

dr Artur Pokropek

Instytut Badań Edukacyjnych

Zespół EWD

Tomasz Żółtak

Uniwersytet Warszawski

Instytut Badań Edukacyjnych

Zespół EWD

Nowe modele jednorocznej EWD

Wstęp

W roku 2012 egzamin gimnazjalny przeprowadzony został według nowych zasad. Po raz pierwszy sprawdzał on opanowanie wiadomości i umiejętności uczniów nauczanych wedle nowej, wprowadzonej w 2009 roku, podstawy programowej. Tak jak poprzednie edycje, składał się z trzech części (humanistycznej, matematyczno-przyrodniczej oraz części dotyczącej języka obcego nowożytnego) jednak struktura każdej z części uległa znacznym zmianom. W nowej formule każda część składa się z dwóch oddzielnych testów. W części humanistycznej jest to test składający się z zadań z języka polskiego oraz drugi test, składający się z zadań z zakresu historii i wiedzy o społeczeństwie. Część matematyczno-przyrodnicza składa się z testu skomponowanego z zadań matematycznych oraz testu z zakresu przedmiotów przyrodniczych: biologii, chemii, fizyki i geografii. Część językowa rozbita została na dwa testy o różnych poziomach trudności: poziomie podstawowym i rozszerzonym. W związku z tymi zmianami wyniki egzaminacyjne zaczęto komunikować w sześciu zakresach (wcześniej w trzech).

Zmiana egzaminu gimnazjalnego wymusiła refleksje nad miarami efektywności kształcenia opartymi na tym egzaminie. Niniejszy tekst przedstawia nowe rozwiązania, jakie zostały wprowadzone do jednorocznego modelu EWD dla szkół gimnazjalnych. Omówienie zmian poprzedzone zostało analizą psychometryczną egzaminu gimnazjalnego, której wyniki implikowały serię decyzji dotyczących modyfikacji modelu.

Egzamin gimnazjalny 2012

Edukacyjna Wartość Dodana dla szkół gimnazjalnych do roku 2012 komunikowana była dla dwóch części egzaminu gimnazjalnego: humanistycznej oraz matematyczno-przyrodniczej. Z języka obcego zrezygnowano ze względu na dużą liczbę języków, które mogą być wybrane przez ucznia, braku pomiaru analogicznej umiejętności pod koniec szkoły podstawowej oraz wątpliwości co do trafności tej miary w dobie powszechnie działających szkół języków obcych oraz prywatnych korepetycji. Nowa formuła egzaminu gimnazjalnego wprowadza cztery miary osiągnięć szkolnych istotnych dla EWD. Taka zmiana zrodziła dwa pytania, które musiały zostać zadane w kontekście rozwoju metody EWD:

1. Czy takie cztery wyniki egzaminacyjne mogą stać się podstawą dla czterech oddzielnych miar EWD?
2. Czy rozbięcie każdej z części egzaminu gimnazjalnego na dwa testy dalej pozwala raportować EWD w odniesieniu do dwóch ogólnych umiejętności?

Aby odpowiedzieć na pierwsze pytanie, posłużyliśmy się analizą rzetelności testów gimnazjalnych, która była szacowana na cztery różne sposoby. Pierwszym z nich była Alfa Cronbacha, klasyczna miara pozwalająca oszacować dolną granicę rzetelności testu. Oznacza to, że rzetelność testu może być nieco wyższa od tej, na jaką wskazuje Alfa Cronbacha. Wartość tego wskaźnika wyraża się wzorem:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_{suma}^2} \right) \quad (1)$$

gdzie:

k - liczba zadań w teście,

σ_i^2 - wariancja dla i -tego zadania,

σ_{suma}^2 - wariancja dla wyniku sumarycznego.

Alfa Cronbacha, jakkolwiek prosta do wyliczenia i szeroko stosowana, ma jednak tę wadę, że nie najlepiej nadaje się do szacowania rzetelności testów, w których poszczególne zadania mają różne długości skal. W takim przypadku będzie przejawiać tendencję do zaniżania oszacowywanych wartości rzetelności (Lord i Novick, 1968).

