

Tomasz Żółtak

Uniwersytet Warszawski

Zespół badawczy EWD

Edukacyjna wartość dodana dla nauczycieli w Polsce: ograniczenia, możliwości, perspektywy¹

Edukacyjna wartość dodana, w skrócie EWD, staje się coraz szerzej znaną i cioną miarą służącą do ewaluacji pracy szkół. W odniesieniu do gimnazjów oszacowania EWD są już dostępne publicznie za pośrednictwem serwisu internetowego. Mogą z nich korzystać dyrektorzy, nauczyciele, organy prowadzące czy rodzice. Szkoły wyposażone zostały dodatkowo w kalkulatory EWD, dzięki którym mogą dokonywać bardziej szczegółowych analiz, takich jak liczenie EWD osobno dla poszczególnych klas, w podziale wg płci czy w podziale wg potencjału edukacyjnego uczniów. W tym kontekście pojawia się pytanie, czy metodologia EWD nie mogłaby zostać adaptowana również na potrzeby ewaluacji pracy nauczycieli.

Jak dotąd problem ten podjął w Polsce jedynie Przemysław Majkut (2008), ale jego wystąpienie na XIV Konferencji Diagnostyki Edukacyjnej należy traktować jedynie jako zagajenie do dyskusji. Jednocześnie problematyka EWD dla nauczycieli jest już od z górą dekady szeroko dyskutowana w Stanach Zjednoczonych, gdzie w większości wdrożonych systemów ewaluacji pracy placówek edukacyjnych wylicza się EWD właśnie na poziomie nauczycieli, a nie szkół (por. Sanders i Horn, 1994; McCaffrey i in., 2003, McCaffrey, 2004). Warto przy tym zaznaczyć, że choć miary EWD są cały czas doskonalone, badacze wskazują na istotne problemy ograniczające możliwość traktowania ich jako jedyne i pewne wskaźniki jakości pracy nauczycieli (m.in. Raudenbush, 2004; Harris, 2008; Rothstein, 2008). Za bardzo kontrowersyjne uważane jest w szczególności opieranie systemów motywacyjnych dla nauczycieli jedynie na wynikach EWD (p. Harris, 2008). Różnice pomiędzy polskim a amerykańskim systemem oświatowym (por. Majkut, 2008, s. 317), zarówno co do organizacji nauczania, jak i roli egzaminów zewnętrznych, są przy tym znaczące i często uniemożliwiają bezpośrednie adaptowanie wniosków wynikających z badań amerykańskich. Rodzi to potrzebę podjęcia badań nad własnościami EWD dla nauczycieli również w specyficznym, polskim kontekście.

Rozważenia wymagają w szczególności trzy kwestie. Po pierwsze, jakie trudności i ograniczenia napotykamy przy próbie zdefiniowania i zmierzenia jakości pracy nauczycieli (a właściwie jego wpływu na relatywny postęp ucznia w zdobywaniu wiedzy i umiejętności testowanych na egzaminach zewnętrznych), w kontekście organizacji nauczania w gimnazjum i formy egzaminów zewnętrznych. Po drugie, jakie techniki estymacji mogą zostać wykorzystane

¹ Badania nad EWD realizowane są w ramach projektu Centralnej Komisji Egzaminacyjnej *Badania dotyczące rozwoju metodologii szacowania wskaźników edukacyjnej wartości dodanej (EWD)*, współfinansowanego przez Unię Europejską w ramach Europejskiego Funduszu Społecznego – Program Operacyjny *Kapitał Ludzki*, Priorytet III *Wysoka jakość systemu oświaty*, Działanie 3.2 *Rozwój systemu egzaminów zewnętrznych*.

do wyliczania EWD dla nauczycieli. Jakie są ich zalety i ograniczenia? Po trzecie, trzeba spróbować ocenić, na ile trafna jest proponowana miara, jaki jest zakres jej stosowalności i praktyczna użyteczność.

W poniższym artykule omówione zostaną przede wszystkim dwa pierwsze z wymienionych przed chwilą problemów. Kwestia oceny trafności proponowanej formy EWD dla nauczycieli jest kwestią bardzo złożoną. Dyskusja na ten temat zostanie tu jedynie rozpoczęta i wymagać będzie uzupełnienia o wyniki dalszych analiz i badań.

Sposób zdefiniowania i pomiaru EWD dla nauczyciela w Polsce

Ogólna logika wyznaczania wartości EWD dla nauczycieli jest bardzo podobna jak w przypadku EWD dla szkół. Biorąc pod uwagę wyniki uczniów na wejściu, przy pomocy modelu regresji określa się przewidywany wynik na wyjściu. Różnice pomiędzy rzeczywistymi wynikami, jakie osiągnęli uczniowie na egzaminie kończącym dany etap kształcenia a wartościami przewidywanymi, *uśrednia się*² w ramach szkół, a w przypadku nauczycieli w ramach uczniów, których dany nauczyciel nauczał i w ten sposób otrzymuje się wskaźniki EWD (p. Dolata, 2007). Przy szacowaniu EWD dla nauczycieli powstają jednak specyficzne problemy. W ramach polskiego systemu oświaty możemy wyróżnić pięć podstawowych:

1. Klasy są nauczane przez kilku nauczycieli.
2. Egzamin są interdyscyplinarne.
3. Liczba godzin dydaktycznych jest różna dla różnych przedmiotów.
4. Aby możliwe było obliczanie EWD dla nauczycieli, w danej szkole nie może zachodzić sytuacja, kiedy ten sam zespół nauczycieli naucza we wszystkich oddziałach.
5. Istnieje segregacja międzyszkolna i międzyoddziałowa.

Pierwsze dwa z przedstawionych problemów dają się podsumować w stwierdzeniu, że na podstawie dostępnych danych egzaminacyjnych trudno jest określić, jaki był wpływ konkretnego nauczyciela na wyniki uzyskane przez ucznia. Gdyby uczniów nauczał tylko jeden nauczyciel, można by przypisać mu całość zanotowanego wpływu. W sytuacji, gdy nauczycieli jest wielu, powstaje problem, kto ma być rozliczany z realizacji poszczególnych celów kształcenia.

Czy nauczycielom da się przyporządkować w sposób rozłączny cele kształcenia? Wydaje się, że o ile może mieć sens rozdzielne traktowanie części humanistycznej i matematyczno-przyrodniczej egzaminu gimnazjalnego, o tyle w ramach tych dwóch szerokich dziedzin bardzo trudno byłoby już dokonać rozłącznych podziałów. Choć formalnie rzecz biorąc, kwestia odpowiedzialności za nauczanie poszczególnych umiejętności pozostaje kwestią otwartą, to jednak nie da się ukryć, że są one rozwijane w ramach wielu przedmiotów, a więc i przez wielu nauczycieli. Na przykład czytanie ze zrozumieniem nie jest

² Określenie *uśrednia się* zostało tu użyte jedynie jako metafora. Jest ono literalnie prawdziwe tylko w odniesieniu do obliczania EWD dla szkół estymowanej jako efekty stałe. W przypadku EWD dla nauczycieli trzeba jeszcze uwzględnić problem nakładania się na tych samych uczniach wpływu kilku nauczycieli. Poza tym przy obliczaniu EWD jako efektów losowych wyniki są dodatkowo ściągane w kierunku zera, w stopniu proporcjonalnym do oszacowanej rzetelności pomiaru. Kwestie te omawiane są nieco szerzej w kolejnej części tekstu, poświęconej metodom estymacji.

domeną wyłącznie języka polskiego. Możemy sobie łatwo wyobrazić sytuację, w której ta umiejętność nabywana jest podczas lekcji historii, gdy uczniowie czytają i interpretują teksty źródłowe. Problem nakładania się nauczanych treści występuje jeszcze wyraźniej w przypadku przedmiotów matematyczno-przyrodniczych. Na przykład umiejętność odczytywania wykresów wydaje się elementem programu, który powinien być uwzględniony zarówno w kontekście nauki matematyki, fizyki, biologii, chemii, jak i geografii.

