatwa do zrozumienia – im mniej uczniów, tym mniej informacji o efektywności, a tym samym mniej pewny wniosek.
- od założonego poziomu ufności - im bardziej zależy nam na pewności, że faktyczna wartość EWD znajdzie się w wyznaczonym przedziale ufności, tym szerszy musimy wyznaczyć przedział. Poziom tej pewności w statystyce nazywamy poziomem ufności. Zwykle przyjmujemy 95% poziom ufności, czyli zgadzamy się na 5% ryzyko popełnienia błędu polegającego na tym, że prawdziwa wartość EWD znajdzie się poza wyznaczonymi wartościami granicznymi. Przedziały ufności określa się zgodnie z obowiązującymi w statystyce procedurami.

Wyznaczone dla EWD przedziały ufności możemy traktować jako regułę decyzyjną. Jeżeli chcemy w sposób odpowiedzialny formułować na podstawie analizy EWD oceny typu: *szkoła A lepiej uczy w zakresie sprawdzanym przez egzamin gimnazjalny od szkoły B*, to warto wiedzieć, jakie jest ryzyko popełnienia błędu. Przedziały ufności pozwalają nam to ryzyko oszacować.

Jeżeli wyznaczmy 95% przedziały ufności EWD dla porównywanych szkół i przedziały te są rozłączne, to ryzyko sformułowania nietrafnej oceny jest niewielkie – poniżej 5%. Gdy przedziały częściowo pokrywają się, formułowanie oceny staje się bardziej ryzykowane. Jeżeli ocena jest doniosła i brzemienna w skutki – należy się od niej powstrzymać.

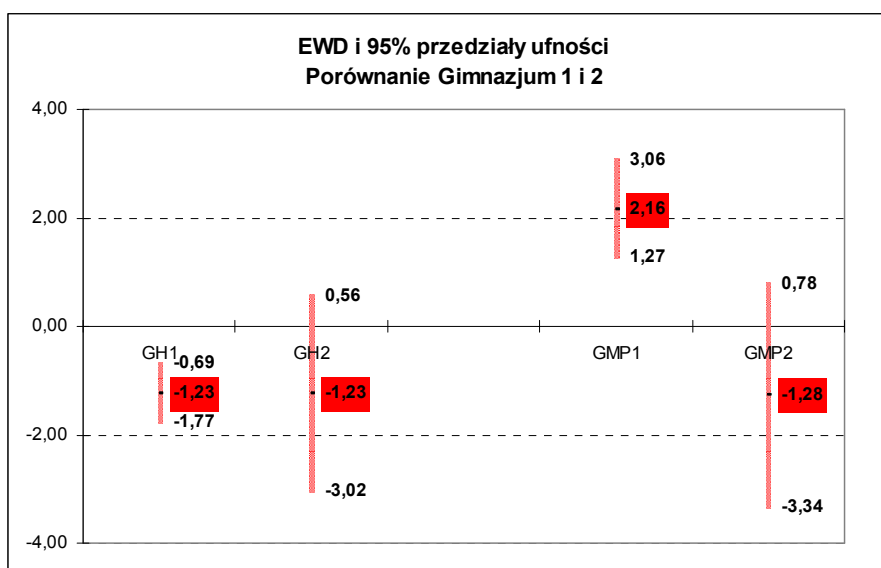
Przykładowo: jeśli przedział ufności EWD dla szkoły A wynosi (1,05; 3,47) a dla szkoły B (-1,17; 0,53), to możemy odpowiedzialnie stwierdzić, że szkoła A ma wyższą EWD niż szkoła B. Natomiast jeżeli dla szkoły C przedział wynosi (-1,93; 1,82), to nie możemy bez dużego ryzyka popełnienia błędu sformułować oceny, że EWD szkoły A jest wyższe niż szkoły C – bo oba przedziały zawierają wspólny zakres wartości (od 1,05 do 1,82).

W przypadku szacowania EWD dla szkoły przyjmujemy 95% poziom ufności (podobnie jak w większości badań naukowych), a dla klas lub innych grup uczniów porównywanych w obrębie jednej szkoły można przyjąć niższy poziom ufności np. 90%.