Jako alternatywną miarę rzetelności, odporną na różnice w długości skali poszczególnych zadań, posłużyliśmy się więc oszacowaniem na podstawie wyników dwuparametrycznego modelu IRT, wyestymowanego w odniesieniu do danych z egzaminu gimnazjalnego. Podobnie jak w KTT, przy podejściu odwołującym się do IRT, wariancję wyniku prawdziwego można rozbić na dwa składniki:

$$D^2(\hat{\theta}) = E[D^2(\theta|x)] + D^2[E(\theta|x)] \quad (2)$$

gdzie:

x oznacza wzór odpowiedzi ucznia (uczniowie o tym samym wzorze odpowiedzi to tacy, którzy uzyskali dokładnie taką samą punktację za każde z zadań w teście), θ prawdziwy poziom umiejętności, a $\hat{\theta}$ przewidywany poziom umiejętności.

Pierwszy składnik odnosi się do niepewności oszacowania parametru umiejętności dla uczniów charakteryzujących się tym samym wzorem odpowiedzi. Wartość oczekiwana tego parametru w przestrzeni możliwych wzorów odpowiedzi to średnia niepewność szacowania umiejętności uczniów. Natomiast drugi komponent to wariancja szacowanych umiejętności uczniów o różnych wzorach odpowiedzi. Tak jak w klasycznej teorii testu, wariancja wyniku jest więc sumą wariancji wynikającej z błędu pomiaru i wariancji wyników prawdziwych. W związku z tym indeks rzetelności może zostać obliczony jako stosunek wariancji wyników prawdziwych do całkowitej wariancji wyników pomiaru (Scheerens i in., 2003: 148):

$$\rho = \frac{D^2[E(\theta|x)]}{D^2(\hat{\theta})} \quad (3)$$

Każdą miarę rzetelności obliczyliśmy w dwóch wariantach: (a) traktując każdy podpunkt zadania jako oddzielne zadanie¹ (jest to sposób szacowania rzetelności przyjęty przez CKE) oraz (b) sumując wyniki uzyskane za poszczególne podpunkty w ramach zadania (jeśli występowały), tworząc jedno zadanie o dłuższej skali. W drugim przypadku zredukowana była oczywiście liczba zadań.

W tabeli 1. przedstawiono wartości szacowanych wskaźników. Wyniki są spójne. Alfa Cronbacha daje najniższe oszacowanie (dolna granica rzetelności), oszacowania na podstawie IRT są nieco wyższe. Szacowanie rzetelności na podpunktach traktowanych jako osobne zadania wyraźnie zwiększa wskaźnik rzetelności w przypadku Alfy Cronbacha, w stosunku do sytuacji gdy wyniki za podpunkty są sumowane w ramach zadań, nie ma to jednak dużego znaczenia w przypadku szacowania rzetelności za pomocą IRT.

Tabela 1. Rzetelność testów gimnazjalnych

		GHP	GHU	GPL	GMP	GPR	GMA
Liczba zadań		56	29	27	49	26	23
(a)	alfa	0.905	0.826	0.848	0.883	0.729	0.841
	IRT	0.915	0.848	0.865	0.912	0.776	0.877
Liczba zadań		47	25	22	47	24	---
(b)	alfa	0.879	0.822	0.757	0.882	0.724	---
	IRT	0.907	0.846	0.838	0.912	0.775	---

GHP - część humanistyczna łącznie; GHU - historia i WOS; GPL - język polski; GMP - część matematyczno-przyrodnicza łącznie; GPR - przyroda; GMA - matematyka.

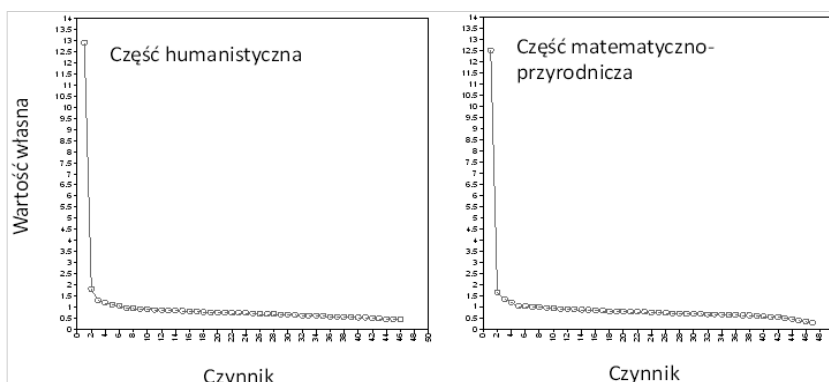
(a) Podpunkty traktowane jako oddzielne zadania. (b) Sumowanie wyników za podpunkty w ramach zadań.

