

Marek Zapieraczyński, Sławomir Sapanowski, Anna Zbidi
Okręgowa Komisja Egzaminacyjna w Łodzi

**Zależność wyników uzyskiwanych przez uczniów
z zadań otwartych od liczby punktów
uzyskanych za zadania zamknięte**

*(Element badań dotyczących porównywalności oceniania prac
ze sprawdzianu w OKE Łódź).*

„Jedynie prawda jest ciekawa”

Józef Mackiewicz

Wyniki egzaminów zewnętrznych decydują w znacznym stopniu o losach uczniów na następnych etapach edukacyjnych. Jako obiektywna miara selekcjonują badaną populację, promując najlepszych poprzez umożliwienie im kontynuowania nauki w wybranych (najlepszych) szkołach.

Na podstawie wyników egzaminów zewnętrznych wyrabiamy i utrwalamy sobie jako ogół społeczeństwa opinie na temat pracy poszczególnych placówek szkolnych oraz poziomu edukacji na obszarze kraju, jego regionów, poszczególnych województw, powiatów czy gmin.

Specjaliści zajmujący się pomiarem dydaktycznym na podstawie wyników egzaminów prognozują wyniki uczniów, jakie powinni oni uzyskiwać na następnych etapach edukacyjnych i wykorzystują to do mierzenia jakości pracy szkół, np. poprzez edukacyjną wartość dodaną (ewd). Nie ma wątpliwości, że informacje, jakie przynoszą egzaminy zewnętrzne i odpowiednie ich wykorzystanie, pozwalają o wiele lepiej niż dotychczas diagnozować sytuację w polskiej oświacie. Bezsprzecznie sześć sesji egzaminacyjnych w szkole podstawowej i gimnazjum dostarczyło dużo materiału porównawczego oraz skłoniło środowisko pedagogiczne i naukowe do różnych refleksji o efektywności działań edukacyjnych.

Obserwacja działań placówek oświatowych pozwala stwierdzić, że zdecydowana większość z nich przykłada dużą wagę do wyników

uzyskiwanych przez uczniów, a także stara się w miarę swoich możliwości, które stopniowo rosną (szkolenia nauczycieli z zakresu pomiaru dydaktycznego), właściwie te wyniki wykorzystywać. Niestety, istnieje grupa placówek uzyskująca wysokie wyniki poprzez stosowanie metod nieformalnych. W trakcie pisania przez uczniów egzaminów nie są przestrzegane procedury ich przeprowadzania, a wyniki to w znacznej mierze efekt podpowiadania oraz ściągania odpowiedzi zarówno do zadań zamkniętych, jak i otwartych. Świadczą o tym protokoły egzaminatorów i przypadki unieważniania egzaminów. Ilość unieważnień nie oddaje w pełni skali problemu, można tu powiedzieć, że to tylko czubek góry lodowej. Wielu wątpliwych sytuacji egzaminacyjnych nie da się udowodnić, także czas na ich zbadanie jest bardzo krótki. Ponadto sankcje, czyli powtórzenie egzaminu, są dla piszących niezbyt dotkliwą karą, a zespoły nadzorujące ponoszą ją w jeszcze mniejszym stopniu. Wszystko to nie jest wystarczającym hamulcem przed pokusą zadbania o interesy uczniów oraz szkoły. W wielu środowiskach spotyka się to nawet ze zrozumieniem. Jak inaczej wytłumaczyć istotny wzrost średniego wyniku uczniów szkoły po wizycie obserwatora reprezentującego lokalny samorząd? W opozycji do przywołanej, zresztą autentycznej, sytuacji, są o wiele częstsze przypadki gwałtownego spadku wyniku szkoły, ale po wizycie obiektywnego obserwatora z Okręgowej Komisji Egzaminacyjnej lub Kuratorium Oświaty. Bardzo często wynik placówki (najczęściej dotyczy to szkół wiejskich), który przez kilka lat plasował się w stanie bardzo wysokim lub najwyższym, spada do średniego, a nawet niżej średniego. Zjawisko to łatwo zaobserwować, przeglądając wyniki szkół na stronach *scholaris*.