Widać więc, że nawet gdybyśmy dysponowali wynikami większej liczby bardziej wyspecjalizowanych dziedzinowo egzaminów, estymowanie EWD nauczyciela w oparciu o połączenie każdego egzaminu z dokładnie jednym nauczycielem (przedmiotem) byłoby problematyczne. Wyliczone EWD mogłoby bowiem nie być *zasługą* danego nauczyciela, lecz nauczycieli przedmiotów *pokrewnych*, rozwijających podobne umiejętności. Z drugiej strony można przypisać wielu nauczycieli do jednego egzaminu, co pozwala uwzględnić fakt, że wszyscy oni mają wpływ na jego wyniki. Pewnym niepożądanym skutkiem takiego działania jest to, że otrzymujemy kilka wskaźników EWD dla nauczycieli, przy czym możliwe jest, by jednemu nauczycielowi można było przypisać wartości EWD dla kilku takich wskaźników. W polskim kontekście problem ten jednak w zasadzie nie występuje.

Przede wszystkim nie mamy narzędzi do pomiaru większej liczby umiejętności (dziedzin). W ramach części humanistycznej i matematyczno--przyrodniczej egzaminu gimnazjalnego na jeden przedmiot przypada zbyt mała liczba pytań, by tworzyły one wystarczająco rzetelne (w kontekście pomiaru EWD dla takiej węższej dziedziny) narzędzie pomiarowe. Z kolei rozłączne przyporządkowanie nauczycieli do części humanistycznej lub matematyczno--przyrodniczej egzaminu gimnazjalnego nie powinno raczej budzić kontrowersji.

Jednocześnie należy zauważyć, że w ramach istniejącego systemu egzaminów zewnętrznych nie ma możliwości wyliczania EWD dla nauczycieli niektórych przedmiotów, jak muzyka, plastyka czy informatyka, gdyż egzamin gimnazjalny po prostu nie obejmuje treści tych przedmiotów. W przypadku języków obcych, wraz z wprowadzeniem części językowej do egzaminu gimnazjalnego, problem ten traci na aktualności, jednak wyliczanie EWD byłoby tu bardzo kontrowersyjne z innego powodu. Jest nim bardzo duże znaczenie pozaszkolnych zajęć językowych w rozwijaniu umiejętności uczniów w zakresie języków obcych (przy czym nie ma oczywiście możliwości, by uwzględnić ten czynnik w analizach).

Kolejny problem wiąże się z tym, że różni nauczyciele pracują z uczniami w różnym wymiarze godzin. Wydaje się przy tym rozsądne założyć, że im więcej czasu nauczyciel spędza z uczniem, tym silniejszy powinien być jego wpływ na wyniki ucznia, tym bardziej, że liczba godzin przeznaczana na poszczególne przedmioty znajduje również wyraźne odbicie w liczbie pytań odpowiadających danym dziedzinom wiedzy na egzaminie gimnazjalnym. W tabeli 1. przedstawiony został przykładowy ramowy plan nauczania dla gimnazjum. Możemy zobaczyć, że nauczyciel języka polskiego spędza z uczniem około 60% wszystkich godzin przeznaczonych na nauczanie przedmiotów humanistycznych. Dla porównania – dla nauczyciela historii jest to tylko około 30% godzin, a dla nauczyciela wiedzy o społeczeństwie zaledwie około 10%.

Wynikają z tego dwa wnioski. Po pierwsze, wyliczane wartości wskaźników EWD dla nauczycieli będą mogły być bezpośrednio porównywane tylko w odniesieniu do innych nauczycieli tego samego przedmiotu, względnie nauczycieli przedmiotów, których naucza się w zbliżonym wymiarze godzin. Po drugie, jeśli wpływ nauczycieli niektórych przedmiotów na wyniki egzaminu jest niewielki, to wiarygodne zrekonstruowanie tego wpływu w ramach miar EWD odwołujących się do wyników egzaminu może być bardzo trudne.

Oczywiście, aby możliwe było wyliczenie wskaźnika EWD, konieczne jest zachowanie ciągłości nauczania. Jeśli nauczyciel przejął nauczanie danego oddziału w trakcie cyklu kształcenia lub był przez dłuższe okresy zastępowany w wykonywaniu swoich obowiązków przez kogoś innego (np. z powodów zdrowotnych), to taki oddział nie może być uwzględniany przy wyliczaniu jego EWD. Analogicznie, z analiz wykluczeni muszą zostać uczniowie, którzy dołączyli do oddziału w późniejszych klasach.

Tabela 1. Przykładowy ramowy plan nauczania dla szkoły gimnazjalnej³

| | I klasa | II klasa | III klasa | | razem (32 tyg./rok) |
|-----------------------------|---------|----------|-----------|---------|------------------------|
| | | | I sem. | II sem. | |
| język polski | 5 | 5 | 4 | 4 | 448 |
| język obcy nowożytny | 3 | 2,5 | 2 | 1 | 224 |
| II język obcy nowożytny | 3 | 2,5 | 2 | 1 | 224 |
| muzyka | 0 | 1 | 0 | 0 | 32 |
| plastyka | 1 | 0 | 0 | 0 | 32 |
| historia | 2 | 2 | 2 | 2 | 192 |
| wiedza o społeczeństwie | 0 | 1 | 1 | 1 | 64 |
| geografia | 2 | 1 | 1 | 1 | 128 |
| biologia | 2 | 1 | 1 | 1 | 128 |
| chemia | 1 | 2 | 1 | 1 | 128 |
| fizyka | 1 | 2 | 1 | 1 | 128 |
| matematyka | 4 | 4 | 5 | 3 | 384 |
| informatyka | 1 | 0,5 | 1 | 1 | 80 |
| wychowanie fizyczne | 4 | 4 | 4 | 4 | 384 |
| edukacja dla bezpieczeństwa | 0 | 0 | 1 | 1 | 32 |
| zajęcia artystyczne | 0,5 | 0,5 | 1 | 1 | 64 |
| zajęcia techniczne | 0 | 0 | 2 | 2 | 64 |
| godziny z wychowawcą | 1 | 1 | 1 | 1 | 96 |
| religia/etyka | 2 | 2 | 2 | 2 | 192 |
| WDŻ | 0,5 | 0,5 | 1 | 0 | 48 |
| preorientacja zawodowa | 0 | 0,5 | 0 | 1 | 32 |
| tygodniowa liczba godzin | 33 | 33 | 33 | 29 | |

³ Tabela opracowana na podstawie dokumentu dostępnego na stronie MEN: http://www.reformaprogramowa.men.gov.pl/images/docs/2_ramowy_plan_nauczania.pdf (dostęp 10.08.2010).

Przejdźmy do problemu czwartego. Trzeba zauważyć, że obliczenie EWD dla nauczycieli w proponowanej formie nie zawsze będzie możliwe. Dla szkół, w których ten sam zespół nauczycieli naucza we wszystkich oddziałach (w szczególności dla szkół jednooddziałowych) nie ma po prostu możliwości rozdzielania od siebie wpływu poszczególnych nauczycieli. Wyliczone w tej sytuacji EWD odnosi się *de facto* do całego zespołu nauczycieli⁴ i nie różnicuje ich między sobą. W przypadku gimnazjów nie jest to jednak problem, który wykluczałby z analizy wiele szkół.