Niższy poziom ufności w tym wypadku wynika z mniejszych kosztów sformułowania nietrafnej oceny. A należy pamiętać, że w pewnych sytuacjach większe ryzyko wiąże się z niedocenianiem różnicy niż jej przecenieniem. Gdy oceniamy szkoły i za oceną idą ważne decyzje, przede wszystkim należy wystrzegać się ocen pochopnych. Ale gdy analiza wykaże na przykład, że w danej szkole relatywnie gorzej nauczeni są uczniowie zdolni, być może warto dmuchać na zimne i podejmować działania, nawet gdy trafności diagnozy nie możemy być pewni.

Włączanie do analiz danych z kolejnych edycji egzaminów zewnętrznych poprawi precyzję szacowania EWD dla szkoły i pozwoli na formułowanie odpowiedzialnych ocen nawet przy mniejszych różnicach między szkołami.

Bardzo ilustratywne jest prezentowanie przedziałów ufności w sposób graficzny. Poniższy wykres przedstawia oszacowania EWD wraz z przedziałami ufności dla dwóch gimnazjów i dwóch części egzaminu



Wykres 4. EWD wraz z przedziałami ufności dla dwóch gimnazjów i dwóch części egzaminu

Na wykresie punkty symbolizują oszacowania EWD (jej wartość jest zapisana na czerwonym tle), a linie pionowe szerokość przedziałów ufności.

Poprawna interpretacja jest następująca: w części humanistycznej (GH1 i GH2) obydwa gimnazja charakteryzuje podobna efektywność nauczania (to samo EWD więc i przedziały ufności się pokrywają). W części matematyczno-przyrodniczej Gimnazjum 1 ma wyższą EWD niż Gimnazjum 2. Co więcej - przedziały ufności tych dwóch szkół (zilustrowane przez czerwone, pionowe linie) nie „zazębiają się”. Zatem w pełni uprawniony statystycznie sposób można wnioskować, że efektywność nauczania w obszarze sprawdzanym przez egzamin matematyczno-przyrodniczy w tych dwóch szkołach jest różna.

Jeśli wartości EWD byłyby różne, ale przedziały ufności częściowo by się pokrywały, to takie wnioskowanie byłoby obarczone dużym ryzykiem popełnienia błędu w ocenie.

Warto także zwrócić uwagę na znacznie szersze przedziały ufności Gimnazjum 2. Może to wynikać z mniejszej liczby uczniów w tej szkole. Jednak jeśli liczba uczniów jest podobna, to wynika z większego zróżnicowania wartości reszt w Gimnazjum 2, co jest dodatkową, interesującą informacją o tej szkole.

5. Problem stabilności oszacowań EWD w czasie

Tak jak już zaznaczyliśmy, ocena jakości pracy szkoły to przedsięwzięcie długofalowe, wymagające kolekcjonowana w systematyczny sposób informacji przez wiele lat. Dotyczy to również EWD i orzekania o efektywności nauczania w zakresie sprawdzanym przez egzaminy zewnętrzne.

Wynik uzyskany przez szkołę w danym roku to tylko próbka efektywności. Z pewnością z roku na rok będziemy obserwować wahania w wartości EWD dla danej szkoły. Ważne jest jednak, jak duże będą te wahania. Jeżeli szacunki EWD dla kolejnych lat nie byłyby z sobą skorelowane, wskazywałyoby to na bezużyteczność metody.

Jak przedstawiają się fakty? Zespół EWD pracujący przy CKE przeanalizował dane dla dwóch edycji egzaminu: panel *sprawdzian 2002 – egzamin gimnazjalny 2005* oraz panel *sprawdzian 2003 – egzamin gimnazjalny 2006*. Korelacja między oszacowaniami EWD dla gimnazjów w 2005 i 2006 roku jest umiarkowana i wynosi około 0,5. Wskazuje to na konieczność – o której już wspominaliśmy powyżej – opierania szacunków EWD na wynikach wieloletnich.

