

**dr Tomasz Żółtak**

Centralna Komisja Egzaminacyjna w Warszawie

Polska Akademia Nauk, Instytut Filozofii i Socjologii

## Nowe kierunki rozwoju polskich wskaźników EWD

### Abstrakt

W tekście przedstawiono wyniki prac nad udoskonaleniem metod obliczania i prezentacji polskich wskaźników EWD dla szkół kończących się maturą, podejmowane w Centralnej Komisji Egzaminacyjnej od 2019 roku. Omówiono w nim motywacje leżące u podstaw opracowania nowej metody obliczania wskaźników, zaproponowane rozwiązania, wyniki symulacyjnych badań własności nowych wskaźników oraz trudności, jakie zdiagnozowane w trakcie analizowania ich własności przy wykorzystaniu do ich obliczania prawdziwych danych egzaminacyjnych.

### Wprowadzenie

Wskaźniki edukacyjnej wartości dodanej (EWD), mające na celu wsparcie oceny jakości pracy szkoły dzięki porównaniu wyników jej uczniów z egzaminów na początku oraz na końcu danego etapu kształcenia, były w Polsce rozwijane w latach 2004–2015 (Dolata i in., 2007, 2013, 2015; Karwowski, 2013). Najpierw obliczano je z wykorzystaniem bardzo prostych metod statystycznych, jedynie w odniesieniu do gimnazjów i dla bardzo wąskiego zakresu przedmiotów. Z czasem stosowane podejście modyfikowano. Wprowadzono nowe, bardziej skomplikowane sposoby modelowania statystycznego, zarówno w aspekcie skalowania wyników egzaminów (Pokropek, 2011; Żółtak, 2015b), jak i na potrzeby analizy związków między wynikami egzaminu *na wejściu* i wynikami egzaminu *na wyjściu* a *efektywnością* pracy szkoły (Pokropek, 2009; Pokropek i Żółtak, 2012; Żółtak, 2015a). Pozwoliło to na zwiększenie zestawu publikowanych wskaźników o odnoszące się do liceów ogólnokształcących i techników oraz na rozszerzenie zakresu przedmiotów uwzględnianych przy ich obliczaniu (Dolata i in., 2015; Żółtak, 2020, s. 232–258).

Po zakończeniu z końcem 2015 roku projektów badawczo-wdrożeniowych finansowanych ze środków europejskich metody obliczania i sposób publikacji polskich wskaźników EWD nie podlegały już poważnym modyfikacjom. Jedynym wyjątkiem stało się tu zaprzestanie publikacji wskaźników dla gimnazjów – w związku ze zmianą struktury polskiego systemu edukacji, przywracającą ośmioletnią szkołę podstawową – oraz związane z tym zaprzestanie publikowania tzw. Kalkulatora EWD, pozwalającego na prowadzenie analiz wewnątrzszkolnych (por. Dolata i in., 2015).

Istniała jednak świadomość pewnych ograniczeń i wad wskaźników EWD w ich obecnej formie. W konsekwencji Wydział Analiz Wyników Egzaminacyjnych CKE podjął w 2019 roku na nowo prace nad rozwojem metod obliczania EWD.

Początkowo prowadzono je w węższym gronie. Wraz z krystalizowaniem się nowego podejścia i z uzyskaniem wyników pierwszych analiz własności metod proponowanych do wdrożenia poszerzono w 2021 roku grono ekspertów zaangażowanych w konsultowanie przygotowywanych rozwiązań: do współpracy zaangażowano dużą część zespołu pracującego nad rozwojem wskaźników EWD w latach 2009–2015. Niniejszy tekst ma za zadanie przybliżyć szerszemu gronu odbiorców wyniki tych prac oraz trudności wiążące się z nimi.

W kolejnych częściach tekstu opisuję kolejno problemy z obecnie wykorzystywanymi wskaźnikami EWD, którym starano się w ramach prowadzonych prac zaradzić, i zaproponowane sposoby rozwiązania tych kwestii wraz z wynikami dostępnych analiz obrazujących, w jakim stopniu udało się to osiągnąć. Artykuł kończy refleksja na temat uwarunkowań pozytywnego przyjęcia proponowanych zmian przez odbiorców wskaźników EWD.