W tym kontekście warto zaznaczyć, że właśnie sposób *krzyżowania* się przydziału nauczycieli do oddziałów ma bardzo duże znaczenie dla możliwości wiarygodnej rekonstrukcji wpływu nauczycieli na wyniki uczniów. Ogólnie rzecz biorąc, najlepiej (z punktu widzenia modelu) aby nauczyciel uczył jak najwięcej klas, a w każdej z nich towarzyszyli mu inni nauczyciele pozostałych przedmiotów. Im mniejsza liczba *konfiguracji* zespołów nauczycielskiego wśród oddziałów, tym trudniej będzie zrekonstruować wpływ poszczególnych nauczycieli.

Wreszcie kwestia segregacji. Ten problem również ma charakter statystyczny. Ostatnie badania (Rothstein, 2009) wskazują na to, iż nielosowy przydział uczniów do szkół i klas może w sposób istotny obarczać błędem wyniki szacowania EWD. Należy przy tym zauważyć, że segregacja międzyszkolna na progu polskich gimnazjów jest znaczna (współczynnik segregacji ~0,14), przy niewielkiej segregacji międzyoddziałowej (gdy kontrolujemy przydział do szkół).

Metody estymacji EWD dla nauczycieli

Wyliczanie EWD dla nauczycieli, podobnie jak w przypadku EWD dla szkół, może być w warunkach polskiego systemu egzaminacyjnego oparte o modele regresji z efektami stałymi lub modele regresji z efektami losowymi. W tym drugim przypadku będziemy mieć do czynienia ze specyficzną podklasą, modelami o częściowo krzyżujących się efektach losowych. Warto dodać, że jest to szczególnie skomplikowana i złożona obliczeniowo klasa modeli.

Zacznijmy od modeli z efektami stałymi. Ich podstawową zaletą jest prostota. Niestety, w kontekście EWD dla nauczycieli konieczne jest zastosowanie nieco bardziej skomplikowanej procedury estymacji niż w przypadku EWD dla szkół. W EWD szkolnym wystarczyło wyznaczyć optymalną postać przyjętej funkcji zależności pomiędzy wynikiem na wejściu a wynikiem na wyjściu (czyli tak dobrać współczynniki regresji a i b , aby zminimalizować wariancję ϵ):

$$y_i = a + b x_i + \epsilon_i$$

⁴ Od strony technicznej można oczywiście uwzględnić takie szkoły w modelu razem z innymi, jednak oszacowane wartości EWD będą dla wszystkich nauczycieli w ramach takiej szkoły identyczne (będą odpowiadały ich *uśrednionemu* wpływowi) i nie będą mogły być interpretowane w ten sam sposób, co pozostałe wyniki.

gdzie:

- y_i – wynik i-tego ucznia na egzaminie gimnazjalnym (jednej części),
- x_i – wynik i-tego ucznia na sprawdzianie,
- a, b – współczynniki równania regresji,
- ε_i – reszta regresji i-tego ucznia.

Następnie EWD szkół można było obliczyć, po prostu sumując reszty regresji uczniów uczęszczających do danej szkoły:

$$EWD_j = 1/n \sum_{ji}$$

gdzie EWD_j jest edukacyjną wartością dodaną dla j -tej szkoły o liczbie uczniów równej n (Dolata, 2007). Operację taką można było przeprowadzić, gdyż przydział uczniów do szkół jest rozłączny. W przypadku nauczycieli warunek rozłączności nie jest jednak zachowany. Oznacza to konieczność włączenia bezpośrednio do modelu, już na etapie estymacji, zmiennych opisujących przypisanie uczniów do szkół/oddziałów i nauczycieli⁵:

$$y_i = a + b x_i + \sum (B^O O_i) + \sum (B^N N_i) + \varepsilon_i$$

gdzie:

- B^O – wektor wag oddziałów (współczynniki równania regresji),
- N_i – wektor opisujący przypisanie i -tego ucznia do oddziału,
- B^N – wektor wag nauczycieli (współczynniki równania regresji),
- N_i – wektor opisujący przypisanie nauczycieli do i -tego ucznia.

W tym przypadku EWD nauczyciela oblicza się jako:

Niestety, ze względu na dużą liczbę parametrów modele te mogą powodować duże problemy w estymacji (przy zastosowaniu typowych metod używanych do estymacji modeli liniowych z efektami stałymi). O ile estymacja modelu w ramach jednej szkoły lub kilku szkół jednocześnie powinna być zwykle możliwa, o tyle przy większej ilości danych mogą występować problemy ze stabilnością oszacowań lub niemożliwość oszacowania wszystkich parametrów modelu.

W praktyce oznacza to, że modele z efektami stałymi mogą znaleźć zastosowanie jedynie w obliczeniach prowadzonych na niewielką skalę, *lokalnie*, np. przez dyrektorów szkół. Przy takim wykorzystaniu trzeba będzie jednak posłużyć się w estymacji postacią zależności pomiędzy wynikiem na wejściu a wynikiem na wyjściu wyestymowaną w ramach szerszego modelu (niewielka liczba obserwacji nie daje dobrych podstaw do estymacji tej zależności). W szczególności wskazane byłoby użycie tej postaci zależności, która została wyliczona w danym roku w ramach modelu jednorocznego EWD dla szkół.

⁵ W przypadku EWD dla szkół również można by włączyć do modelu informacje o przypisaniu uczniów do szkół już na etapie estymacji, co (ze względu na rozłączność tego podziału) doprowadziłoby do tych samych wyników, co obliczanie średniej z reszt regresji.

Model EWD dla nauczycieli przybiera w takich okolicznościach postać:

$$y_i = f(x_i) + \Sigma (B^O O_i) + \Sigma (B^N N_i) + \varepsilon_i$$

gdzie $f(x_i)$ oznacza postać zależności wyestymowaną w ramach modelu EWD jednorocznego⁶. Model ten trzeba estymować jako model bez stałej regresji (jest ona *zaszyta* wewnątrz $f(x_i)$) i nie powinna być wyliczana *po raz drugi*).

Warto jeszcze zaznaczyć, że w odniesieniu do zastosowań na niewielką skalę, modele z efektami stałymi powinny dawać wyniki bardzo zbliżone do tych, jakie uzyskałoby się, stosując dużo bardziej zaawansowane modele z częściowo krzyżującymi się efektami losowymi. W sytuacji, gdy mamy do czynienia z segregacją, modele z efektami stałymi będą jednak gorzej oszacowywały wartości błędów standardowych EWD.

Przy estymowaniu EWD dla nauczycieli na dużą skalę jedynym sprawdzającym się rozwiązaniem jest zastosowanie modeli wielopoziomowych (zwanymi też modelami mieszanymi efektów), w których EWD nauczycieli traktowane są jako czynniki losowe, a nie stałe. Po pierwsze, ich założenia lepiej pasują do modelowanego zjawiska. Po drugie, w odróżnieniu od typowych metod estymacji modeli z efektami stałymi, są one w stanie zapewnić stabilność (i w ogóle możliwość uzyskania) wyników przy modelowaniu na dużych zbiorach danych.

Ponieważ przydział nauczycieli do uczniów nie jest rozłączny, o modelach EWD dla nauczycieli możemy mówić jako o modelach częściowo krzyżujących się efektów losowych. Jest to szczególna klasa modeli wielopoziomowych, bardzo złożona w estymacji (p. Skrondal i Rabe-Hesketh, 2004; Pinheiro i Bates, 2000). Niestety możliwość praktycznego zastosowania tego typu modeli do wyliczania EWD dla nauczycieli w Polsce ogranicza fakt, że do ich estymacji niezbędne jest specjalistyczne oprogramowanie statystyczne. Właściwie wyklucza to tego typu modele z użycia w ramach analiz prowadzonych na małą skalę, w szczególności wewnątrzszkolnych (jak już jednak wspomniano, w tych ostatnich wystarczające będą raczej modele z efektami stałymi).