W teście matematycznym wszystkie zadania były punktowane na skali 0-1.

Za dopuszczalną granicę rzetelności skal przyjętą w pomiarze edukacyjnym przyjęło się traktować wartość 0,9 dla skali głównej oraz 0,8 dla podskal. Jeżeli traktować poszczególne testy egzaminu gimnazjalnego jako podskale, tylko test przyrodniczy nie spełnia tego kryterium (bez względu na to, jaką metodą liczona jest rzetelność). Ponieważ wyniki edukacyjnej wartości dodanej nie są komunikowane w odniesieniu do uczniów, a w odniesieniu do większych agregatów (szkoła lub w przypadku wskaźników jednorocznych oddział), kryterium minimalnej wartości rzetelności może jednak zostać nieco obniżone. Wartości oscylujące wokół 0,75 wydają się być wystarczająco wysokie na potrzeby szacowania EWD. Można zatem powiedzieć, że nowe testy gimnazjalne są wystarczająco rzetelne, aby mogły się stać się podstawą dla czterech oddzielnych miar EWD.

Rzetelność liczona dla poszczególnych części egzaminu gimnazjalnego (humanistycznej i matematyczno-przyrodniczej) oscyluje w granicach 0,9, co pozwala wierzyć, że choć każda z nich powstała w wyniku połączenia dwóch testów, tworzą one wewnętrznie spójne skale. Aby potwierdzić to przypuszczenie, odwołaliśmy się do eksploracyjnej analizy czynnikowej (EFA) dla zmiennych porządkowych. Na rysunku 1. przedstawiony został wykres osypiska uzyskany za pomocą tej analizy.

¹ Wyjątek stanowiły zadania 15. i 24. z testu z historii i wiedzy o społeczeństwie, gdyż w zależności od wersji testu ich podpunkty miały różne schematy punktowania. W związku z tym nie ma możliwości, aby rozpatrywać wyniki tych zadań w rozbiciu na podpunkty.



Rysunek 1. Wykresy osypiska dla dwóch części egzaminu gimnazjalnego

Wykres osypiska wskazuje na wartości własne kolejnych czynników, a więc pośrednio na odsetek wariacji wyników, który może zostać opisany przy ich pomocy. Jak widać, w obydwu przypadkach znaczna część wariacji umiejętności mierzonych obydwoma testami przypisywana jest pierwszemu czynnikowi. Kolejne czynniki charakteryzują się znacznie mniejszymi wartościami własnymi, a ich uwzględnienie nie poprawia radykalnie ilości wyjaśnianej przez model wariacji. Wyniki te uzasadniają traktowanie sumarycznej liczby punktów w ramach każdej z części egzaminu gimnazjalnego jako wskaźników odnoszących się do ogólnych umiejętności humanistycznych lub matematyczno-przyrodniczych.

Wyniki przedstawionych analiz sugerują, iż egzamin gimnazjalny może być wykorzystany w modelowaniu EWD co najmniej na dwa sposoby. Po pierwsze, egzamin może stać się bazą dla czterech wskaźników odnoszących się do szczegółowych umiejętności mierzonych osobnymi testami w każdej z części egzaminu (przyroda, matematyka, język polski, historia z WOS). Po drugie, można zachować ogólne miary umiejętności, znane ze wcześniejszych edycji egzaminu: humanistyczną i matematyczno-przyrodniczą, gdyż dwa testy z jednej części są dosyć spójne i dają się przedstawić w jednowymiarowej formie.