W tym miejscu zasadne jest postawienie pytania. Czy wynik egzaminu zewnętrznego, np. Sprawdzianu 2007, dla populacji krajowej jest rzetelny i adekwatny do umiejętności przez nią prezentowanych? Biorąc pod uwagę wcześniej wymienione czynniki, wydaje się, że wynik ten jest zawyżony. Jaka jest skala tego zawyżenia? Aby to określić, należy przeprowadzić celowe badanie, które zdecydowanie obiektywniej odda obraz zjawiska niż nasze intuicyjne szacunki. Przeprowadzenie takiego badania wiąże się jednak z pewnymi problemami logistyczno-finasowymi. Jak miałoby ono wyglądać?

Można je przeprowadzić podczas następnej sesji sprawdzianu w kwietniu 2008 roku. Aby jego wyniki były wiarygodne, należy nim objąć 10% populacji dobranej celowo z uwzględnieniem lokalizacji placówki (wieś, małe miasto, średnie miasto i duże miasto) oraz wyników osiągniętych w poprzednich latach (stanin). Do każdej wskazanej placówki powinien pojechać obserwator OKE lub KO. Badanie należy przeprowadzić w każdym województwie. Taki zabieg niemal w 100% zapewniłby

uzyskanie obiektywnych wyników za zadania zamknięte. Do tych wyników można by przyrównać wyniki z zadań zamkniętych pozostałej części populacji i określić stopień ich wiarygodności. Wyżej opisane działania powinny rozwiązać problem wyników z zadań zamkniętych testu egzaminacyjnego. Pozostaje jednak równie istotna, także wagowo, kwestia punktacji za zadania otwarte. W OKE Łódź przeprowadzono badanie dotyczące tego problemu, niestety ograniczenia finansowe zawęziły jego zakres. Jednak wnioski płynące po jego przeprowadzeniu wydają się istotne i warte potwierdzenia na większej populacji.

Po pierwsze, po przeprowadzonym badaniu odnotowano znaczące rozbieżności w ocenianiu zadań otwartych w poszczególnych zespołach egzaminatorów. Wywołało to dużą konsternację, ponieważ zakłóciło poważnie wrażenie obrazu pracy zespołów egzaminatorów, jaki można było wynieść, bazując na dwóch testach diagnostycznych przeprowadzonych w trakcie sesji egzaminacyjnej 2006. Oczywiście zadaniem testów diagnostycznych był ograniczony monitoring pracy zespołów, a nie ocena ich pracy, tym niemniej ich przeprowadzenie i wyniki mogły wytworzyć iluzję właściwej i co ważniejsze porównywalnej pracy zespołów. Według nas tak nie jest. Można oczywiście dyskutować, kiedy mamy do czynienia ze sprawdzianem porównywalnym, a kiedy różnice są na tyle duże, że trzeba podjąć działania je niwelujące.

Wyraźnym sygnałem świadczącym o tym, że wyniki uczniów mogą być zaburzone w większym stopniu niż możemy przypuszczać, było porównanie wyników uczniów z powiatu łowickiego w roku 2006 i 2007. W roku 2006 średni wynik dla uczniów z przywołanego powiatu wyniósł 25,32 punktu i był o 0,16 wyższy od średniej w województwie łódzkim. Analizując strukturę wyników w całym kraju, jako miarę najbardziej obiektywną, oszacowaliśmy oczekiwany wynik dla tego powiatu na podobnym poziomie. Obejmował on zakres od 25 do 25,5 punktu. Wynik osiągnięty empirycznie zmieścił się w tym przedziale.