W ramach modelu z częściowo krzyżującymi się efektami losowymi wynik ucznia na wyjściu opisany będzie jako:

$$y_{ji} = \gamma_{00} + \gamma_{01} x_{ji} + u_j^O + \Sigma u_j^N + \varepsilon_{ji}$$

gdzie:

y_{ji} – wynik i -tego ucznia z j -tego oddziału na egzaminie gimnazjalnym (jednej części),

x_{ji} – wynik i -tego ucznia z j -tego oddziału na sprawdzianie,

γ_{00}, γ_{01} – efekty stałe (opisujące zależność wyniku egz. gim. od wyniku sprawdzianu – taką samą dla całej badanej populacji),

u_j^O – efekt losowy j -tego oddziału,

Σu_j^N – suma efektów losowych nauczycieli uczących j -ty oddział,

ε_{ji} – reszta regresji i -tego ucznia z j -tego oddziału.

⁶ Model właśnie w tej postaci estymuje *Kalkulator EWD*, przy czym zamiast informacji o przydziale uczniów do oddziałów i nauczycieli znajdują się tam zmienne wybrane przez użytkownika.

Włączanie do modelu efektów losowych dla oddziałów – możliwe jest też dodanie efektów dla szkół – ma na celu wybicie z efektów nauczycielskich wpływu segregacji międzyszkolnej i międzyoddziałowej (uczniów). Trzeba jednak pamiętać, że w praktyce zawsze występować będzie pewna arbitralność w dekompozycji wpływu pomiędzy nauczycieli a oddziały (szkoły). W szczególności, jeśli występuje również segregacja nauczycieli ze względu na ich umiejętności, to część wpływu nauczycieli będzie przypisywana właśnie wpływowi oddziałów (szkół). Problem ten dotyczy zresztą w jeszcze większym stopniu modeli, które efekty nauczycieli i oddziałów modelują jako efekty stałe (McCaffrey i.in, 2003).

Metody oceny trafności EWD dla nauczycieli w Polsce

Kwestia oceny trafności proponowanej metody obliczania EWD dla nauczycieli w Polsce jest dosyć skomplikowana. Składają się na to dwa problemy. Pierwszy z nich to konieczność przeprowadzenia badań empirycznych, które będą wymagały zebrania informacji o przydziale nauczycieli do oddziałów, w tej chwili nie ma bowiem dostępu do tego rodzaju danych. Badań takich jak dotąd nie przeprowadzono, choć zespół edukacyjnej wartości dodanej zamierza się podjąć ich realizacji w niedalekiej przyszłości. Dopiero na podstawie ich wyników możliwe będzie ostateczne rozstrzygnięcie, na ile dobrze model EWD dla nauczycieli sprawdza się w rzeczywistości.

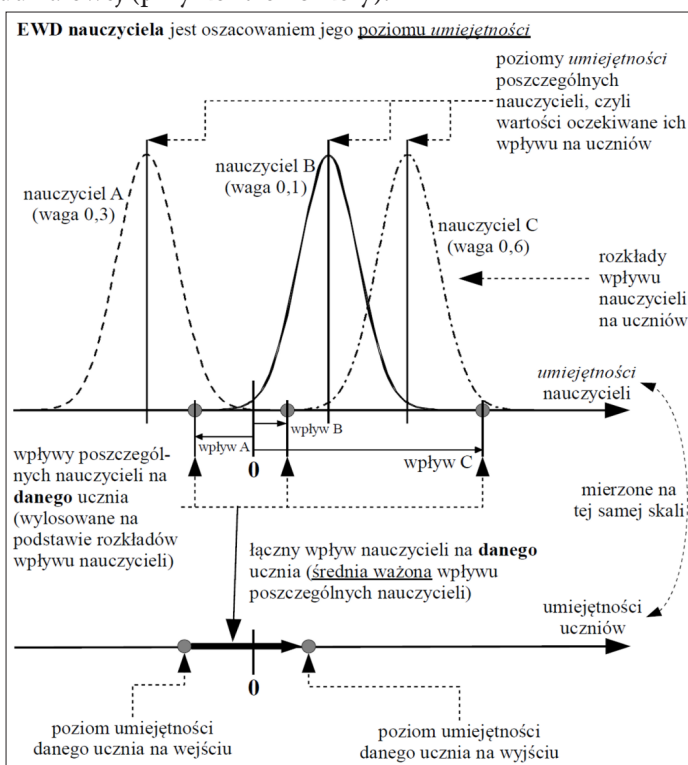
W tej chwili można jedynie przeprowadzić analizy symulacyjne. Choć poprawność zachowania modelu w ramach symulacji nie gwarantuje, że sprawdzi się on również na danych rzeczywistych (symulacje zawsze są uproszczonym modelem rzeczywistości), to jednak mogą one już na wstępnym etapie obnażyć niektóre słabości modelu. W ramach symulacji mamy też możliwość lepszego sprawdzenia, jak dokładność oszacowań zależy od wartości poszczególnych parametrów modelu.

Niemniej również przy badaniach symulacyjnych napotykamy w polskim kontekście poważne utrudnienia. Symulacje powinny w możliwie dokładny sposób odwzorowywać sytuację rzeczywistego badania (wyliczenia EWD). Rzecz w tym, że modelowanie EWD nauczycielskiego w najbliższej przyszłości (jeśli do tego dojdzie) będzie jako całość procesem bardzo złożonym. Nie będzie się ono bowiem odbywać *globalnie*, na poziomie ogólnokrajowym, lecz *lokalnie*, najprawdopodobniej na poziomie pojedynczej szkoły, być może kilku szkół. W takim *lokalnym* modelowaniu będą jednak musiały zostać wykorzystane informacje z modelu wyestymowanego *globalnie*, czyli postaci zależności pomiędzy wynikami sprawdzianu a wynikami egzaminu gimnazjalnego. Symulacja mająca dobrze oddawać specyfikę opisywanej sytuacji musi więc uwzględniać w sobie tę dwustopniowość: 1) estymacja *globalna* EWD dla szkół, 2) estymacja *lokalna* EWD dla nauczycieli przy uwzględnieniu informacji z 1). Jakość odwzorowania założonego wpływu nauczycieli należy oczywiście sprawdzać w odniesieniu do wyników uzyskanych w drugim kroku. W oczywisty sposób komplikuje to schemat symulacji, utrudniając jej przygotowanie i przeprowadzenie.

Symulacje globalnego modelu EWD dla nauczycieli

Mimo wszystko wydaje się, że badania nad własnościami proponowanego modelu EWD dla nauczycieli najlepiej zacząć od kwestii relatywnie prostych. W związku z tym przeprowadzony został szereg symulacji odpowiadających sytuacji, gdyby EWD dla nauczycieli estymowane było *globalnie*, w skali całego kraju (podobnie jak EWD dla szkół).

W ramach modelu symulacji przygotowywano losowe zbiory danych, za każdym razem losując parametry dla 1000 szkół, z których każda składała się z 2–6 oddziałów (w zależności od wariantu symulacji; w ramach jednego wariantu wszystkie szkoły miały tę samą liczbę oddziałów), po 20 uczniów każdy. Umiejętności uczniów losowano z rozkładu normalnego o wartości oczekiwanej zero i odchyleniu standardowym jeden, przy czym przypisywano ich do oddziałów i szkół w ten sposób, by zachowane zostały przyjęte (jako parametr symulacji) poziomy segregacji międzyszkolnej i międzyoddziałowej. Wartości współczynników segregacji przyjęto w ramach przeprowadzonych symulacji jako stałe i ustalono na poziomie zbliżonym do rzeczywistych wartości dla polskich gimnazjów: 0,14 dla segregacji międzyszkolnej i 0,1 dla segregacji międzyoddziałowej (przy kontroli szkoły).