### **Ograniczenia dotychczas stosowanych metod obliczania polskich wskaźników EWD**

Kwestię wyraźnego związku pomiędzy wartościami polskich wskaźników EWD a średnimi wynikami uczniów szkół *na wejściu* rozpoznano już dosyć dawno (Żółtak, 2011, 2015a, s. 47–49). Jednocześnie stwierdzono, że ma ona nieliniowy przebieg, co sprawia, że jej adekwatne wyeliminowanie okazuje się trudne. Zachodziły też uzasadnione obawy, że uwzględnienie średnich wyników *na wejściu* do przewidywania wyników *na wyjściu* w modelu EWD wytrąci z obliczanych wskaźników również pewną część prawdziwego zróżnicowania jakości pracy szkół (McCaffrey i in., 2003, s. 68–75; OECD, 2008, s. 125–139; Pokropek, 2013; Raudenbush i Willms, 1995). Waga problemu zwiększyła się, gdy (wobec zmian w strukturze systemu oświaty) wskaźniki EWD zaczęto publikować wyłącznie dla szkół kończących się maturą, gdyż w liceach ogólnokształcących związek między średnimi wynikami *na wejściu* a wartościami EWD okazuje się znacznie silniejszy – średnia wartości współczynnika korelacji liniowej dochodzi tu do 0,74 dla wskaźnika odnoszącego się do przedmiotów matematyczno-przyrodniczych, wobec 0,56 – w gimnazjach, a 0,54 – w technicach (Żółtak, 2020, s. 324). Konsekwencją tego faktu stał się też bardzo silny związek wskaźników EWD ze średnimi wynikami uczniów *na wyjściu* (średnia korelacja Pearsona równa 0,92 dla przedmiotów matematyczno-przyrodniczych), co sprawia, że wskaźniki EWD niosą ze sobą relatywnie niewiele takich informacji o szkołach, jakich nie przekazywałyby już wyniki egzaminów końcowych. Dlatego za jeden z podstawowych postulatów wobec nowo opracowywanych metod obliczania wskaźników EWD przyjęto osłabienie związku między średnimi wynikami *na wejściu* a EWD.

Kolejną kwestią, co do której było wiadomo, że jest problematyczna, ale w toku dotychczasowego rozwoju wskaźników EWD (nie tylko w Polsce) nie zdawano sobie w pełni sprawy z jej konsekwencji, okazało się zróżnicowanie wielkości błędów pomiaru umiejętności uczniów za pomocą wyników egzaminów (McCaffrey i in., 2004, s. 73; Żółtak, 2015a, s. 58). Tymczasem analizy przeprowadzone przez Bartosza Kondratka w latach 2021–2022 z wykorzystaniem tzw. wartości potencjalnych (ang. *plausible values*, zob. Kondratek i Pokropek, 2015, s. 27)

wykazały, że przebieg nieliniowego związku między przewidywaniami umiejętności uczniów na egzaminie gimnazjalnym i na maturze, uzyskanymi za pomocą dotychczas wykorzystywanych estymatorów EAP, pozwala się zasadniczo wytłumaczyć większą wielkością błędu pomiaru szczególnie niskich i szczególnie wysokich umiejętności (w porównaniu z umiejętnościami zbliżonymi do średniej). Prowadzi to do wniosku, że wspomniany wyżej nieliniowy przebieg, którego adekwatne modelowanie stało się znacznym wyzwaniem technicznym i metodologicznym, okazuje się statystycznym artefaktem. Odpowiednim sposobem poradzenia sobie z nim stałoby się z kolei zwrócenie się przy obliczaniu wskaźników EWD ku jakiejś formie modelowania *latentnego*, czy to z wykorzystaniem *wartości potencjalnych*, czy też – modeli równań strukturalnych. Chociaż pewne próby w tym kierunku już podjęto w przeszłości, nie doprowadziły one do rozwiązań nadających się do wdrożenia (Żółtak, 2015a, s. 58–62).

Wiąże się z tym kolejne zagadnienie: wykorzystanie w skalowaniu wyników egzaminów adekwatnych sposobów modelowania związku między poziomem umiejętności a prawdopodobieństwem udzielenia poprawnej odpowiedzi na konkretne pytanie. Ze względu na ograniczenia oprogramowania, które wykorzystywane było do tej pory do skalowania wyników egzaminów na potrzeby obliczania polskich wskaźników EWD, niemożliwe było użycie tzw. trzyparametrycznego modelu logistycznego (3PL), pozwalającego uwzględnić podejmowanie prób zgadnięcia odpowiedzi na pytania testowe przez uczniów o niskim poziomie umiejętności (por. Żółtak, 2015a, s. 23–34). Jest to o tyle istotne, że w zależności od rodzaju wykorzystanego dla danego zadania modelu IRT zmienia się przebieg związku między poziomem umiejętności a wielkością błędu pomiaru (por. Kondrątek i Pokropek, 2015, s. 27–28). Zatem jeśli określone zadania istotnie są narażone na zgadywanie, to uwzględnienie tego faktu ma zasadnicze znaczenie dla możliwości adekwatnego wymodelowania związku między umiejętnościami uczniów *na wejściu* a umiejętnościami uczniów *na wyjściu* – na podstawie danych informujących o punktacji uzyskanej za poszczególne zadania egzaminacyjne.

