

**Aleksandra Jasińska-Maciążek**

Uniwersytet Warszawski

## **Metody oceniania efektywności nauczania w Polsce i na świecie – zarys historyczny<sup>1</sup>**

W dyskursie dotyczącym edukacji coraz więcej uwagi poświęca się w ostatnich latach problematyce efektywności nauczania i równości szans edukacyjnych. Rośnie zapotrzebowanie na wiarygodne miary podsumowujące jakość nauczania w szkołach. Jest to związane z odchodzeniem od biurokratycznego modelu zarządzania edukacją na rzecz zwiększania autonomii szkół i nauczycieli, decentralizacji zarządzania i finansowania edukacji oraz wprowadzenia elementów konkurencji między szkołami (Sitek, 2015). Wzrasta też znaczenie testów osiągnięć szkolnych jako narzędzia ewaluacji placówek edukacyjnych oraz monitorowania systemu oświaty (EACEA, 2010).

Wyniki pomiaru osiągnięć szkolnych jednak nie wystarczają do trafnej oceny efektywności nauczania, gdyż w dużym stopniu zależą one od czynników, na które szkoła nie ma wpływu, takich jak zdolności i uprzednie osiągnięcia uczniów, ich środowisko rodzinne i rówieśnicze. Dlatego rozwijane są metody oceny efektywności nauczania uwzględniające fakt, że szkoły różnią się pod względem cech uczniów oraz szerszego kontekstu, w którym pracują. Ich celem jest uchwycenie tej części wyników nauczania, którą możemy przypisać działaniom szkoły, a wyłączenie tego, co od szkoły niezależne. Najpopularniejszą obecnie na świecie metodą oceny efektywności są modele edukacyjnej wartości dodanej (EWD), które pozwalają na szacowanie postępów edukacyjnych uczniów.

Metoda EWD była przez dekadę rozwijana i zaimplementowana także w polskim systemie edukacji. Dostosowano ją do panujących rozwiązań oraz zbadało jakość proponowanych modeli. Ustawa z dnia 14 grudnia 2016 roku<sup>2</sup> wprowadziła w Polsce od września 2017 roku nowy ustrój szkolny. Jest to dobry moment na podsumowania. Artykuł opisuje to, co już się zdarzyło w zakresie mierzenia efektywności nauczania, u progu zmiany, której rezultaty dopiero poznamy; także te dotyczące sposobów ewaluacji nauczania w szkołach. W artykule przedstawiono kluczowe polskie wydarzenia i prace dotyczące tego zagadnienia. Opis ten poprzedzono zarysem najważniejszych faktów i przykładów zastosowań modeli efektywności nauczania na świecie, co pomoże w lepszym zrozumieniu sytuacji naszego kraju. Artykuł nie jest wyczerpującym przeglądem historycznym. Ma on na celu pokazanie różnych sposobów myślenia o badaniu efektywności nauczania oraz naświetlenie ich uwarunkowań i konsekwencji.

<sup>1</sup> Artykuł powstał na podstawie fragmentu pracy doktorskiej przygotowanej pod kierunkiem dr. hab. Romana Dolaty, prof. UW: A. Jasińska-Maciążek, (2017). *Modele oceny szkolnej efektywności nauczania*. Warszawa: Uniwersytet Warszawski, Wydział Pedagogiczny.

<sup>2</sup> Ustawa z dnia 14 grudnia 2016 r. – Prawo oświatowe (Dz.U. poz. 59) oraz Ustawa z dnia 14 grudnia 2016 r. – Przepisy wprowadzające ustawę – Prawo oświatowe (Dz.U. poz. 60).

## Początki badań nad efektywnością nauczania i przykłady zastosowań modeli jej oceny w systemach ewaluacyjnych na świecie

Pytanie o to, co robić, by lepiej i skuteczniej nauczać, było i jest jednym z czynników rozwoju dydaktyki (Denek, 1971). Poszczególne szkoły pedagogiczne starały się odpowiedzieć na nie w różny sposób. Jan Fryderyk Herbart, stworzył na początku XIX wieku system nazwany szkołą tradycyjną. Przyjmując perspektywę psychologiczną, wyznaczył stopnie formalne, według których miało odbywać się skuteczne nauczanie. John Dewey swoje poglądy wywiódł z pragmatyzmu, nadając działaniu i odkrywaniu kluczowe znaczenie w procesie poznania. Zasady nauczania wyprowadził z praw myślenia, wyszczególniając pięć etapów, przez które przechodzi uczeń podczas rozwiązywania problemu. Także w różnych kierunkach „nowego wychowania” podkreślano, że dobre nauczanie musi uwzględniać potrzeby i zainteresowania dziecka, odwoływać się do jego aktywności, przeżyć i dawać mu szansę na twórcze działanie. W ciągle aktualnej teorii wielostronnego uczenia się Wincenty Okoń zintegrował dotychczasowe podejścia, uznając, że uczenie się obejmuje przyswajanie, odkrywanie, przeżywanie i działanie (Okoń, 1965).