Rys. 1. Schemat ilustrujący sposób wyliczania łącznego wpływu nauczycieli na umiejętności uczniów

Założono, że uczniowie są nauczani przez trzech nauczycieli, przy czym stopień ich oddziaływania na ucznia jest proporcjonalny do liczby nauczanych godzin. Pierwszy nauczyciel uczący dany oddział otrzymał przy tym wagę 0,6, drugi 0,3, a trzeci 0,1. Odzwierciedla to w przybliżeniu sytuację, z jaką mamy do czynienia w odniesieniu do nauczania przedmiotów humanistycznych w gimnazjum. Poziom *umiejętności* nauczyciela losowano z rozkładu normalnego o wartości oczekiwanej zero. Nauczyciele byli losowo przydzielani do oddziałów (oddzielnie w ramach każdej z trzech kategorii-wag nauczycieli). Założono, że nie występuje segregacja międzyszkolna nauczycieli ze względu na ich poziom *umiejętności*. Liczba oddziałów, które nauczał każdy nauczyciel ustalana była parametrami modelu (trzema – po jednym dla każdego rodzaju nauczycieli) i różniła się w zależności od wariantu symulacji. Jeśli liczba oddziałów w szkole nie była wielokrotnością wartości tego parametru, to ostatniemu nauczycielowi (danego rodzaju) przydzielanemu do każdej szkoły przypisywana była mniejsza liczba oddziałów.

Wpływ nauczyciela na uczniów zdefiniowano jako zmienną losową o rozkładzie normalnym o wartości oczekiwanej równej poziomowi *umiejętności* danego nauczyciela. Uwzględniano w ten sposób fakt, że ten sam nauczyciel może mieć różny wpływ na wyniki, w zależności od ucznia. Dla każdego ucznia, którego uczył dany nauczyciel, losowano wartość z takiego rozkładu, mnożono przez wagę danego nauczyciela i dodawano do poziomu *umiejętności* ucznia. Po wykonaniu tej operacji dla wszystkich nauczycieli uzyskiwano w wyniku poziom *umiejętności* uczniów na wyjściu.

W symulacji uwzględniony został również fakt, że *umiejętności* uczniów nie są dostępne bezpośrednio, lecz za pośrednictwem egzaminów, które nie są doskonale rzetelne. Proces egzaminowania symulowano pod założeniami modelu Rasha (p. Jakubowski, Pokropek, 2009) przy wykorzystaniu dwóch testów (jeden test na wejściu, drugi na wyjściu) składających się z 50 pytań każdy. Trudność pytań losowano z rozkładu normalnego o parametrach odpowiadających rozkładowi *umiejętności* uczniów odpowiednio na wejściu i na wyjściu. W efekcie uzyskiwano symulowane wyniki egzaminacyjne uczniów.

Sumę wariancji rozkładu *umiejętności* nauczycieli i wariancji wpływu nauczycieli na uczniów dobrano w ten sposób, aby korelacja pomiędzy wynikiem egzaminu na wejściu i wynikiem egzaminu na wyjściu była zbliżona do tej, która jest notowana w rzeczywistości pomiędzy wynikiem sprawdzianu a wynikiem egzaminu gimnazjalnego⁷ (wartość współczynnika korelacji Pearsona $\sim 0,720$). Kwestią otwartą pozostaje, jaka powinna być proporcja pomiędzy wariancją *umiejętności* nauczycieli a wariancją wpływu nauczycieli na uczniów. W odróżnieniu od innych parametrów modelu nie mamy tu możliwości dokonania oszacowania w oparciu o dane empiryczne, choć wydaje się, że wariancja *umiejętności* nauczycieli powinna być większa (a nawet wyraźnie większa). Jest to o tyle istotny problem, że od wartości tej proporcji może również zależeć dokładność oszacowań EWD z modelu – wydaje się, że im większe zróżnicowanie nauczycieli między sobą, a mniejsze zróżnicowanie ich wpływu na

⁷ O wartości tej korelacji decyduje właśnie suma ww. wariancji. Dla wartości korelacji nie jest za to istotne, w jakiej proporcji składają się na tę sumę wariancja *umiejętności* nauczycieli i wariancja wpływu nauczycieli na uczniów.

uczniów, tym dokładniejsze powinny być oszacowania. W związku z tym proporcja pomiędzy ww. wariancjami włączona została do symulacji jako parametr, którego wpływ testowano.

EWD dla nauczycieli wyliczane było w ramach modelu wielopoziomowego z losowymi efektami dla oddziałów i nauczycieli oraz efektem stałym dla wyniku egzaminu na wejściu (jako zmiennej zależnej używano wyniku egzaminu na wyjściu). Dla porównania estymowano też analogiczny model w sytuacji pomiaru umiejętności *doskonale rzetelnym* testem – w tym przypadku zmienną zależną był poziom umiejętności uczniów na wyjściu, a efekt stały związany był z poziomem umiejętności uczniów na wejściu.

W ramach symulacji badano wpływ na dokładność oszacowania EWD nauczycieli⁸ dwóch czynników, wspomnianej przed chwilą proporcji wariancji *umiejętności* nauczycieli do wariancji wpływu nauczycieli na uczniów oraz stopnia *krzyżowania się* przydziału nauczycieli do oddziałów.

W przypadku proporcji wariancji analizę przeprowadzono dla czterech wartości:

- 1, 2, 5, 10 (tj. wariancja *umiejętności* nauczycieli równa wariancji wpływu nauczycieli na uczniów oraz dwu-, pięcio- i dziesięciokrotnie większa od niej).

W przypadku *krzyżowania się* przydziału nauczycieli do oddziałów brak jest syntetycznej miary pozwalającej określić jego stopień. Za to, z jak korzystną/niekorzystną sytuacją mamy tu do czynienia, decyduje w badanym przypadku kombinacja aż czterech parametrów: liczby oddziałów w szkole oraz liczby oddziałów nauczanych przez nauczycieli każdego rodzaju (3 parametry). Do analizy wybrano cztery kombinacje ww. parametrów (w kolejności od sytuacji najbardziej do najmniej korzystnej):

- 6 oddziałów w szkole, każdy nauczyciel naucza 3 oddziały;
- 4 oddziały w szkole, nauczyciel z wagą 0,6 naucza 2 oddziały, nauczyciel z wagą 0,3 naucza 3 oddziały, a nauczyciel z wagą 0,1 naucza wszystkie 4 oddziały;
- 4 oddziały w szkole, nauczyciel z wagą 0,6 naucza 1 oddział, pozostali nauczyciele 3⁹;
- 2 oddziały w szkole, nauczyciel z wagą 0,6 naucza 1 oddział, pozostali nauczyciele 2.

Na rys. 2. w sposób graficzny zaprezentowane zostały możliwe przypadki (a właściwie klasy przypadków) przydziału nauczycieli do klas w ramach przyjętych wariantów symulacji. Zaczernienie komórki wskazuje na przydział nauczyciela do danego oddziału. Przy poszczególnych przypadkach podano prawdopodobieństwa ich wystąpienia, gdy przydział nauczycieli do oddziałów następuje losowo (jak w omawianej symulacji).

⁸ Pamiętajmy, że odnosi się to do wykorzystania modelu EWD dla nauczycieli estymowanego *globalnie*. Wyniki te nie mogą być bezpośrednio odnoszone do sytuacji, gdy wyliczanie EWD dla nauczycieli będzie się odbywać *lokalnie* z wykorzystaniem informacji o postaci zależności pomiędzy wynikiem na wejściu a wynikiem na wyjściu.