Podobnej analizie poddaliśmy rezultat powiatu w roku 2007. Wynik łowickiego to 28,61 punktu przy średniej województwa 26,61 punktu. O wybitności uzyskanego przez Łowiczan poziomu świadczyło przewyższenie średniego wyniku krajowego dla wielkiego miasta (28,1). Pragniemy zaznaczyć, że wynik sprawdzianu analizowany w skali kraju (miara obiektywna) wskazuje wyraźnie na zależność osiągniętych wyników od czynnika demograficznego – im większa miejscowość, jeśli chodzi o lokalizację placówki, tym wyższe wyniki.

W powiecie łowickim 35 % dzieci uczęszcza do szkół w mieście średniej wielkości (20 – 100 tys. mieszkańców), a 65 % do szkół wiejskich.

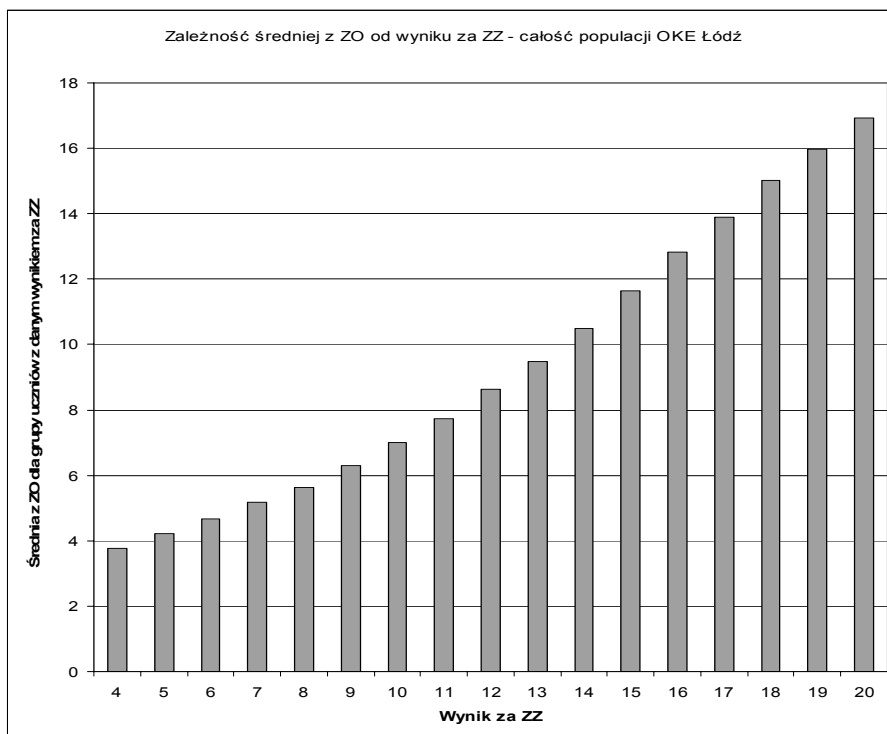
Wykorzystując te dane, oszacowaliśmy przedział oczekiwanego wyniku dla powiatu od 26 do 26,5 punktu. Jak widać wynik empiryczny 28,61 wyraźnie odstaje od naszych oczekiwań. Kolejne wątpliwości wzbudziło porównanie łatwości sprawdzianu w powiecie między rokiem 2006 a 2007. Różnica ta w skali kraju wynosi 1,3 punktu (rok 2006 – średnia 25,3 punktu, rok 2007 – 26,6 punktu) w powiecie łowickim to 3,3 punktu.

Oczywiście w naszych rozważaniach uwzględniliśmy też możliwość eksplozji intelektualnej populacji łowickiej z rocznika 1994. Wykonaliśmy jednak także pewne przeliczenia i porównania, bazując na posiadanych danych wynikowych.