Skala prezentacji wyników egzaminacyjnych

Jednocześnie ze zmianą liczby dostarczanych wskaźników wprowadzona została istotna zmiana w sposobie prezentacji wyników egzaminacyjnych i wartości jednorocznych miar EWD. Dotychczas były one przedstawiane na skali „surowych” punktów egzaminacyjnych. Od bieżącego roku wprowadzona zostaje, znana już z trzyletnich wskaźników EWD dla gimnazjów, prezentacja wyników egzaminacyjnych na znormalizowanej skali o średniej 100 i odchyleniu standardowym 15 (podobnie jak w przypadku wskaźników trzyletnich; por. Pokropek, 2009). Zmiana ta powinna ułatwić porównywanie wyników z różnych lat oraz z różnych testów, zwłaszcza że w ramach nowej formuły egzaminu gimnazjalnego poszczególne testy różnią się co do liczby punktów, jakie można w nich uzyskać².

² Taka sama rozpiętość skali testów nie jest oczywiście warunkiem wystarczającym, aby móc bezpośrednio porównywać ich surowe wyniki, niemniej jest do tego warunkiem koniecznym.

W wyniku normalizacji wynik danego ucznia przedstawiany jest w odniesieniu do wyników wszystkich zdających, przy założeniu, że w populacji rozkład umiejętności badanej przez test jest normalny. Przeliczenie dokonywane jest tzw. metodą Hazena (Barnett 1974) w oparciu o wzór:

$$U(X = x_i) = 100 + 15 \cdot \Phi^{-1} \left(\frac{N(X \leq x_i) - \frac{N(X=x_i)}{2} - 0,5}{n} \right) \quad (4)$$

gdzie:

$U(X=x_i)$ - wynik znormalizowany dla wyniku surowego równego x_i ,

Φ^{-1} - funkcja odwrotna do dystrybuanty rozkładu normalnego standaryzowanego,

$N(X \leq x_i)$ - liczba zdających z wynikiem niższym niż x_i ,

$N(X = x_i)$ - liczba zdających z wynikiem surowym równym x_i ,

n - liczba zdających.

Ponieważ liczba osób, które uzyskują na sprawdzianie najniższe wyniki, jest bardzo niewielka, na potrzeby normalizacji zdecydowaliśmy się wszystkie osoby, które otrzymały z tego egzaminu nie więcej niż 3 punkty traktować tak, jakby otrzymały ten sam wynik. Zapobiega to przypisywaniu pojedynczym osobom ekstremalnie dalekich od średniej wyników znormalizowanych i wpływa pozytywnie na własności modeli jednorocznej EWD.

Wynik znormalizowany wyższe niż 100 wskazuje, że uczeń uzyskał z danego testu rezultat wyższy niż średnia wyników wszystkich zdających. Wyniki znormalizowane można też łatwo przeliczyć na rozpowszechnioną w polskiej edukacji skalę staninową. Wiąże się to w oczywisty sposób z utratą znacznej ilości informacji o uczniach, ale może być pomocne przy oswojaniu się z interpretacją wyników na nowej skali. Przypisanie wyników na skali 100;15 do staninów pokazuje tabela 2. Wskaźniki EWD prezentowane są na skali o średniej 0 i jednostce równej jednostce skali wyników egzaminacyjnych.

Tabela 2. Przeliczanie wyników na skali 100;15 na skalę staninową

Zakres wyników na skali 100;15	<73,75	73,75-81,25	81,25-88,75	88,75-96,25	96,25-103,75	103,75-111,25	111,25-118,75	118,75-126,25	>126,25
Stanin	1	2	3	4	5	6	7	8	9

Warto zaznaczyć, że znormalizowane wyniki dla dwóch części egzaminu gimnazjalnego (humanistycznej i matematyczno-przyrodniczej) wyliczane są na podstawie rozkładu sumy wyników surowych z odpowiednich dwóch testów, a nie z przekształcenia ich wyników znormalizowanych. W szczególności znormalizowany wynik danej części egzaminu gimnazjalnego nie jest średnią z wyników znormalizowanych dwóch tworzących daną część testów. Zależność pomiędzy znormalizowanymi wynikami testów tworzących daną część i znormalizowanym wynikiem dla tej części jako całości wyniku z łącznego rozkładu wyników tych dwóch testów, jest złożona i nie daje się łatwo opisać.