⁹ Oczywiście w praktyce oznacza to, że w każdej szkole jest po dwóch nauczycieli o wagach 0,3 i 0,1, z których jeden naucza trzy oddziały, a drugi tylko jeden.

| Wariant 6; 3-3-3: najbardziej korzystny przypadek krzyżowania (P=339/400) | | | | | | | | | | |
|---|---|---|---|---|--|---|---|---|---|--|
| | A | B | C | D | E | F | | | | |
| n0,6_1 | | | | | | | | | | |
| n0,6_2 | | | | | | | | | | |
| n0,3_1 | | | | | | | | | | |
| n0,3_2 | | | | | | | | | | |
| n0,1_1 | | | | | | | | | | |
| n0,1_2 | | | | | | | | | | |
| Wariant 6; 3-3-3: mniej korzystny przypadek krzyżowania (P=3/20) | | | | | | | | | | |
| | A | B | C | D | E | F | Przypadek ten jest mniej korzystny, gdyż przydział do oddziałów nauczycieli dwóch rodzajów (tu o wagach 0,3 i 0,1) pokrywa się. | | | |
| n0,6_1 | | | | | | | | | | |
| n0,6_2 | | | | | | | | | | |
| n0,3_1 | | | | | | | | | | |
| n0,3_2 | | | | | | | | | | |
| n0,1_1 | | | | | | | | | | |
| n0,1_2 | | | | | | | | | | |
| Wariant 4; 2-3-4: dość korzystne krzyżowanie (P=1) | | | | | | | | | | |
| | A | B | C | D | | | | | | |
| n0,6_1 | | | | | | | | | | |
| n0,6_2 | | | | | | | | | | |
| n0,3_1 | | | | | | | | | | |
| n0,3_2 | | | | | | | | | | |
| n0,1 | | | | | | | | | | |
| Wariant 4; 1-3-3: bardziej korzystny przypadek krzyżowania (P=3/4) | | | | | Wariant 4; 1-3-3: niekorzystny przypadek krzyżowania (P=1/4) | | | | | |
| | A | B | C | D | | A | B | C | D | |
| n0,6_1 | | | | | n0,6_1 | | | | | |
| n0,6_2 | | | | | n0,6_2 | | | | | |
| n0,6_3 | | | | | n0,6_3 | | | | | |
| n0,6_4 | | | | | n0,6_4 | | | | | |
| n0,3_1 | | | | | n0,3_1 | | | | | |
| n0,3_2 | | | | | n0,3_2 | | | | | |
| n0,1_1 | | | | | n0,1_1 | | | | | |
| n0,1_2 | | | | | n0,1_2 | | | | | |
| Wariant 2; 1-2-2: bardzo niekorzystne krzyżowanie (P=1) | | | | | | | | | | |
| | A | B | | | | | | | | |
| n0,6_1 | | | | | | | | | | |
| n0,6_2 | | | | | | | | | | |
| n0,3 | | | | | | | | | | |
| n0,1 | | | | | | | | | | |

Rys. 2. Schematy ilustrujące krzyżowanie się przydziału nauczycieli do oddziałów dla różnych wariantów symulacji

Należy wspomnieć, że dla wariantu z 6 klasami możemy jeszcze mieć do czynienia z przypadkiem (niepokazanym na rysunku), kiedy nauczyciele dzielą się na dwa rozłączne zespoły, tak że nie ma żadnego *krzyżowania się*. Jest to oczywiście sytuacja skrajnie niekorzystna¹⁰, jednak jej wystąpienie jest mało prawdopodobne ($P=1/400$).

Oba czynniki skrzyżowano ze sobą, uzyskując łącznie 16 kombinacji. Dla każdej z takich kombinacji symulacja przeprowadzona została 30 razy, a miary dopasowania uśrednione, w celu usunięcia efektów związanych z nietypowym rozlosowaniem się trudności pytań w testach. Łącznie wykonano 480 przebiegów symulacji.

Uzyskane w wyniku modelowania wartości EWD nauczycieli były porównywane z rzeczywistym (wylosowanym w procesie przygotowywania danych) poziomem ich *umiejętności*. Wykorzystano tu dwie miary:

- współczynnik korelacji liniowej Pearsona pokazujący siłę związku liniowego,
- współczynnik korelacji rangowej t_b Kendalla pokazujący, w jakim stopniu zgodne ze sobą są uporządkowania nauczycieli ze względu na prawdziwy poziom *umiejętności* i ze względu na wyestymowane EWD.

Dodatkowo analizie poddano szerokość estymowanych przedziałów ufności dla EWD nauczycieli.

Wyniki symulacji zestawione zostały w tabelach 2–4. Wyraźnie różnicują się one ze względu na wagę nauczyciela. Z jednej strony należy zauważyć, że dla nauczycieli z wagą 0,6 udaje się całkiem trafnie zrekonstruować ich poziom *umiejętności*. Rozpatrując zależność w kategoriach związku liniowego, osiągamy bardzo dobre rezultaty, w zależności od wariantu wartość współczynnika korelacji Pearsona waha się od 0,857 do aż 0,929 (dla testów z 50 pytaniami). Siła zależności zmienia się przy tym w bardzo niewielkim stopniu dla różnych wartości parametrów – w szczególności dla proporcji wariancji, choć również wpływ *krzyżowania się* przydziału nauczycieli do oddziałów jest raczej niewielki. Także rzetelność testu ma niewielkie znaczenie dla siły notowanej zależności.

¹⁰ Formalnie, jak było wspomniane we wcześniejszej części tekstu, estymujemy wtedy EWD całego zespołu nauczycieli (rozdzielone po równo pomiędzy nich). Wartość ta nie powinna być więc interpretowana jako *pełnoprawne* indywidualne EWD nauczyciela (choć jeśli występuje nauczyciel o dominującym wpływie - jak tu *nauczyciel 0,6* - to wartość takiego EWD będzie mieć wiele wspólnego właśnie z jego poziomem *umiejętności*), jednak aby nie komplikować dodatkowo procedury symulacji, nie eliminowano tego typu sytuacji.

Tabela 2. Współczynniki korelacji Pearsona pomiędzy EWD a rzeczywistym poziomem umiejętności nauczycieli

| lp. | parametry | | | naucz. 0,6 | | naucz. 0,3 | | naucz. 0,1 | |
|-----|------------------|------------------|----------------------|------------|---------|------------|---------|------------|---------|
| | l. oddz./ szkoła | l. oddz./ naucz. | prop. D ² | testy 50 | testy ∞ | testy 50 | testy ∞ | testy 50 | testy ∞ |
| 1 | 6 | 3-3-3 | 10 | 0,929 | 0,937 | 0,722 | 0,739 | 0,516 | 0,621 |
| 2 | 6 | 3-3-3 | 5 | 0,927 | 0,936 | 0,716 | 0,735 | 0,494 | 0,594 |
| 3 | 6 | 3-3-3 | 2 | 0,923 | 0,933 | 0,708 | 0,729 | 0,439 | 0,531 |
| 4 | 6 | 3-3-3 | 1 | 0,917 | 0,929 | 0,693 | 0,717 | 0,395 | 0,467 |
| 5 | 4 | 2-3-4 | 10 | 0,923 | 0,936 | 0,723 | 0,763 | 0,202 | 0,206 |
| 6 | 4 | 2-3-4 | 5 | 0,921 | 0,934 | 0,713 | 0,755 | 0,202 | 0,206 |
| 7 | 4 | 2-3-4 | 2 | 0,915 | 0,930 | 0,688 | 0,734 | 0,207 | 0,211 |
| 8 | 4 | 2-3-4 | 1 | 0,903 | 0,921 | 0,655 | 0,703 | 0,197 | 0,202 |
| 9 | 4 | 1-3-3 | 10 | 0,892 | 0,907 | 0,545 | 0,554 | 0,191 | 0,198 |
| 10 | 4 | 1-3-3 | 5 | 0,889 | 0,904 | 0,537 | 0,548 | 0,194 | 0,197 |
| 11 | 4 | 1-3-3 | 2 | 0,878 | 0,897 | 0,537 | 0,548 | 0,187 | 0,192 |
| 12 | 4 | 1-3-3 | 1 | 0,861 | 0,885 | 0,531 | 0,543 | 0,192 | 0,193 |
| 13 | 2 | 1-2-2 | 10 | 0,885 | 0,903 | 0,556 | 0,564 | 0,181 | 0,180 |
| 14 | 2 | 1-2-2 | 5 | 0,885 | 0,901 | 0,557 | 0,567 | 0,187 | 0,191 |
| 15 | 2 | 1-2-2 | 2 | 0,875 | 0,893 | 0,553 | 0,563 | 0,180 | 0,186 |
| 16 | 2 | 1-2-2 | 1 | 0,857 | 0,881 | 0,542 | 0,556 | 0,175 | 0,181 |