Ostatnia problematyczna kwestia nie ma – w odróżnieniu od wcześniejszych – charakteru statystycznego, lecz wiąże się ze sposobem wykorzystania wskaźników EWD przez ich użytkowników. Dla liceów ogólnokształcących i dla techników publikowano do tej pory po cztery różne wskaźniki EWD: dwa z nich odnosiły do pojedynczych przedmiotów obowiązkowo zdawanych przez wszystkich maturzystów (język polski, matematyka), a dwa – do szeroko zdefiniowanych grup przedmiotów (humanistyczne, matematyczno-przyrodnicze), obejmujących również przedmioty obowiązkowe, wymienione wcześniej. O ile objęcie jednym wskaźnikiem wielu przedmiotów, w tym jednego przedmiotu obowiązkowego, miało tę zaletę, że umożliwiało opublikowanie wyników dla bardzo dużej liczby szkół, o tyle stawało się również przyczyną trudności z ich interpretacją. Aby zrozumieć przyczyny uzyskania konkretnej wartości wskaźnika przez szkołę, należało posiłkować się tabelą opisującą liczbę zdających poszczególne przedmioty maturalne – w ujęciu ogólnym: dany przedmiot miał tym większy wpływ na uzyskaną wartość EWD, im większa była liczba zdających go uczniów, którzy zostali uwzględnieni w analizach (Dolata i in., 2015). Wyizolowanie – na podstawie takiego wskaźnika – informacji o pojedynczym przedmiocie było właściwie niemożliwe. W konsekwencji przy porównywaniu szkół, które znacząco

różniły się profilem przedmiotów maturalnych wybieranych przez ich uczniów, trudno było określić, z czego dokładnie wynikają ewentualne różnice prezentowanych wartości wskaźników EWD i średnich wyników matury.

## Propozycje zmian

### Skalowanie wyników egzaminów

Propozycje zmian metod obliczania wskaźników EWD przygotował Bartosz Kondratek. W zakresie sposobu skalowania wyników egzaminów objęły one przede wszystkim wykorzystanie w tym celu programu *uirt* (Kondratek, 2022), umożliwiającego modelowanie zgadywania odpowiedzi na pytania zamknięte, w tym automatyczne wykrywanie zadań (spośród zadanego zestawu zadań zamkniętych), w których zjawisko to występowało.

### Zakres przedmiotowy wskaźników

Druga poważna zmiana wiązała się z propozycją odejścia od obliczania wskaźników uwzględniających wiele przedmiotów maturalnych jednocześnie na rzecz obliczania EWD oddzielnie dla każdego z nich. Pozwoliłoby to znacznie uprościć interpretację publikowanych wskaźników, chociaż – oczywiście – odbyłoby się to kosztem ograniczenia liczby szkół, dla których poszczególne wskaźniki będą dostępne. Aby to zilustrować, przeprowadzono analizy z wykorzystaniem danych egzaminacyjnych obejmujących roczniki absolwentów 2020–2022. Sprawdzono w nich, jak kształtowałyby się liczba szkół, dla których publikowano by trzyletnie wskaźniki EWD, w stosunku do liczby szkół, dla których są dostępne wskaźniki z przedmiotów obowiązkowych. Tylko dla biologii i geografii w liceach ogólnokształcących i geografii w technikach ten odsetek wyniósłby około 50%. W liceach byłoby to jeszcze około 37% dla chemii i około 25% dla historii, dla WOS-u i dla fizyki. Z kolei w technikach byłoby to jeszcze około 10% dla biologii, dla fizyki i dla informatyki. Z kolei wskaźniki z informatyki dla liceów, a z chemii, z historii i z WOS-u dla techników byłyby dostępne jedynie dla kilkudziesięciu szkół w skali całego kraju.