Zasadniczym pytaniem, jakie sobie postanowiliśmy, było: jaka jest zależność punktów uzyskanych przez uczniów za zadania otwarte od ilości punktów zdobytych za zadania zamknięte. Porównanie wyników w powiatach wykazało duże różnice w strukturze uzyskanych punktów. Średnia różnica między ilością punktów za oba typy zadań w całym okręgu wyniosła 3,08 punktu. Dla poszczególnych powiatów różnice wahały się od 2,31 do 3,52 punktu. Czy można na tej podstawie wyciągać konkretne wnioski? Właściwie od razu przychodzi myśl o łagodności i surowości oceniania zadań otwartych przez egzaminatorów. Mała różnica (2,31) to łagodne ocenianie, duża różnica (3,52) to ocenianie surowe. Tu pojawia się kolejne pytanie. Czy i jaki wpływ na analizę ma sytuacja, gdy zadania zamknięte zostały ściągnięte? Tego wykluczyć nie można, trudno sobie jednak wyobrazić, aby problem ten dotyczył wszystkich uczniów w powiecie. Prawdopodobne jest także to, że to patologiczne zjawisko dotyczy w porównywalnym stopniu wszystkich powiatów. Można je również rozciągnąć na zadania otwarte. Zakładając, że ściąganie zaburza wynik egzaminu stosunkowo równomiernie na każdym terenie, przyjęliśmy wynik za zadania zamknięte jako miarę stosunkowo obiektywną.

Intuicja podpowiada nam, że istnieje zależność pomiędzy liczbą punktów za zadania otwarte a punktami zdobytymi za zadania zamknięte z tego samego testu. Wątpliwości budzi tylko to, czy zależność tę można zapisać w postaci funkcji. A jeśli tak, to w jaki sposób ta funkcja wiąże ze sobą badane wielkości?

Przyjrzyjmy się wykresowi 1., na którym przedstawiono zależność średniej ilości punktów za zadania otwarte dla grupy uczniów z konkretną ilością punktów za zadania zamknięte.

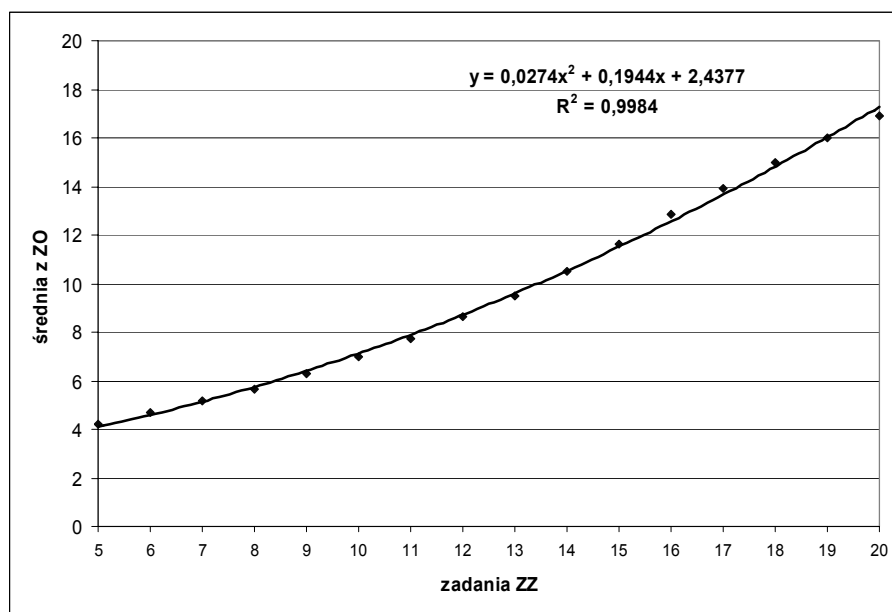


Wykres 1. Zależność średniej z ZO od wyniku za ZZ z całości populacji OKE Łódź.

Łatwo zauważyć, że wysokości kolejnych słupków nie są przypadkowe. Umożliwiło to nam wyłonienie narzędzia, które pozwala dokładnie wyznaczyć zależność wyniku uzyskanego za zadania otwarte od ilości punktów z zadań zamkniętych. Podstawowy warunek, jaki musi być spełniony, aby można je zastosować, to odpowiednio duża grupa wyników uczniowskich. Oczywiście im większa, tym lepiej. Populacje powiatowe są wystarczająco liczne i spełniają założone kryterium. Przejdźmy zatem do samego narzędzia. Jest ono stosunkowo proste, jego działanie polega na przyporządkowaniu wynikom za zadania zamknięte całej populacji (OKE Łódź) wyników za zadania otwarte.