Normalizacja wyników egzaminu gimnazjalnego przeprowadzana jest w oparciu o dane otrzymane przez Zespół EWD od OKE w wyniku zamówienia na połączone wyniki egzaminu gimnazjalnego i sprawdzianu (zawiera on również

wyniki egzaminu gimnazjalnego uczniów, którym nie udało się przyłączyć wyników sprawdzianu). Normalizacja wykonywana jest z uwzględnieniem wszystkich zdających egzamin gimnazjalny po raz pierwszy, z wyjątkiem uczniów szkół dla dorosłych, specjalnych i przyszpitalnych (szkoły te identyfikowane są na podstawie prowadzonej przez Zespół EWD bazy gimnazjów). Ponieważ wyniki sprawdzianu w zbiorze połączonych wyników egzaminacyjnych nie są kompletne (ze względu na niemożliwość przyłączenia wyników sprawdzianu niektórym osobom oraz występowanie uczniów o wydłużonym toku kształcenia), normalizacja wyników sprawdzianu przeprowadzana jest w oparciu o odpowiedni zbiór wyników otrzymywany przez CKE od OKE w czerwcu każdego roku, na potrzeby przygotowania sprawozdania ze sprawdzianu. Przy normalizacji uwzględniane są wszystkie obserwacje z tego zbioru, gdyż brak jest bazy pozwalających określić charakter poszczególnych szkół podstawowych i wyłączyć uczniów placówek specjalnych czy przyszpitalnych.

Modele jednorocznej EWD 2012

W modelach jednorocznej EWD 2012 po raz pierwszy uwzględnieni zostali uczniowie gimnazjów, których cykl kształcenia uległ wydłużeniu o rok. W odniesieniu do nich estymowana jest odrębna linia regresji określająca zależność pomiędzy wynikami sprawdzianu a wynikami danego testu lub danej części egzaminu gimnazjalnego. Wielkość efektów opisujących wpływ płci i dysleksji estymowana jest z kolei łącznie, biorąc pod uwagę jednocześnie uczniów o normalnym i wydłużonym o rok toku kształcenia. Formalnie model jednorocznej EWD 2012 dla danego testu, lub części, egzaminu gimnazjalnego opisuje równanie:

$$Y_i = w(X_i, tok_i) + b_p płęć_i + b_{dg} dysl_g_i + b_{ds} dysl_s_i + b_{dgs} (dysl_g_i \cdot dysl_s_i) + r_i \quad (5)$$

gdzie:

Y_i - wynik i -tego ucznia z danego testu lub części egzaminu gimnazjalnego,

X_i - wynik i -tego ucznia ze sprawdzianu,

tok - tok kształcenia i -tego ucznia (0 - normalny, 1 - wydłużony),

$w(X_i, tok_i)$ - wielomian opisujący zależność pomiędzy wynikami sprawdzianu a wynikami danego testu lub części egzaminu gimnazjalnego,

$płęć_i$ - płeć i -tego ucznia (0 - męczyzna, 1 - kobieta),

$dysl_g_i$ - posiadanie przez i -tego ucznia zaświadczenia o dysleksji na egzaminie gimnazjalnym (0 - nie, 1 - tak),

$dysl_s_i$ - posiadanie przez i -tego ucznia zaświadczenia o dysleksji na sprawdzianie (0 - nie, 1 - tak),

r_i - reszta regresji dla i -tego ucznia.

Funkcja $w(X_i, tok)$ jest z kolei postaci:

$$w(X_i, tok_i) = b_0 + b_1 X_i + \dots + b_k X_i^k + b_{w0} tok + b_{w1} X_i tok + \dots + b_{wk} X_i^k tok \quad (6)$$

a więc linia regresji dla uczniów o standardowym toku kształcenia³ opisywana jest wzorem:

$$w(X_i, tok = 0) = b_0 + b_1 X_i + \dots + b_k X_i^k \quad (7)$$

a dla uczniów o toku kształcenia wydłużonym o rok wzorem:

$$w(X_i, tok = 1) = (b_0 + b_{w0}) + (b_1 + b_{w1})X_i + \dots + (b_k + b_{wk})X_i^k \quad (8)$$