Na potrzeby interpretacji przytoczonych wyników należy zastrzec, że otwartą kwestią pozostaje, jaki wpływ na wybieralność przedmiotów maturalnych będzie mieć zmiana formuły tego egzaminu, wchodząca właśnie w życie. Jednak można bezpiecznie przyjąć, że duża część wskaźników będzie dostępna dla relatywnie niewielkiej części szkół. Z drugiej strony – poprzez zastosowanie takich wskaźników można uzyskać porównanie ze sobą właśnie tych szkół, które specjalizują się w kształceniu w określonym przedmiocie (a więc większa jest w nich liczba zdających go uczniów). Pozwala to zakładać, że wskaźniki EWD obliczane oddzielnie dla poszczególnych przedmiotów stałyby się dla szkół znacznie bardziej użyteczne w kontekście prób oceny skuteczności praktyk dydaktycznych, które są w nich stosowane.

### Model EWD

Najpoważniejsza z proponowanych zmian dotyczy wdrożenia *latentnych* wskaźników EWD, obliczanych w sposób dający możliwość w pełni adekwatnego kontrolowania błędu pomiaru umiejętności za pomocą egzaminów.



Wymaga to użycia modelu statystycznego, który będzie obejmował jednocześnie tzw. część pomiarową, opisującą związek między poziomem umiejętności a wynikami uzyskanymi za poszczególne zadania danego egzaminu, oraz tzw. część strukturalną, opisującą związek między umiejętnościami uczniów *na wyjściu* a umiejętnościami uczniów *na wejściu*, efektywnością pracy szkół (której oszacowaniem jest wartość EWD) i ewentualnymi dodatkowymi zmiennymi kontrolnymi (dla polskich wskaźników EWD: płcią i posiadaniem na egzaminie zaświadczenia o dysleksji).

Aby uprościć estymację tak złożonego modelu, zdecydowano o przyjęciu podejścia polegającego na estymowaniu najpierw parametrów jego części pomiarowej – skaluje się oddzielnie wyniki egzaminu *na wyjściu* i *na wejściu* (przy czym – aby uwzględnić uczniów o toku kształcenia wydłużonym o rok – konieczne staje się tu przeprowadzenie skalowania jednocześnie dla dwóch kolejnych roczników egzaminu *na wejściu*). Następnie wartości parametrów (oraz ich macierze kowariancji) w ten sposób uzyskane są przekształcane liniowo tak, aby zagwarantować wartość oczekiwaną 0 i odchylenie standardowe 1 wśród grupy uczniów uwzględnianych przy obliczaniu danego wskaźnika EWD. W dalszej estymacji, odnoszącej się do części strukturalnej modelu, są już one traktowane jako dane (stałe niepodlegające estymacji). Ta procedura w swoim ogólnym kształcie okazuje się analogiczna do podejścia rozwijanego w ostatnich latach w odniesieniu do modeli strukturalnych, w których obserwowane wskaźniki są ciągłe (a nie – jak w przypadku egzaminów – binarne lub porządkowe), określanego jako *global structural after measurement* (Rosseel i Loh, 2022).

Bartosz Kondratek zaproponował bardzo poważne zmiany w odniesieniu do części strukturalnej modelu, którą można zestawić z dotychczas stosowaną formą modeli EWD. W odróżnieniu od wcześniej stosowanego podejścia, w którym wyniki egzaminu *na wyjściu* były przewidywane ze względu na wyniki egzaminu *na wejściu*, nowa propozycja jest w swojej ogólnej formie analogiczna do tzw. *layered model*, zaproponowanego przez Williama Sandersa i dosyć szeroko stosowanego w USA do obliczania wskaźników EWD nauczycieli (McCaffrey i in., 2004; Sanders i Horn, 1994). Jednostką obserwacji w takim modelu jest wynik konkretnego egzaminu konkretnego ucznia. W przypadku polskich wskaźników EWD każdy uczeń jest więc reprezentowany przez dwie obserwacje. Jedna opisuje jego umiejętności podczas egzaminu *na wejściu*, a druga opisuje jego umiejętności podczas egzaminu *na wyjściu*. Obie te umiejętności są w ramach modelu przewidywane ze względu na efekty opisujące: a) średnie umiejętności uczniów danej szkoły *na wejściu*, b) średnią zmianę relatywnego poziomu umiejętności uczniów danej szkoły między egzaminem *na wejściu* a egzaminem *na wyjściu* (oszacowanie wielkości tego efektu będzie w ramach tego podejścia wartością wskaźnika EWD) oraz c) składnik błędu indywidualnego, odpowiedzialny za opisanie odchylenia poszczególnych obserwacji od sumy wcześniej wymienionych efektów (przy czym w odniesieniu do wyników egzaminu *na wejściu* wartość efektu związanego ze zmianą przyjmuje się – oczywiście – za równą 0).