Jest prawdopodobnie kilka przyczyn niezbyt dużej korelacji między szacunkami EWD w kolejnych latach, np.: niedoskonałość danych, na których opieramy wyliczenia. Zarówno wyniki sprawdzianu jak i egzaminu gimnazjalnego – jak każdego zresztą pomiaru – obarczone są błędem. Warto podjąć badania nad rzetelnością testów i podjąć działania zmierzające do ograniczenia takich źródeł błędu pomiaru jak efekt egzaminatora czy efekt ściągania. Na wartość metody EWD z pewnością rzutują również problemy związane z trafnością testów. Na zmienność oszacowań EWD w kolejnych latach z pewnością ma wpływ także czynnik nauczycielski. Wiadomo, że efektywność nauczania zależy przede wszystkim od nauczycieli. W większych gimnazjach co roku skład kadry nauczycielskiej jest inny. To też z pewnością sprawia, że EWD szkoły z roku na rok może się zmieniać. Ale gdyby korelacje między kolejnymi oszacowaniami EWD były doskonałe, czyli przybierały wartość

1, byłyby to wynik przygnębiający. Oznaczałoby to, że efektywność jest niezmienna i nie ma „przestrzeni zmiany” – szkoły nie miałyby szansy na poprawę.

6. Zastosowanie EWD do analiz wewnętrzzkolnych

Do tej pory skupialiśmy się na problemach ewaluacji zewnętrznej efektywności nauczania i szacowaniu edukacyjnej wartości dla całej szkoły. Jednak metoda EWD może i powinna być stosowana też do analiz wewnętrzzkolnych. Analizy takie mogą być wykonywane przez szkoły na własne potrzeby, lub przez nadzór pedagogiczny, by wspierać szkoły w poprawie jakości pracy.

W Stanach Zjednoczonych, Anglii, Francji czy Australii wskaźnik EWD wykorzystywany jest od wielu lat zarówno w ewaluacji zewnętrznej jak i w pracy pedagogicznej w szkole. EWD jest wartościowym wskaźnikiem dla **wewnętrzzkolnej oceny i planowania działań** podejmowanych przez szkołę w celu podniesienia jakości kształcenia.

Takie analizy ze względu na swoje zastosowanie nie wymagają tak dużej precyzji jak szacowanie EWD dla szkoły – przyjmuje się, że w tym przypadku wystarczy rozpatrywać 90% przedziały ufności.

Oto typy analiz, jakie można wykonywać w gimnazjum za pomocą metody EWD⁷. Ich lista nie jest z pewnością wyczerpująca i przedstawia jedynie przykłady możliwych zastosowań EWD. Wszystko zależy od potrzeb i pomysłowości użytkowników tej metody.

1) Analiza rozkładu reszt

Choć w dotychczasowych rozważaniach EWD definiowaliśmy jako średnią arytmetyczną reszt, to pogłębiona analiza efektywności wymaga przyjrzenia się pełnemu rozkładowi wyników w szkole. Za daną średnią mogą się kryć bardzo różne rozkłady reszt, czyli tak naprawdę różne edukacyjne wartości dodane. W analizie rozkładu bardzo ważne jest też pojęcie **wartości odstających**, czyli znacząco odbiegających od pozostałych wartości w próbie. W wypadku EWD wartości odstające to reszty drastycznie różniące się od pozostałych reszt w szkole. Na przykład, w 56 osobowym gimnazjum wartości reszt 54 uczniów mieszczą się w przedziale (-10 ; +10). Wyniki dwu pozostałych to +19,5 oraz -21,1. Te wartości uznamy za odstające, bo znacznie różnią się od pozostałych. Można przypuszczać, że wartości odstające są wynikiem działania czynników losowych, więc wskazane jest minimalizowanie ich wpływu na szacowanie EWD szkoły.

2) Analiza EWD dla klas lub nauczycieli

Dla szkoły niezwykle ważne jest szacowanie EWD dla poszczególnych klas lub nauczycieli. Pozwala to lepiej poznać czynniki odpowiedzialne za wynik szkoły, a tym samym trafniej zaplanować program naprawczy. Analiza wyników dla klas pozwoli też ocenić efekty przyjętego w szkole sposobu dzielenia uczniów na klasy.

3) Analiza efektu płci

Komu lepiej wiedzie się w szkole, dziewczętom, czy chłopcom? Oczywiście w skali kraju EWD dla dziewcząt i chłopców jest takie samo, ale w danej szkole mogą wystąpić znaczące różnice. Choć w polskiej debacie edukacyjnej problemy równości szans edukacyjnych chłopców i dziewcząt są słabo jeszcze obecne, z pewnością w najbliższych

⁷ Na stronie CKE znajduje się Kalkulator EWD 2006, który pozwala obliczyć reszty potrzebne do poniższych analiz za rok 2006

latach problem będzie traktowany coraz poważniej, a tym samym śledzenie różnic w tym zakresie w efektywności nauczania nabierze społecznego znaczenia.