Problemem, który należy rozstrzygnąć, jest wybór odpowiedniej wartości parametru k , czyli stopnia wielomianu. Przyjęto tu zasadę, że jest to najmniejsza liczba naturalna, dla której model z linią regresji opisywaną wielomianem tego stopnia spełnia dwa warunki:

1. Linia regresji jest monotoniczna (rosnąca) w przedziale możliwych wyników sprawdzianu, zarówno dla uczniów o standardowym, jak i wydłużonym o rok toku kształcenia.
2. Wyniki testu „linktest”, w odmianie resztowej (por. Cameron i Trivedi, 2010), zarówno dla uczniów o standardowym, jak i wydłużonym o rok toku kształcenia wskazują, że dalsze podnoszenie stopnia wielomianu nie przyniesie istotnej statystycznie poprawy dopasowania modelu do danych.

Kryterium drugie wymaga tu nieco szerszego omówienia. „Linktest” jest testem diagnostycznym, pozwalającym na ocenę, czy model może zostać udoskonalony poprzez włączenie do niego jako dodatkowego predyktora nieliniowego przekształcenia którejś z już wykorzystywanych zmiennych niezależnych lub ich kombinacji liniowej. W klasycznej postaci przeprowadza się go, rozpatrując istotność współczynnika b_{l2} regresji postaci:

$$Y_i = b_{l0} + b_{l1} \hat{Y}_i + b_{l2} \hat{Y}_i^2 + r_{li} \quad (9)$$

gdzie:

$$\hat{Y}_i = w(X_i, tok = t) + b_{p} płęć_i + b_{dg} dysl_g_i + b_{as} dysl_s_i + b_{ags}(dysl_g_i \cdot dysl_s_i) \quad (10)$$

to wartość zmiennej zależnej przewidywana dla i -tego ucznia w ramach modelu poddawanego testowi. Równanie (9) i (10) muszą być przy tym rozpatrywane oddzielnie dla uczniów o standardowym i wydłużonym toku kształcenia.

Istotność statystyczna współczynnika b_{l2} sugeruje potrzebę wprowadzenia do modelu dodatkowej zmiennej, będącej nieliniowym przekształceniem którejś ze zmiennych już obecnych w modelu. Nieistotność wskazuje, że takie działanie nie doprowadzi do istotnej poprawy jakości dopasowania modelu do danych.

Pewną komplikacją w przypadku modeli EWD gimnazjów stanowi, że jedyną zmienną, co do której zakładamy możliwość wprowadzania do modelu jej przekształconej nieliniowo postaci, jest wynik ucznia ze sprawdzianu. W związku z tym rozsądne wydaje się zastosowanie takiej modyfikacji „linktestu”, która

³ Dokładnie dla mężczyzn, którzy nie posiadali zaświadczenia o dysleksji ani na egzaminie gimnazjalnym, ani na sprawdzianie.

pozwaląaby ocenić zasadność tego działania, nie biorąc jednocześnie pod uwagę skutków ewentualnego nieliniowego przekształcenia innych zmiennych niezależnych obecnych w modelu. Postulat ten spełnia odmiana „linktestu” postaci:

$$Y_i - \left[b_p \text{płeć}_i + b_{dg} \text{dysl}_{g_i} + b_{ds} \text{dysl}_{s_i} + b_{dgs} \left(\text{dysl}_{g_i} \cdot \text{dysl}_{s_i} \right) \right] = \\ = b_{l0} + b_{l1} w(X_i, \text{tok} = t) + b_{l2} [w(X_i, \text{tok} = t)]^2 + r_{li} \quad (11)$$

Po lewej stronie równania mamy więc wyniki testu lub części egzaminu gimnazjalnego, które skorygowano o wpływ przypisywany płci i dysleksji. Po prawej stronie znalazły się z kolei przewidywanie związane bezpośrednio z wynikami sprawdzianu, $w(X_i, \text{tok}=t)$ oraz jego kwadrat (a poza tym stała b_{l0} i reszta r_{li} , które jednak same w sobie nie są obiektem naszego zainteresowania). Wyniki testu interpretujemy tak samo jak w standardowej odmianie „linktestu”, sprawdzając istotność współczynnika b_{l2} . Również w tym przypadku test musi zostać przeprowadzony oddzielnie dla uczniów o standardowym i wydłużonym toku kształcenia.