Poważne wyzwanie stanowi estymacja parametrów takiego modelu z uwzględnieniem niepewności pomiarowej dotyczącej zarówno poziomu umiejętności poszczególnych uczniów, jak i oszacowań parametrów części pomiarowej modelu.

W tym celu Bartosz Kondratek napisał specjalny pakiet języka Python – *pvreg* – w którym została zaimplementowana jego autorska metoda szacowania błędów standardowych oszacowań wymienionych wcześniej efektów części strukturalnej modelu z wykorzystaniem uproszczonego podejścia MCMC (ang. *Markov Chain Monte Carlo*). Pozwala ona również uzyskać wartości potencjalne umiejętności poszczególnych uczniów podczas egzaminu *na wejściu* i podczas egzaminu *na wyjściu*, obliczone przy założeniu, że związek między tymi umiejętnościami został adekwatnie opisany przez założony model strukturalny.

Analiza własności wskaźników obliczonych z wykorzystaniem zaproponowanej metody w odniesieniu do uczniów kończących licea i technika w latach 2020–2022 pokazała, że wskaźniki te wyraźnie się różnią od dotychczas stosowanych. Większe zróżnicowanie wartości nowo proponowanych wskaźników EWD nie jest zaskoczeniem, gdyż jest cechą charakterystyczną dla użycia modeli *latentnych* (Żółtak, 2015a, s. 58–62). Jednak – w związku z uwzględnieniem w modelu błędu pomiaru związanego z wynikami egzaminów – towarzyszy temu zwiększenie się wartości błędów standardowych wskaźników, czyli – w konsekwencji – powiększenie się obszarów ufności (elips prezentowanych użytkownikom). Łączny efekt obu tych zmian to wyraźne zmniejszenie się liczby szkół, które można uznać za mające EWD istotnie statystycznie różne od 0.

Co bardzo istotne, nowo proponowane wskaźniki okazały się jedynie umiarkowanie silnie powiązane z dotychczas obliczanymi – dla wskaźników trzyletnich obejmujących roczniki maturzystów 2020–2022 korelacje Pearsona wyniosły 0,51 i 0,64 – odpowiednio – dla liceów i dla techników w odniesieniu do języka polskiego, a 0,65 i 0,81 w odniesieniu do matematyki (po uwzględnieniu tylko tych szkół, których wskaźniki są prezentowane). Różnice między wskaźnikami układają się w sposób systematyczny – w nowo proponowanym podejściu niższe wartości wskaźników EWD typowo odpowiadają szkołom o wyższych średnich wynikach *na wejściu*, a wartości wyższe okazują się typowe dla szkół o niższych wynikach *na wejściu*. Wiąże się z tym dramatyczna wręcz zmiana siły związku wskaźników EWD ze średnimi wynikami *na wejściu*. Dla wskaźników matematycznych w nowo proponowanym podejściu taki związek właściwie nie występuje, a dla języka polskiego okazuje się wręcz negatywny (korelacja Pearsona równa  $-0,37$  zarówno w liceach, jak i w technikach).

O ile ta ostatnia własność proponowanych wskaźników niejako wychodzi na przeciw oczekiwaniom odnośnie do pożądaných zmian własności wskaźników EWD, o tyle pojawiają się poważne wątpliwości, czy nie jest to jednak zmiana zbyt duża. Rozpatrzmy tę kwestię czysto teoretycznie: w systemie, w którym rekrutacja do szkół na danym etapie edukacji jest konkursowa i prowadzona na podstawie wyników egzaminacyjnych, należałoby się raczej spodziewać pozytywnego związku między średnimi wynikami *na wejściu* a jakością pracy szkoły (por. Żółtak, 2020, s. 125–128), chociaż zapewne dalece mniej silnego niż w przypadku obecnie wykorzystywanych wskaźników EWD. Jeśli miałyby on nie występować lub wręcz być negatywny, musiałyby to się zapewne wiązać z wadliwie działającymi mechanizmami przepływu informacji o jakości pracy szkół do rodziców tych uczniów, i samych uczniów, którzy kończą wcześniejszy etap kształcenia. Trudno bowiem sobie wyobrazić, aby osoby mogące wybrać

szkołę – z racji uzyskanych wysokich wyników egzaminacyjnych – świadomie wybierały placówki dające im mniejsze szanse na uzyskanie dobrych postępów w nauce na kolejnym etapie edukacji.