4) EWD w grupach uczniów o różnym potencjale edukacyjnym

Niezwykle interesującą dla szkoły analizą może być przyjrzenie się edukacyjnej wartości dodanej w grupach uczniów lepszych i słabszych. Z kim lepiej szkoła pracuje, dla kogo ma bardziej efektywną ofertę? To informacja, która pozwoli szkole wiele dowiedzieć się o sobie. To również bardzo ważny problem z perspektywy polityki równych szans edukacyjnych.

5) EWD w grupach uczniów dojeżdżających do szkoły

Wprowadzenie gimnazjów ożywiło nie nowy spór o to, co lepsze dla uczniów: czy szkoła mała, gorzej wyposażona, oddalona od lokalnych centrów kulturowych, ale blisko miejsca zamieszkania dziecka, czy szkoła duża, lepiej wyposażona, funkcjonująca w korzystniejszym środowisku kulturowym, ale wymagająca czasochłonnego dowożenia? Szkoła lub organ prowadzący może na bieżąco śledzić skutki różnych rozwiązań organizacyjnych szacując EWD dla odpowiednich grup uczniów.

6) EWD w grupach uczniów objętych zajęciami wyrównawczymi

Analiza EWD może przynieść cenne informacje na temat skuteczności zajęć wyrównawczych. Czy intensyfikacja tych zajęć, nowe rozwiązania organizacyjne lub metodyczne przynoszą pożądane efekty? Na takie pytania może odpowiedzieć systematyczne monitorowanie EWD w różnych grupach uczniów.

7) EWD w grupach uczniów pochodzących z różnych szkół podstawowych

To ostatni przykład problemu, który można zgłębić za pomocą metody EWD. Dla niektórych gimnazjów może to być ważne i interesujące.

Podsumowanie

Najbardziej fundamentalny spór o metodę edukacyjnej wartości dodanej wiąże się z pytaniem o kształt polskiego systemu oświaty. Czy wzmacniać tendencje decentralistyczne, pluralizm programowy i zakorzenianie się szkoły w lokalnej społeczności, czy uruchamiać mechanizmy rynkowe w oświacie, zrezygnować z rejonizacji, postawić na rywalizację i regulację oświaty przez wybory rodziców, czy może centralizować i wzmacniać tendencje uniformistyczne przez, na przykład, wzmacnianie systemu egzaminów państwowych?

EWD wpisuje się raczej w tendencje centralistyczne i choć znalazłaby swoje miejsce również w liberalnej wizji oświaty, z pewnością kłóci się ze strategią pierwszą. Z jednej strony wizja szkoły otwartej na różnorodne, często subtelne „impulsy sterujące” płynące od różnych partnerów społecznych w jej lokalnym środowisku, a z drugiej - szkoła poddana jednolitemu, silnemu nadzorowi pedagogicznemu, którego orężem staje się nowy wskaźnik jakości jej pracy - edukacyjna wartość dodana.

Choć strategia centralistyczna na pierwszy rzut oka wygląda odpychająco, pamiętajmy, że tak kluczowy cel polityki oświatowej jak podtrzymywanie spójności społecznej, a w tym polityka równych szans edukacyjnych, nie da się skutecznie realizować zarówno w obrębie strategii wspólnotowej jak i rynkowej. Zarówno szkoła - element wspólnoty lokalnej, jak i szkoła - dostawca usług nie gwarantuje realizacji tych celów. To dylematy warte dyskusji.

Metoda edukacyjnej wartości dodanej jest instrumentem polityki oświatowej. Można oceniać go w oderwaniu od innych narzędzi, ale bardziej racjonalne jest myślenie w kategoriach porównawczych – czy instrument X jest lepszy od instrumentu Y?