Zdecydowanie gorzej wygląda sytuacja w przypadku nauczycieli z wagą 0,3. Tu wartość współczynnika korelacji Pearsona waha się od 0,542 do 0,722, a zróżnicowanie wartości parametrów modelu okazuje się mieć zdecydowanie większy wpływ na siłę notowanej zależności. Szczególnie chodzi tu o natężenie *krzyżowania*. W przypadku nauczycieli z wagą 0,1 sprawa wygląda już zdecydowanie źle. Przy tak niewielkim wpływie nauczyciela na wyniki ucznia trafna rekonstrukcja poziomu *umiejętności* nauczyciela jest już właściwie niemożliwa. Jednocześnie warto zauważyć, jak duża jest w tym wypadku rola czynnika *krzyżowania* się przydziału nauczycieli do oddziałów dla trafności EWD.

Jeśli od rozpatrywania zależności liniowych przejdziemy do sprawdzania, jak dobrze przy pomocy EWD można zrekonstruować uporządkowanie nauczycieli ze względu na poziom ich umiejętności, to możemy zauważyć spadek trafności naszych przewidywań. Choć dla nauczycieli o wadze 0,6 wartości korelacji rangowej τ_b są jeszcze dosyć wysokie, o tyle dla pozostałych nauczycieli uporządkowanie ze względu na EWD w dużej mierze nie pokrywa się z rzeczywistością. Można przy tym zauważyć zdecydowanie większą wrażliwość korelacji rangowej (w stosunku do korelacji Pearsona) na zmiany parametrów symulacji, zwłaszcza porządku wariacji.

Tabela 3. Współczynniki korelacji rangowej τ_b Kendalla pomiędzy EWD a rzeczywistym poziomem umiejętności nauczycieli

| lp. | parametry | | | naucz. 0,6 | | naucz. 0,3 | | naucz. 0,1 | |
|-----|---------------------|---------------------|-------------------------|-------------|----------------|-------------|----------------|-------------|----------------|
| | l. oddz./ szkoła | l. oddz./ naucz. | prop. D ² | testy 50 | testy ∞ | testy 50 | testy ∞ | testy 50 | testy ∞ |
| 1 | 6 | 3-3-3 | 10 | 0,760 | 0,774 | 0,513 | 0,530 | 0,341 | 0,423 |
| 2 | 6 | 3-3-3 | 5 | 0,757 | 0,771 | 0,509 | 0,527 | 0,327 | 0,402 |
| 3 | 6 | 3-3-3 | 2 | 0,749 | 0,765 | 0,501 | 0,520 | 0,286 | 0,352 |
| 4 | 6 | 3-3-3 | 1 | 0,739 | 0,759 | 0,486 | 0,507 | 0,256 | 0,307 |
| 5 | 4 | 2-3-4 | 10 | 0,750 | 0,772 | 0,516 | 0,553 | 0,128 | 0,130 |
| 6 | 4 | 2-3-4 | 5 | 0,746 | 0,768 | 0,506 | 0,545 | 0,130 | 0,133 |
| 7 | 4 | 2-3-4 | 2 | 0,736 | 0,760 | 0,483 | 0,525 | 0,132 | 0,135 |
| 8 | 4 | 2-3-4 | 1 | 0,717 | 0,745 | 0,455 | 0,496 | 0,125 | 0,129 |
| 9 | 4 | 1-3-3 | 10 | 0,703 | 0,723 | 0,365 | 0,371 | 0,121 | 0,125 |
| 10 | 4 | 1-3-3 | 5 | 0,698 | 0,719 | 0,360 | 0,368 | 0,123 | 0,124 |
| 11 | 4 | 1-3-3 | 2 | 0,683 | 0,709 | 0,360 | 0,368 | 0,119 | 0,122 |
| 12 | 4 | 1-3-3 | 1 | 0,661 | 0,693 | 0,354 | 0,363 | 0,121 | 0,122 |
| 13 | 2 | 1-2-2 | 10 | 0,694 | 0,717 | 0,376 | 0,382 | 0,115 | 0,113 |
| 14 | 2 | 1-2-2 | 5 | 0,693 | 0,715 | 0,376 | 0,384 | 0,120 | 0,123 |
| 15 | 2 | 1-2-2 | 2 | 0,678 | 0,703 | 0,373 | 0,381 | 0,115 | 0,119 |
| 16 | 2 | 1-2-2 | 1 | 0,655 | 0,685 | 0,363 | 0,374 | 0,113 | 0,117 |

Rozważania dotyczące trafności można by podsumować stwierdzeniem, że rokując one dosyć dobrze dla możliwości szacowania EWD dla nauczycieli najważniejszych przedmiotów: polskiego i matematyki, choć w odniesieniu do pozostałych nauczycieli obliczanie EWD może być dosyć ryzykowne. Możliwość szacowania EWD dla nauczycieli przedmiotów, takich jak: historia, biologia, chemia, fizyka czy geografia wydaje się być silnie uwarunkowana wielkością szkoły i istniejącym w niej sposobem przydzielania nauczycieli do oddziałów. Szacowanie EWD dla nauczycieli pozostałych przedmiotów należałoby uznać za właściwie niemożliwe.

O ile dobrze jest wiedzieć, że generowane przez model predykcje są co do zasady trafne (przynajmniej w warunkach symulacji), o tyle trzeba zdać sobie sprawę, że w realnej sytuacji przy określaniu wartości EWD będziemy musieli bazować na procedurach wnioskowania statystycznego. W tym kontekście warto sprawdzić, jaka jest dokładność szacowania wyliczana w ramach modelu. W tabeli 4. zestawione zostały średnie odchylenia standardowe EWD (wskazujące na dokładność oszacowania EWD dla nauczycieli) w ramach modelu oraz odsetki nauczycieli, których EWD można uznać za istotnie statystycznie różne od zera przy estymacji przedziałowej na poziomie ufności $\gamma=0,95$.