Bliższe przyjrzenie się temu problemowi od strony empirycznej wykazało, że wśród szkół rekrutujących uczniów o szczególnie wysokich wynikach egzaminów nie ma w ogóle takich, które w ramach nowo proponowanego podejścia uzyskiwałyby pozytywne wartości wskaźników EWD. Jest to związane z faktem, że pod względem formalnym zaproponowany *model strukturalny* można opisać jako wariant tzw. modelu wzrostu (p. Żółtak, 2020, s. 285–293), odznaczający się tą własnością, że ustala on *oczekiwania* co do średnich umiejętności uczniów danej szkoły *na wyjściu* w tym samym punkcie (gdzie chodzi o rozkład danego egzaminu wystandaryzowany w ramach analizowanej populacji) co średnie umiejętności uczniów tej szkoły *na wejściu*. To oczekiwanie przyjmuje się arbitralnie w tym sensie, że wynika ono z formalnych własności założonego *modelu strukturalnego*, a nie z własności analizowanych danych. Zatem – tak jak zachodzi to w polskich liceach charakteryzujących się szczególnie dobrym naborem – empirycznie może się okazać, że dla pewnego zakresu średnich umiejętności *na wejściu* nie ma żadnych szkół, które zdołałyby tak postawione *oczekiwania* przekroczyć lub chociażby wypełnić. Jednocześnie należy zauważyć, że takiego założenia nie przyjmowano w dotychczas stosowanych modelach EWD, w których współczynnik, przez jaki należy przemnożyć umiejętności *na wejściu*, aby uzyskać (optymalne ze względu na związki między zmiennymi zachodzące w danych) oczekiwane umiejętności *na wyjściu*, był parametrem modelu podlegającym estymacji.

Kwestia, w jakich okolicznościach założenie zawarte w modelu zaproponowanym przez Bartosza Kondratka i opisane powyżej staje się zasadne, a w jakich nie, wymaga dalszej, pogłębionej refleksji. Wydaje się jednak, że może się ono okazać problematyczne w sytuacji, gdy oba modelowane egzaminy mierzą nieco inne umiejętności (por. Żółtak, 2020, s. 285–287). Można jeszcze odnotować, że było ono krytykowane również w przypadku amerykańskiego *layered model*, w odniesieniu do którego zaproponowano modyfikację pozwalającą z tego założenia zrezygnować (McCaffrey i in., 2004).

Obecnie prace nad przygotowaniem nowej metody obliczania wskaźników EWD są ukierunkowane na wykorzystanie *wartości potencjalnych* wygenerowanych przez program *uirt* w modelach regresji mieszanych efektów o formie zbliżonej do tych, jakie były dotychczas wykorzystywane. Takie rozwiązanie pozwala na uwzględnienie błędów pomiaru przy jednoczesnym uzyskaniu większej elastyczności co do specyfikacji formy modelu stosowanego do obliczania samych wartości wskaźników EWD. Dodatkowych analiz wymaga jednak kwestia stopnia, w jakim konkretna forma *modelu strukturalnego* założonego w programie *uirt* w ramach procedury prowadzącej do wygenerowania *wartości potencjalnych* (gdzie wykorzystuje się niezmiennie model zaproponowany przez Bartosza Kondratka) wpływa (lub nie) na własności statystyczne tychże *wartości potencjalnych* w późniejszych analizach, a w konsekwencji – na wartości uzyskiwanych wskaźników EWD.