Szczególne znaczenie dla rozwoju badań nad efektywnością nauczania miały prace Burrhusa Frederica Skinnera z połowy XX wieku poświęcone warunkowaniu instrumentalnemu i nauczaniu programowanemu. Stwierdził on, że nabywanie wiadomości i umiejętności wymaga podzielenia materiału na dużą liczbę małych kroków, po których musi następować wzmocnienie (Skinner, 1968). Poglądy Skinnera, w których przedstawiał nauczanie jako bodziec (przyczynę), a uczenie się jako reakcję (skutek), skierowały uwagę badaczy na proces nauczania-uczenia się oraz interakcje między nauczycielem i uczniem (Fenstermacher i Soltis, 1998). Koncepcja nauczania programowanego wzbudziła zaś w Polsce na tyle duże zainteresowanie, że podjęto badania nad efektywnością tej metody (Denek, 1968, 1971; Kupisiewicz, 1966).

Inny sposób myślenia o szukaniu odpowiedzi na pytania o efektywność nauczania pojawił się wraz z reakcjami środowiska naukowego i politycznego na amerykańskie badania podsumowane w raporcie *Equality of Educational Opportunity* (Coleman i in., 1966), zwanym też raportem Colemana. Jeden z wniosków w sposób szczególnie przebił się do opinii publicznej. Okazało się bowiem, że przyczyn zróżnicowania wyników nauczania należy szukać przede wszystkim w cechach uczniów, takich jak zdolności i charakterystyki ich rodzin. Znacznie mniej z tych różnic daje się wyjaśnić czynnikami szkolnymi. Podobne wnioski przedstawili Christopher Jencks ze współautorami w książce opublikowanej kilka lat później (Jencks i in., 1972). Wyniki te, podsumowywane stwierdzeniem, że „szkoła nie ma znaczenia”<sup>3</sup>, poruszyły środowisko naukowe, przyczyniając się do rozwoju tradycji badań nad szkolną efektywnością nauczania (*school effectiveness research*, SER). W podejściu tym koncentrowano się na różnicach między szkołami, zakładając, że w jednych uczy się efektywniej niż w innych. Punktem wyjścia tych badań był więc nie sam proces

---

<sup>3</sup> *schools make no difference*

czy metoda nauczania, jak we wspomnianych pracach dydaktyków, ale wyniki nauczania. Zakładano, że wpływa na nie wiele czynników szkolnych i pozaszkolnych, które będzie można uchwycić, badając różnice między szkołami.

Początkowo, w odpowiedzi na raport Colemana, badacze starali się dowieść, że szkoła ma znaczenie, koncentrowali się w szczególności na placówkach osiągających wysokie wyniki w pracy z uczniami pochodzącymi z niekorzystnych środowisk. Szczególne znaczenie przypisuje się amerykańskim badaniom Ronalda Edmondsa (1979) oraz Wilbura Brookovera i współpracowników (1979), które pozwoliły zbudować pięcioczynnikowy model korelatów efektywnego kształcenia. W Wielkiej Brytanii pierwszym głośnym badaniem w tej tradycji było badanie Michaela Ruttera (1979), w którym zdefiniowano czynniki mające i czynniki niemające znaczenia dla efektywności nauczania. Późniejsze prace polegały na testowaniu kolejnych hipotez dotyczących uwarunkowań wyników kształcenia oraz rozwoju metodologii statystycznego opisu złożonej rzeczywistości szkolnej (Creemers, 2007; Teddlie i Stringfield, 2007). Wyniki badań efektywności nauczania stawały się często inspiracją dla działań służących poprawie jakości nauczania (*school improvement research*, SIR), gdyż obu nurtom przyswiecał ten sam podstawowy cel – doskonalenie szkoły.