W przypadku modeli jednorocznych dla zdających egzamin gimnazjalny w 2012 r. zastosowanie omówionych kryteriów doprowadziło do wyboru wielomianu stopnia trzeciego dla modelu opisującego część humanistyczną i wielomianów piątego stopnia w przypadku pozostałych modeli. Wyestymowane wartości współczynników modeli zostały zestawione w tabeli 3.

Tabela 3. Współczynniki modeli jednorocznej EWD gimnazjów 2012

Zmienna	GHP	GHU	GPL	GMP	GPR	GMA
b_0 (stała)	101,13	-10,17	99,75	4,91	95,33	-155,48
b_1 (spr.)	-1,77	6,68	-0,26	4,90	-0,11	13,78
b_2 (spr. ²)	2,67E-02	-1,89E-01	-3,72E-02	-1,27E-01	-1,81E-02	-3,10E-01
b_3 (spr. ³)	-9,14E-05	2,44E-03	9,03E-04	1,52E-03	3,80E-04	3,29E-03
b_4 (spr. ⁴)		-1,41E-05	-6,71E-06	-7,84E-06	-2,25E-06	-1,61E-05
b_5 (spr. ⁵)		3,05E-08	1,65E-08	1,45E-08	4,15E-09	2,92E-08
b_p (płeć-kobieta)	0,89	-2,43	4,35	-2,50	-1,78	-2,66
b_{dg} (dysleksja egz. gimn.)	1,80	1,78	1,56	2,03	1,90	1,89
b_{ds} (dysleksja spr.)	-5,69	-4,55	-6,06	-5,12	-4,55	-4,79
b_{dgs} (dysleksja egz. gimn. i spr.)	0,46	-0,21	1,01	0,03	-0,10	-0,05
b_{w0} (wydł. tok kształcenia)	8,63	-89,17	-163,45	-172,92	-267,56	65,96
b_{w1} (interakcja: wydł. tok kształc. - spr.)	-0,24	5,51	9,60	10,96	16,55	-3,21
b_{w2} (interakcja: wydł. tok kształc. - spr. ²)	2,12E-03	-1,40E-01	-2,24E-01	-2,76E-01	-4,05E-01	5,11E-02
b_{w3} (interakcja: wydł. tok kształc. - spr. ³)	-1,28E-05	1,82E-03	2,59E-03	3,45E-03	4,88E-03	-2,48E-04

Zmienna	GHP	GHU	GPL	GMP	GPR	GMA
b_{ws} (interakcja: wydł. tok kształc. - spr. ⁴)		-1,19E-05	-1,51E-05	-2,13E-05	-2,91E-05	-8,51E-07
b_{ws} (interakcja: wydł. tok kształc. - spr. ⁵)		3,10E-08	3,49E-08	5,13E-08	6,80E-08	7,10E-09
R ²	0,595	0,498	0,563	0,596	0,467	0,551
N	377 805	377 808	377 805	377 710	377 715	377 710

GHP - część humanistyczna łącznie; GHU - historia i WOS; GPL - język polski; GMP - część matematyczno-przyrodnicza łącznie; GPR - przyroda; GMA - matematyka.

Technicznie rzecz ujmując, EWD jednoroczne dla wybranych grup (szkół lub oddziałów szkolnych) liczona jest jako średnia reszt regresji. Reszta regresji jest różnicą między wynikiem prawdziwym a wynikiem przewidywanym przez model: $r_i = y_i - \hat{y}_i$.

Diagnostyka modeli

Diagnostyka graficzna wyliczonych modeli wskazuje, że dobrze spełniają one założenia regresji MNK. W żadnym z modeli nie odnotowujemy związków pomiędzy wartościami przewidywanymi a średnimi wartościami reszt regresji ani poziomem ich rozproszenia. Jedynie w przypadku rozkładów reszt odnajdujemy niewielkie odchylenia od założenia o normalności, jednak nie mają one istotnego wpływu na jakość wyliczonych modeli.