## Możliwości oddziaływania nowych wskaźników EWD na system edukacji

Z myślą o ewentualnych zmianach w sposobie obliczania i prezentacji polskich wskaźników EWD warto jeszcze rozważyć, jak te zmiany mogą wpłynąć na społeczny odbiór wskaźników oraz na popularność ich wykorzystania w ramach systemu edukacji. W istocie – o zakresie obecnego wykorzystania wskaźników EWD wiemy bardzo niewiele. Ostatnie badania na ten temat przeprowadzono przed niemal 10 laty, w ramach ewaluacji projektów unijnych poświęconych rozwojowi wskaźników EWD. Można przy tym założyć, że w ostatnich latach popularność wskaźników EWD raczej się zmniejszała, chociażby ze względu na – związane ze zmianami w strukturze kształcenia – zakończenie publikowania wskaźników EWD gimnazjów. Wiadomo bowiem, że w przeszłości to w szkołach tego typu korzystanie z nich było najbardziej rozpowszechnione (Bąbiak i in., 2014; Miszczuk i Karmiński, 2015). Bezspornym faktem – ograniczającym zainteresowanie ze strony dyrektorów i nauczycieli – pozostaje brak odniesień do wskaźników EWD w oficjalnych dokumentach opisujących wymagania państwa wobec szkół. Również działalność szkoleniową i promocyjną dotyczącą wykorzystania EWD prowadzono po 2015 roku jedynie w bardzo ograniczonym zakresie.

W tym kontekście wydaje się, że o ile proponowane zmiany mają potencjał uczynienia polskich wskaźników EWD obiektywnie bardziej użytecznymi – czy to poprzez dostarczenie bardziej szczegółowych informacji w rozbiciu na poszczególne przedmioty, czy to poprzez poprawę trafności tych wskaźników, jak również ograniczenie (wedle wszelkiego prawdopodobieństwa nadmiernie silnego) związku między EWD a średnimi wynikami *na wejściu*, o tyle nie jest wcale oczywiste, że nieuchronnie przełożą się one na subiektywne poczucie użyteczności ze strony ich potencjalnych użytkowników. Konieczne w tym celu stałoby się opracowanie odpowiednich materiałów informacyjnych i szkoleniowych, a być może również całego programu szkoleń. Znaczenie przygotowania odpowiedniej strategii komunikacji odnośnie do wprowadzanych zmian będzie tym większe, im większa okaże się skala zmian wartości (i ogólniej: własności) nowo wprowadzanych wskaźników EWD w porównaniu z obecnie stosowanymi. Zaniedbanie tej kwestii może skutkować utratą zaufania (i zainteresowania) ze strony obecnych użytkowników przy jednoczesnym niepozyskaniu nowych. Tymczasem wcześniejsze badania nad sposobami wykorzystania wskaźników EWD przez szkoły wskazują, że budowa kultury wykorzystania tych wskaźników w sposób zaangażowany, a nie wyłącznie fasadowy, okazuje się zadaniem trudnym i przynoszącym ograniczone efekty nawet przy relatywnie dużym natężeniu działalności upowszechniającej i szkoleniowej (Bąbiak i in., 2014; Kędracka i in., 2015; Miszczuk i Karmiński, 2015).