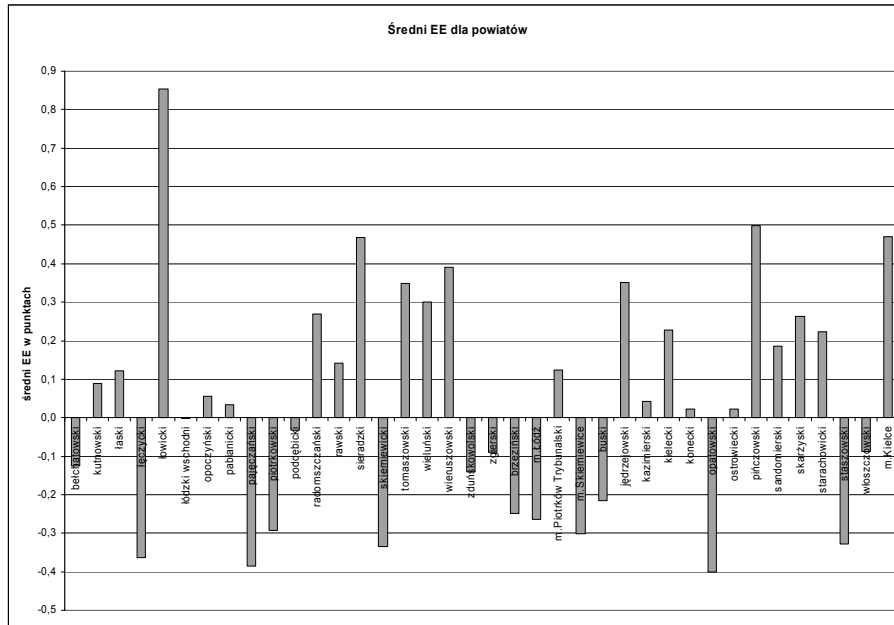
Łatwo zauważyć, że na wykresie została pominięta grupa najniższych wyników za zadania zamknięte. Jest to zabieg celowy związany z ich małą przydatnością, wyniki te wynikają bardziej ze zgadywania przez uczniów niż ich wiedzy. Właściwie dopiero o uczniach, którzy uzyskali powyżej 5 punktów, można powiedzieć, że jest to efekt ich wiedzy. Przebieg funkcji sygnalizuje, że zależność zmiennych jest bardzo wyraźna. Nie jest to funkcja liniowa, a kwadratowa. Współczynnik dopasowania krzywej teoretycznej do

danych empirycznych (tzw. R2) jest bardzo wysoki i wynosi 0,998. Można to interpretować w ten sposób, że teoretyczna krzywa (opisywana funkcją kwadratową) w 99,8% odpowiada danym uzyskanym z badanej populacji.