Niestety należy zanotować, że wyliczone na podstawie modeli jednoroczne wskaźniki EWD gimnazjów pozostają w wyraźnym, a niepożądanym związku ze średnimi wynikami sprawdzianu uczniów szkoły. W zależności od wskaźnika wartość współczynnika korelacji liniowej osiąga tu wartość od 0,243 do 0,318. O ile w przypadku części matematyczno-przyrodniczej są to wartości porównywalne z tymi, jakie notowano w przypadku jednorocznych wskaźników EWD gimnazjów wyliczanych w latach wcześniejszych, o tyle w przypadku wskaźników humanistycznych siła tych związków kształtowała się wcześniej na nieco niższym poziomie (w granicach 0,160-0,190) (Żółtak, 2011).

Podsumowanie

Tekst przedstawia nowe rozwiązania, jakie zostały wprowadzone do jednorocznego modelu EWD dla szkół gimnazjalnych. Użytkownicy EWD jednorocznego od 2012 mogą korzystać z sześciu miar EWD, dwóch ogólnych: EWD matematyczno-przyrodniczego i humanistycznego (tak jak we wcześniejszych latach) oraz czterech szczegółowych: EWD dla języka polskiego, EWD dla historii i WOS, EWD dla matematyki oraz EWD dla przedmiotów przyrodniczych. Takie rozwiązanie przyjęte zostało po wprowadzeniu nowej formuły egzaminu gimnazjalnego i zgodne jest z psychometrycznymi właściwościami nowych egzaminów. Poza zwiększeniem liczby wskaźników stosowana wcześniej metodologia EWD uległa kilku istotnym zmianom. Do modelu wprowadzeni zostali uczniowie o przedłużonym toku nauki, umożliwia to szacowanie EWD szkół również na podstawie wyników uzyskiwanych przez tych uczniów, co we wcześniejszych wersjach EWD nie było

możliwe. Wyniki egzaminów gimnazjalnych w modelu jednorocznym EWD od roku 2012 będą normalizowane przed fazą modelowania i wystandaryzowane do skali o średniej 100 i odchyleniu standardowym 15, co upodabnia model jednoroczny do modelu trzyletniego. Opisane zmiany zwiększają możliwości wykorzystania EWD w celach ewaluacyjnych i stanowią istotny krok w rozwoju metody EWD w Polsce. Należy zaznaczyć, iż wprowadzone zmiany mają też negatywne aspekty związane z tym, iż nowe miary nie będą w pełni porównywalne ze wskaźnikami szacowanymi wcześniej. Jest to jednak cena za otwarcie nowych możliwości, jakie dają nowe wskaźniki.

Bibliografia:

1. Barnett, V. 1975. *Probability Plotting Methods and Order Statistics*. Journal of the Royal Statistical Society: Series C (Applied Statistics): 24(1), s. 95-108.
2. Cameron, T. C. i P. K. Trivedi 2010, *Microeconometrics using Stata*, Gichigan: StataPress.
3. Lord F. M., Novick M. R., Birnbaum A., 1968, *Statistical theories of mental test scores*, Reading: Addison-Wesley.
4. Pokropek A., *Metody statystyczne wykorzystywane w szacowaniu trzyletnich wskaźników egzaminacyjnych* [w:] Niemierko B., Szmigel M.K. (red.), *Badania międzynarodowe i wzory zagraniczne w diagnostyce edukacyjnej*, PTDE, Kraków 2009, s. 137-153.
5. Scheerens J., Glas C. A. W., Thomas S. M., 2003, *Educational evaluation, assessment, and monitoring: a systemic approach*, Lisse: Swets and Zeitlinger.
6. Żółtak T., *Znaczenie informacji o średnim wyniku uczniów na wejściu dla własności jednorocznych wskaźników EWD gimnazjów* [w:] Niemierko B., Szmigel M.K. (red.), *Ewaluacja w edukacji: koncepcje, metody, perspektywy*, PTDE, Kraków 2011, s. 505-523.