Tabela 4. Porównanie dokładności oszacowania EWD i odsetka nauczycieli, których EWD istotnie statystycznie różniło się od zera (estymacja przedziałowa na poziomie ufności $\gamma=0,95$)

| lp. | nauczyciel 0,6 | | nauczyciel 0,3 | | nauczyciel 0,1 | |
|-----|----------------|-------------------------|----------------|-------------------------|----------------|-------------------------|
| | $E(os_{EWD})$ | EWD ist. stat. $\neq 0$ | $E(os_{EWD})$ | EWD ist. stat. $\neq 0$ | $E(os_{EWD})$ | EWD ist. stat. $\neq 0$ |
| 1 | 1,953 | 44,9% | 1,793 | 7,2% | 0,749 | 0,3% |
| 2 | 1,860 | 44,1% | 1,699 | 7,0% | 0,719 | 0,2% |
| 3 | 1,708 | 41,7% | 1,557 | 6,3% | 0,646 | 0,0% |
| 4 | 1,546 | 39,5% | 1,389 | 4,9% | 0,588 | 0,0% |
| 5 | 1,950 | 44,6% | 1,783 | 6,1% | 0,536 | 0,0% |
| 6 | 1,871 | 43,9% | 1,726 | 5,5% | 0,453 | 0,0% |
| 7 | 1,811 | 38,9% | 1,606 | 3,7% | 0,681 | 0,0% |
| 8 | 1,663 | 34,9% | 1,452 | 2,5% | 0,538 | 0,0% |
| 9 | 2,854 | 1,2% | 2,165 | 0,8% | 0,778 | 0,0% |
| 10 | 2,636 | 2,5% | 2,058 | 0,7% | 0,532 | 0,0% |
| 11 | 2,445 | 1,1% | 1,840 | 0,6% | 0,582 | 0,0% |
| 12 | 2,156 | 0,9% | 1,633 | 0,8% | 0,521 | 0,0% |
| 13 | 2,748 | 3,5% | 1,073 | 0,0% | 1,332 | 0,0% |
| 14 | 2,427 | 5,2% | 1,079 | 0,0% | 1,104 | 0,2% |
| 15 | 2,484 | 1,8% | 1,384 | 0,0% | 0,906 | 0,0% |
| 16 | 2,160 | 0,9% | 1,123 | 0,0% | 1,125 | 0,0% |

Uzyskane wyniki są bardzo jednoznaczne. Po pierwsze, wyjąwszy nauczycieli o wadze 0,6 bardzo trudno jest o istotne statystycznie różnice. W szczególności nauczyciele o wadze 0,1 są z punktu widzenia wnioskowania statystycznego właściwie nierozróżnialni od siebie nawzajem. Po drugie, nawet w przypadku nauczycieli przedmiotów wiodących rozróżnialność jest możliwa do uzyskania jedynie w warunkach dobrego *krzyżowania się* przydziału nauczycieli do oddziałów. Czynnikiem ten ma nawet większe znaczenie niż liczba nauczanych uczniów (oddziałów). Warto przy tym zauważyć, że w sprzyjających warunkach EWD od 35% do 45% nauczycieli o wadze 0,6 mogło zostać sklasyfikowane jako istotnie większe, lub istotnie mniejsze od zera. Dla nauczycieli z wagą 0,3 było to już niestety tylko od 2,5% do 7%.

Podsumowanie

W niniejszym artykule zaproponowany został sposób wyliczania EWD dla nauczycieli w Polsce. Przedstawione zostały przy tym podstawowe problemy, z którymi trzeba się zmierzyć, konstruując tego typu wskaźnik. Opisana propozycja ma jednak wciąż charakter roboczy. Aby na jej podstawie mogło powstać w pełni wartościowe narzędzie, konieczne są dalsze badania i testy.

Na podstawie pierwszych analiz, których wyniki przedstawione zostały w ostatniej części niniejszego opracowania, można jednak wysnuć pierwsze wnioski. Przede wszystkim widać, że EWD dla nauczycieli może być wartościowym

narzędziem przede wszystkim w odniesieniu do nauczycieli dwóch najważniejszych przedmiotów w gimnazjum: języka polskiego i matematyki. W przypadku nauczycieli przedmiotów, takich jak: biologia, chemia, fizyka, geografia, czy historia praktyczne wykorzystanie EWD może już napotykać duże trudności. Choć w ramach przeprowadzonych analiz modele wykazały się dosyć dobrą trafnością rekonstruowania efektów *tego rodzaju* nauczycieli, o tyle bardzo rzadko występują w ich przypadku różnice efektywności, które w ramach reguł wnioskowania statystycznego można uznać za istotne. Nie ulega za to wątpliwości, że nie ma podstaw, by wyliczać EWD dla nauczycieli przedmiotów nauczanych jedynie w niewielkim wymiarze godzin. Uzyskiwane wskaźniki są tu zupełnie niemiarodajne.

Dodatkowo należy zaznaczyć, że jakość EWD dla nauczycieli bardzo silnie będzie zależeć od specyfiki konkretnej szkoły. Dla uzyskania dobrych oszacowań trudne do przecenienia znaczenie ma *krzyżowanie się* przydziału nauczycieli do oddziałów. Oczywiście czynnik ten jest pośrednio związany np. z wielkością gimnazjum¹¹, jednak w sporej mierze zależy także od działań dyrekcji. Świadome zarządzanie przez dyrektora zasobami ludzkimi, optymalizowane właśnie pod kątem efektywności obliczania wskaźników, wydaje się mieć bardzo duże znaczenie dla powodzenia wprowadzania w szkole EWD dla nauczycieli.

Bibliografia:

1. Dolata R. (red.). 2007. *Edukacyjna wartość dodana jako metoda oceny efektywności nauczania na podstawie wyników egzaminów zewnętrznych*. CKE, Warszawa.
2. Harris, Douglas. 2008. *Would Accountability Based on Teacher Value-Added Be Smart Policy? An Examination of the Statistical Properties and Policy Alternatives*. Materiały National Conference on Value-Added Modeling, University of Wisconsin-Madison 22-24.04.2008.
3. Jakubowski M., Pokropek A. 2009. *Badając egzaminy: Podejście ilościowe w badaniach edukacyjnych*. CKE, Warszawa.
4. Majkut P. 2008. *Szacowanie edukacyjnej wartości dodanej (EWD) dla nauczyciela: możliwości i ograniczenia* [w:] B. Niemierko, M. K. Szmigel (red.) *Uczenie się i egzamin w oczach nauczyciela* Materiały XIV Konferencji Diagnostyki Edukacyjnej, Opole 26-28.09.2008, Polskie Towarzystwo Diagnostyki Edukacyjnej, s. 313-320.
5. McCaffrey, Daniel, J.R. Lockwood, Daniel Koretz, Laura Hamilton. 2003. *Evaluating Value-Added Models for Teacher Accountability*. Rand Corporation.
6. McCaffrey, Daniel, J. R. Lockwood, Daniel Koretz, Thomas Louis, Laura Hamilton. 2004. „Models for Value-Added Modeling of Teacher Effects”. *Journal of Educational and Behavioral Statistics* 29 (1), s. 67-101.
7. Pinheiro, Jose, Douglas Bates. 2000. *Mixed Effects Models in S and S-Plus*. Springer-Verlag, New York.

¹¹ W skrajnym przypadku szkoły jednooddziałowej nie ma po prostu żadnej możliwości krzyżowania. Ogólnie można zaś stwierdzić, że im więcej oddziałów (i nauczycieli), tym większe pole manewru przy przydzielaniu nauczycieli do oddziałów.

8. Raudenbush, Stephen. 2004. „What Are Value-Added Models Estimating and What Does This Imply for Statistical Practice?”. *Journal of Educational and Behavioral Statistics* 29 (1), s. 121-129.
9. Rothstein, Jesse. 2008. „Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement”. *NBER Working Papers* 14442. National Bureau of Economic Research, Inc.
10. Rothstein, Jesse. 2009. „Student Sorting and Bias in Value-Added Estimation: Selection on Observables and Unobservables”. *Education Finance and Policy* 4 (4): s. 537-571.
11. Sanders, William, Sandra Horn. 1994. „The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment”, *Journal of Personnel Evaluation in Education* 8 (3), s. 299-311.
12. Skrondal, Anders, Sophia Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. CRC Press.

E. PRZEDMIOTY HUMANISTYCZNE