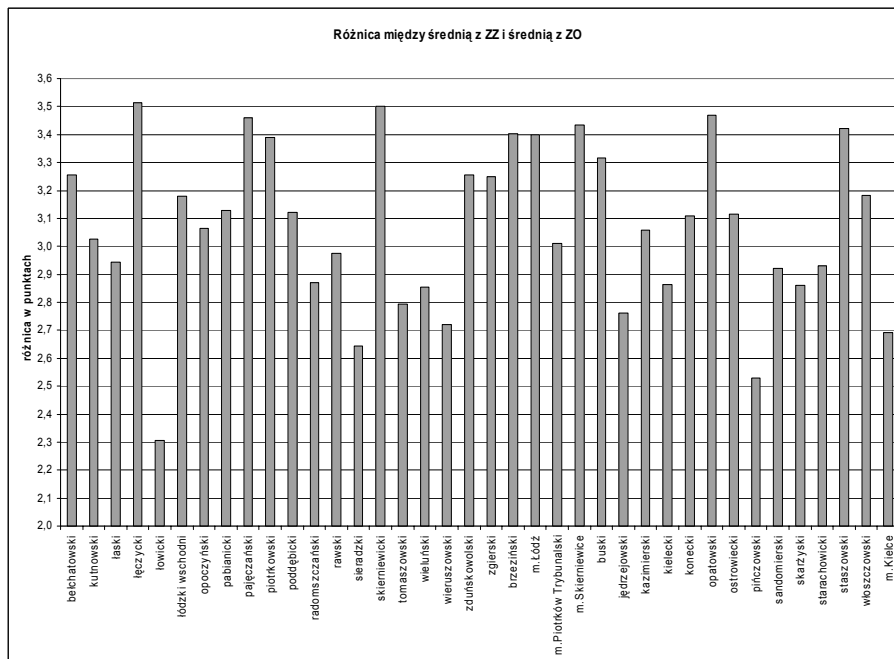


Wykres 2. Zależność średniej z ZO od wyniku za ZZ.

W tym momencie z łatwością można już było wyznaczyć efekt egzaminatora (EE) dla wyników w poszczególnych powiatach – wykres 3. (Efekt egzaminatora obliczamy jako średnią odchyłeń wyników uczniowskich z danego powiatu od oczekiwanej wartości obliczonej na podstawie funkcji uzyskanej z wykresu 2.) Jest on wartościowo bardzo zbliżony do różnic między wynikiem za zadania zamknięte i otwarte – Wykres 4.