W realiach polskiego systemu oświaty, od 2002 roku zaczął funkcjonować system egzaminów zewnętrznych – potężne narzędzie polityki oświatowej. Czy wykorzystujemy jego możliwości? Czy śledzimy realne skutki? To szerszy kontekst namysłu nad wartością metody

EWD. Z niego wynika pytanie, czy wskaźniki EWD są lepszą miarą efektywności nauczania w danej szkole niż średnia arytmetyczna wyniku egzaminu? Dotychczasowe wyniki analiz pozwalają odpowiedzieć twierdząco.

Dotychczasowe prace nad EWD w Polsce można podsumować następująco:

1. Metoda edukacyjnej wartości dodanej może być z powodzeniem stosowana do oceny efektywności nauczania w gimnazjach i szkołach maturalnych w zakresie sprawdzanym na egzaminach zewnętrznych. Tworzenie modeli szacowania EWD dla gimnazjów weszło już w fazę pilotażową. W wypadku szkół maturalnych potrzeba jeszcze dalszych prac, by doprowadzić metodę do fazy pilotażowej.

2. Wskaźniki EWD wzbogacają repertuar miar jakości pracy szkoły i powinny być traktowane jako komplementarne wobec nich.

3. Dysponujemy bardzo bogatym repertuarem metod statystycznych szacowania EWD – od najprostszych modeli regresyjnych – przedstawionych w tym tekście - po wyrafinowane modele hierarchiczne. W praktyce jednak różnice między oszacowaniami EWD za pomocą różnych metod nie są duże. Znacznie ważniejsza jest dostępność i jakość danych - głównie egzaminacyjnych - użytych w modelach szacowania EWD.

4. Wskaźniki EWD obliczane na podstawie danych jednorocznych są umiarkowanie stabilne w czasie. Znacząca zmienność EWD szkoły w kolejnych latach z jednej strony pokazuje konieczność szacowania przedziałowego tego wskaźnika (przedziały ufności), z drugiej strony wyznacza kierunek prac nad modelami szacowania. Dla potrzeb ewaluacji wewnątrzszkolnej powinny być stosowane modele jednoroczne, ale dla ewaluacji zewnętrznej należy stosować modele wieloletnie.

5. Badania nad trafnością metody EWD w odniesieniu do gimnazjów wykazały, że miary efektywności nauczania wykorzystujące metodę EWD są zdecydowanie słabiej powiązane z uwarunkowaniami środowiskowymi pracy szkoły w stosunku do nieprzetworzonych wyników egzaminów. Kontrola wpływu czynników środowiskowych pozwala traktować EWD jako zdecydowanie lepszą miarę efektywności nauczania. Jednak brak jest mocnych argumentów na rzecz prawomocności uogólniania znaczenia wskaźnika EWD na całość procesu dydaktycznego. Badania wskazały, że istnieją znaczące korelacje EWD z niezależnymi miarami jakości nauczania, ale pokazały również, że wiele empirycznych wskaźników jakości opartych na opiniach i ocenach uczniów i rodziców nie potwierdza jednoznacznie przewagi miar EWD nad surowymi wynikami egzaminów. Dopóki dalsze badania nie dostarczą mocnych przesłanek dla takich uogólnień, należy wskaźniki EWD interpretować jako miarę efektywności nauczania w zakresie sprawdzanym na egzaminach zewnętrznych. By podkreślić tę zawężającą interpretację warto rozważyć zastąpienie powszechnie na świecie stosowanej nazwy *edukacyjna wartość dodana* terminem *egzaminacyjna wartość dodana*.

6. W ewaluacji zewnętrznej metoda EWD powinna być używana przede wszystkim do identyfikacji szkół o szczególnie niskiej efektywności nauczania oraz szkół o najwyższym poziomie tej cechy. Identyfikacja przez organy nadzoru pedagogicznego tych pierwszych powinna prowadzić do uruchamiania działań naprawczych, a rozpoznanie drugich – do wskazywania szkół, których praca może być źródłem inspiracji dla innych.

7. Metoda EWD może być z powodzeniem stosowana do ewaluacji wewnątrzszkolnej. EWD pozwala analizować takie problemy jak: efektywność nauczania w poszczególnych klasach, skutki stosowanych metod dzielenia uczniów na oddziały, problemy równości szans edukacyjnych różnych grup uczniów, ocena skuteczności różnych działań zmierzających do podniesienia jakości nauczania.