**Bibliografia**

- Bąbiak I., Matuszczak K. i Zielonka P. (2014), *Raport z ewaluacji wewnętrznej projektu EWD*, Instytut Badań Edukacyjnych.
- Dolata R., Hawrot A., Humenny G., Jasińska A., Koniewski M., Majkut P. i Żółtak T. (2013), *Trafność metody edukacyjnej wartości dodanej dla gimnazjów*, Instytut Badań Edukacyjnych.
- Dolata R., Hawrot A., Humenny G., Jasińska-Maciążek A., Rappe A., Stożek E. i Żółtak T. (2015), *Metoda edukacyjnej wartości dodanej w Polsce [w:] R. Dolata i M. Sitek (red.), Raport o stanie edukacji 2014. Egzamininy zewnętrzne w polityce i praktyce edukacyjnej* (s. 145–219), Instytut Badań Edukacyjnych.
- Dolata R., Jakubowski M., Pokropek A., Rappe A. i Stożek E. (2007), *Edukacyjna wartość dodana jako metoda oceny efektywności nauczania*, Centralna Komisja Egzaminacyjna.
- Karwowski M. (red.) (2013), *Ścieżki rozwoju edukacyjnego młodzieży – szkoły pogimnazjalne. Trafność wskaźników edukacyjnej wartości dodanej dla szkół maturalnych*. Wydawnictwo IFiS PAN.
- Kędracka E., Rappe A. i Stożek E. (2015), *Opis poziomów wykorzystania wskaźników EWD przez szkoły [w:] G. Mazurkiewicz (red.), Ucząca się szkoła. Od rozwoju jednostek do rozwoju wspólnoty*, Wydawnictwo Uniwersytetu Jagiellońskiego.
- Kondratak B. (2022), *uirt: A command for unidimensional IRT modeling*, „The Stata Journal. Promoting Communications on Statistics and Stata”, vol. 22(2) (s. 243–268), <https://doi.org/10.1177/1536867X221106368>
- Kondratak B. i Pokropek A. (2015), *Teoria odpowiedzi na pozycje testowe. Jednowymiarowe modele dla cech ukrytych o charakterze ciągłym [w:] A. Pokropek (red.), Modele cech ukrytych w badaniach edukacyjnych, psychologii i socjologii. Teoria i zastosowania* (s. 15–32), Instytut Badań Edukacyjnych.
- McCaffrey D.F., Lockwood J.R., Koretz D.M. i Hamilton L.S. (2003), *Evaluating Value-Added Models for Teacher Accountability*. RAND.
- McCaffrey D.F., Lockwood J.R., Koretz D.M., Louis T.A. i Hamilton L. (2004), *Models for Value-Added Modeling of Teacher Effects*, „Journal of Educational and Behavioral Statistics”, vol. 29(1) (s. 67–101).
- Miszczuk E. i Karmiński M. (2015), *Badania na temat wykorzystania edukacyjnej wartości dodanej (EWD) przez szkoły*, Instytut Badań Edukacyjnych.
- OECD (2008), *Measuring Improvements in Learning Outcomes. Best Practices to Assess the Value-Added of Schools*, OECD.
- Pokropek A. (2009). *Metody statystyczne wykorzystywane w szacowaniu trzyletnich wskaźników egzaminacyjnych [w:] M.K. Szmigel i B. Niemierko (red.), Badania międzynarodowe i wzory zagraniczne w diagnostyce edukacyjnej. XV Konferencja Diagnostyki Edukacyjnej, Kielce 4–6.12.2009* (s. 137–153), Grupa Tomami.
- Pokropek A. (2011), *Matura z języka polskiego. Wybrane problemy psychometryczne [w:] B. Niemierko i M.K. Szmigel (red.), Ewaluacja w edukacji. Koncepcje, metody, perspektywy XVII Konferencja Diagnostyki Edukacyjnej, Kraków 23–25.09.2011* (s. 438–450), Grupa Tomami.
- Pokropek A. (2013), *Efekt rówieśników w nauczaniu szkolnym*, Instytut Badań Edukacyjnych.
- Pokropek A. i Żółtak T. (2012), *Nowe modele jednorocznej EWD [w:] M.K. Szmigel i B. Niemierko (red.), Regionalne i lokalne diagnozy edukacyjne. XVIII Konferencja Diagnostyki Edukacyjnej, Wrocław, 21–23.09.2012 r.* (s. 178–187), Grupa Tomami.
- Raudenbush S.W. i Willms J.D. (1995), *The Estimation of School Effects*, „Journal of Educational and Behavioral Statistics”, vol. 20(4) (s. 307–335).
- Rossee Y. i Loh W.W. (2022), *A structural after measurement approach to structural equation modeling*, „Psychological Methods”, <https://doi.org/10.1037/met0000503>

- Sanders W. i Horn S.P. (1994), *The Tennessee Value-Added Assessment System (TVA-AS). Mixed-Model Methodology in Educational Assessment*, „Journal of Personnel Evaluation in Education”, vol. 8(3) (s. 299–311).
- Żółtak T. (2011), *Znaczenie informacji o średnim wyniku uczniów na wejściu dla własności jednorocznych wskaźników EWD gimnazjów* [w:] B. Niemierko i M.K. Szmiigel (red.), *Ewaluacja w edukacji. Koncepcje, metody, perspektywy. XVII Konferencja Diagnostyki Edukacyjnej*, Kraków 23–25.09.2011 (s. 505–523). Grupa Tomami.
- Żółtak T. (2015a), *Statystyczne modelowanie wskaźników Edukacyjnej Wartości Dodanej. Podsumowanie polskich doświadczeń 2005–2015*, Instytut Badań Edukacyjnych.
- Żółtak T. (2015b), *Wykorzystanie IRT w skalowaniu kompozycyjnej miary osiągnięć maturalnych* [w:] A. Pokropek (red.), *Modele cech ukrytych w badaniach edukacyjnych, psychologii i socjologii. Teoria i zastosowania* (s. 308–319), Instytut Badań Edukacyjnych.
- Żółtak T. (2020), *Edukacyjna wartość dodana: Teoria i zastosowania, kontrowersje i perspektywy*. <https://doi.org/10.5281/ZENODO.6965603>