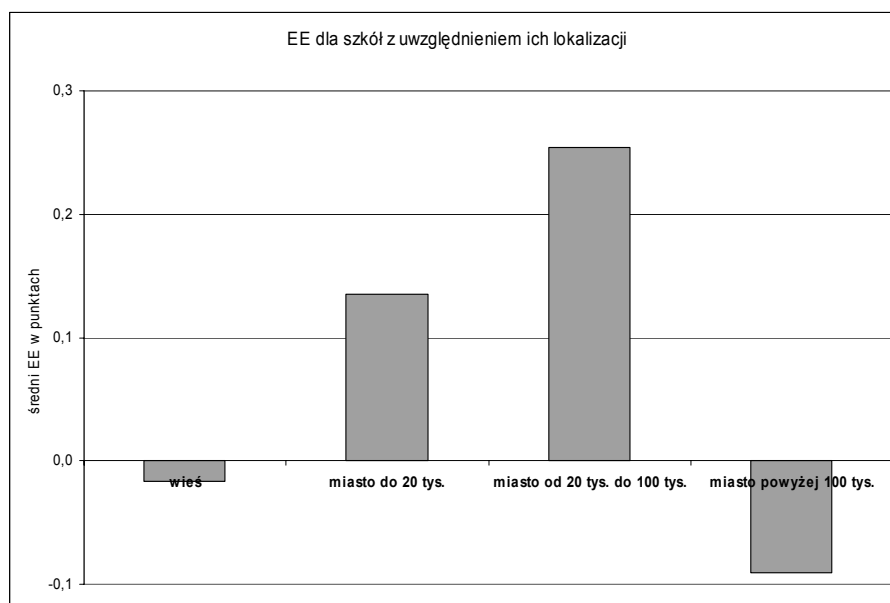
Wykres 3. Efekt egzaminatora



Wykres 4. Różnica między średnimi wynikami za zadania zamknięte i otwarte

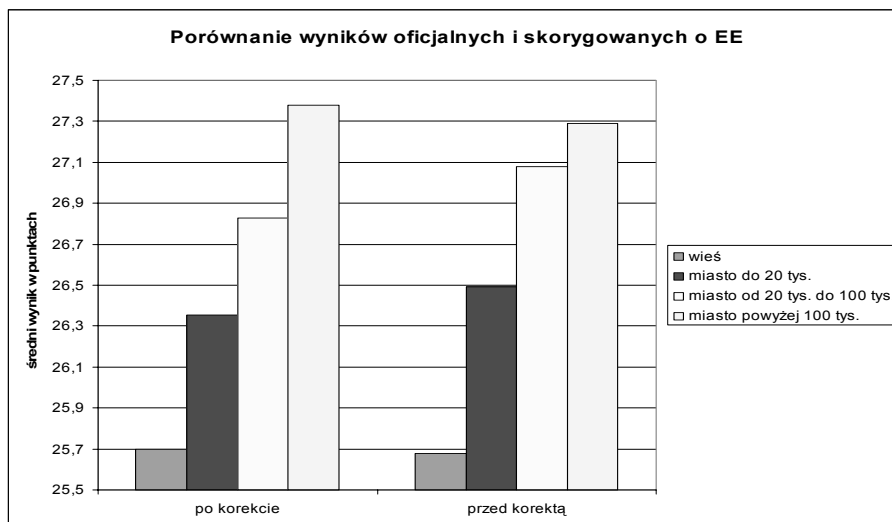
Maksymalna różnica między powiatami wynosi 1,26 punktu. Nie można oczywiście stwierdzić, że egzaminatorzy sprawdzający prace uczniów z powiatu łowickiego „dołożyli” im 0,86 punktu, a w powiecie opatowskim „zabrali” 0,4 punktu, ponieważ średnia 0 jest średnią wszystkich powiatów OKE Łódź. Badanie prowadzone w komisji, którego częścią jest ta prezentacja, wskazuje, że wszystkie wyniki uczniowskie w powiatach są mniej lub więcej przewartościowane. W tej sytuacji bardziej adekwatne wydaje się stwierdzenie: w łowickim wynik został zawyżony o ok. 1,3, a w opatowskim o 0,4, w innych powiatach analogicznie, z uwzględnieniem danych z wykresu 3.

Posługując się wyłonionym przez nas narzędziem „zależności wyniku zadań otwartych od liczby punktów uzyskanych za zadania zamknięte”, prześledziliśmy wyniki populacji okręgu łódzkiego z uwzględnieniem lokalizacji placówki.



Wykres 5.

Bez trudu zauważamy, które grupy zostały ocenione łagodniej, a które surowiej i jakiej skali jest to różnica. Wyniki dla poszczególnych grup po skorygowaniu o efekt egzaminatora (EE) zdecydowanie bardziej korelują z danymi krajowymi (najbardziej obiektywnymi). Ilustruje to następny wykres.



Wykres 6.

Wydaje się, że wyłonione narzędzie rokuje bardzo dużą skuteczność w zakresie kontrolowania i usprawniania porównywalności sprawdzania prac ze sprawdzianu w szkole podstawowej oraz egzaminu gimnazjalnego. Można również prawdopodobnie wykorzystać je na pozostałych poziomach (nie przeprowadziliśmy jeszcze takiej symulacji). Optymalnym byłoby zastosowanie narzędzia do zbadania wyników uczniów uzyskanych przez 10% populacji, o której wspominaliśmy wcześniej. A gdyby jeszcze dla wspomnianej grupy zastosowano czynniki obiektywizujące sprawdzenie zadań otwartych (uczyniono to już w badaniu na terenie OKE Łódź) w bardzo wiarygodnym stopniu można by było rzetelnie określić poziom umiejętności danej populacji w całym kraju.

Jest mało prawdopodobne, aby potwierdziło się, że dwie populacje wielkomijskie mogą różnić się aż o 3 punkty za cały test, jeżeli w umiejętności czytania różnica ta wyniosła ok. 0,3 punktu.

Bibliografia:

1. Dolata R., , *Edukacyjna wartość dodana, czyli jak wykorzystywać wyniki egzaminów zewnętrznych do oceny efektywności nauczania*, Warszawa 2007.
2. *Scholaris. Internetowe Centrum Zasobów Edukacyjnych MEN*. Wyniki sprawdzianu szóstoklasistów i egzaminu gimnazjalnego 2006 i 2007.