8. Krytyczna analiza metody EWD pozwoliła zidentyfikować kluczowe problemy metody.

a. Jakość metody EWD zależy od dwóch czynników, które są wobec niej zewnętrzne. Są to program kształcenia i wiedza o czynnikach decydujących o opanowywaniu przez uczniów celów kształcenia, co łącznie można nazwać teorią kształcenia. Tylko wtedy, gdy model szacowania EWD jest zbudowany na dobrej teorii kształcenia, wskaźniki mogą być w pełni wartościowe. Podstawowe zadanie do wykonania to rozwój programów nauczania i doskonalenie testów egzaminacyjnych. Obecnie wartość metody EWD jest ograniczona przez słabość standardów wymagań egzaminacyjnych i - powiązaną z tym - problematyczną trafność narzędzi pomiarowych.

b. Ponieważ jest wysoce prawdopodobne, że takie zasoby szkoły, jak potencjał edukacyjny uczniów, wyposażenie dydaktyczne czy korzystny skład społeczny szkoły są powiązane z efektywnością nauczania, do modeli szacowania EWD należy bardzo ostrożnie włączać dodatkowe zmienne. Choć mamy przykłady stosowania tzw. modeli kontekstowych szacowania EWD (Anglia), to rozwiązania takie należy wdrażać z dużą ostrożnością.

c. Precyzji wskaźników EWD zagraża niedostateczna rzetelność testów. Z punktu widzenia metody EWD kluczowe są dwa źródła błędu pomiaru: efekt egzaminatora i efekt ściągania.

Każda nowa informacja o jakości nauczania ma szansę wpływać na procesy decyzyjne, gdy jest właściwie rozumiana. Oznacza to, że wdrożenie metody EWD to wielkie przedsięwzięcie szkoleniowe. Bardzo dużo już w tym zakresie zrobiono, jeszcze więcej jest do wykonania.

Bibliografia

- Ball, S., J. (1993) Education Markets, Choice and Social Class: The Market as a Class Strategy in the UK and USA. *British Journal of Sociology of Education*, Vol. 14, Issue 1, pp. 3-20.
- Dolata R. (2002) Procedury rekrutacji i dzielenia uczniów na oddziały w gimnazjach - próba oceny z perspektywy nierówności społecznych w edukacji. W: *Zmiany w systemie oświaty. Wyniki badań empirycznych*. Instytut Spraw Publicznych, Warszawa.
- Dolata R. (2006) Edukacyjna wartość dodana w komunikowaniu wyników egzaminów zewnętrznych. *Egzamin. Biuletyn Badawczy CKE*, 8, s. 9-20.
- Dolata R. (2006) Efektywność nauczania w gimnazjach miasta X. Analiza edukacyjnej wartości dodanej. *Egzamin. Biuletyn Badawczy CKE*, 8, s. 28-37.
- Dolata R., Murawska B., Putkiewicz E., Żytko M. (1997) *Monitorowanie osiągnięć szkolnych jako metoda doskonalenia edukacji. Zarys metody oraz przykłady zastosowań w edukacji początkowej*. Warszawa, Wydawnictwo Żak.
- Hanushek E., A. (1986) The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature*, vol. 24, pp. 1141-1177.
- Jakubowski M. (2006) Empiryczna analiza metod szacowania edukacyjnej wartości dodanej. *Egzamin. Biuletyn Badawczy CKE*, 8, s. 83-112.
- Jakubowski M. (2006) Metody szacowania edukacyjnej wartości dodanej. *Egzamin. Biuletyn Badawczy CKE*, 8, s. 67-82.
- Pokropek A. (2006) Analiza trafności wskaźnika edukacyjnej wartości dodanej. *Egzamin. Biuletyn Badawczy CKE*, 8, s. 113-136.
- Rothstein R. (2004) *Class and schools. Using social, economic, and educational reform to close the Black-White achievement gap*. Washington, Teachers College, Columbia University.
- Schagen I., Hutchinson D. (2003) Adding value in educational research – the marriage of data and analytical power. *British Educational Research Journal*, vol. 29, no. 5.
- Vasta R., Haith M., M., Miller S., A. (1995) *Psychologia dziecka*. Warszawa. WSiP.
- Gorard S., Fitz J. (2000) Markets and stratification: a view from England and Wales. *Educational Policy*, vol. 14, no. 3, pp. 405-428.