Amerykańscy badacze koncentrowali się równocześnie na rozwoju metodologii szacowania efektu szkoły (czyli tej części zmienności wyników nauczania, którą można przypisać działaniom szkoły) i badaniu jego właściwości (wielkości, stabilności w czasie, trafności itp.). Duże znaczenie dla rozwoju tych starań miały prace ekonomistów, odnoszących się krytycznie do metodologii badań *Equality of Educational Opportunity* (ich argumenty podsumowano np. w: Heckman i Neal, 1996). W kolejnych latach zajęli się oni rozwojem analiz, które w sposób bardziej adekwatny ujęłyby związki między zasobami szkoły, a osiągnięciami edukacyjnymi uczniów. W 1971 roku Eric Hanushek (1971) zaproponował model, w którym osiągnięcia edukacyjne na zakończenie etapu kształcenia są funkcją uprzednich osiągnięć ucznia, jego zdolności, wpływów środowiska rodzinnego, środowiska rówieśniczego i czynników szkolnych oddziałujących na ucznia w okresie nauczania w danej placówce. Podkreślił znaczenie uwzględnienia w modelu uprzednich osiągnięć, bowiem niosą one za sobą informację o istotnych dla wyników nauczania różnorodnych czynnikach, również tych, o których nie wiemy lub których nie możemy wprost uwzględnić w analizach. W zastosowanym modelu uwzględnił także zmienne zero-jedynkowe opisujące przydział uczniów do nauczycieli, by oszacować tę część zmienności wyników testowych, którą można przypisać działaniom i cechom nauczycieli<sup>4</sup>. Podejście takie nazywano badaniem efektu całkowitego i były to pierwsze modele edukacyjnej wartości dodanej (EWD), choć współcześnie stosuje się inny sposób ich liczenia. Podobne badania prowadzili także inni ekonomiści (np. Murnane, 1975).

<sup>4</sup> Miało to na celu oszacowanie *efektu nauczyciela*, analogicznego do wspomnianego *efektu szkoły*. W ramach tego nurtu badań interesowano się bowiem znaczeniem czynników szkolnych dla osiągnięć uczniów lub zawężano przedmiot zainteresowania do czynników, które są powiązane z nauczycielem.

Hanushek (1971) już w pierwszych tekstach poświęconych tej problematyce zaznaczał, że osiągnięcia edukacyjne należałoby rozpatrywać szerzej niż wyniki testów wiadomości i umiejętności, jednak z czasem badacze zaczęli koncentrować się na tej części efektów kształcenia. Kompromis ten wynikał m.in. z faktu, że inne wyniki kształcenia nie dają się łatwo zmierzyć. Częściowo przyczynił się też do tego jeden z najbardziej wpływowych w historii Stanów Zjednoczonych raportów – *A Nation at Risk* (Gardner i in., 1983). Autorzy opisali w nim problemy amerykańskiej oświaty, a zdiagnozowany stan przedstawili jako zagrożenie dla rozwoju kraju. Podkreślili m.in. potrzebę diagnozy osiągnięć szkolnych uczniów z wykorzystaniem standaryzowanych testów. Publikacja ta odwróciła kierunek myślenia o pomiarze jakości szkół z ilości dostarczanych szkole zasobów na ocenę efektów nauczania, czyli osiągnięć szkolnych uczniów na zakończenie edukacji (Guthrie i Springer, 2004). Dodatkowo raport *A Nation at Risk* stworzył odpowiedni klimat polityczny do finansowania zbierania danych edukacyjnych, co pozwoliło na rozwój systemów monitorowania edukacji wykorzystujących dane, a w szczególności pomiary osiągnięć szkolnych. W systemach tych naturalnie zaczęto zestawiać wyniki szkół, a wraz z tym zrodziła się potrzeba sprawiedliwego porównywania, uwzględniającego środowisko rodzinne uczniów. W budowaniu odpowiednich wskaźników zaczęto więc korzystać z doświadczeń badań efektywności nauczania. W niektórych stanach wprowadzono konsekwencje dla szkół za uzyskiwane wyniki, rozpoczynając politykę rozliczalności (*accountability*) (Kochan, 2007).

Pierwsze na świecie systemy ewaluacyjne wykorzystujące wskaźniki efektywności nauczania (metodę EWD) rozwinęły się właśnie w Stanach Zjednoczonych w połowie lat 80. XX w. w Karolinie Południowej oraz na obszarze okręgu szkolnego Dallas. W systemie ewaluacyjnym *Dallas Value-Added Accountability System* (DVAAS) wskaźniki służyły do tworzenia rankingów placówek, a na początku lat 90. stały się podstawą przyznawanych szkołom gratyfikacji finansowych. Początkowo wskaźniki były szacowane na podstawie prostej metodologii, którą z czasem rozwinęto i, co warte podkreślenia, w modelach wykorzystano szeroko rozumiane efekty kształcenia uwzględniające, poza wynikami testów osiągnięć, wskaźniki frekwencji, promocji do następnej klasy czy przedwczesnego kończenia nauki szkolnej. Kontrolowano także różnicowanie charakterystyk uczniów i środowiska lokalnego szkoły (Webster i Mendro, 1997).

Również w Karolinie Południowej przyznawano szkołom zachęty finansowe na podstawie wskaźników EWD. Miary te były wyznaczane na poziomie szkół, jednak w modelach nie uwzględniano dodatkowych zmiennych kontekstowych, takich jak status społeczno-ekonomiczny uczniów, co spotykało się z krytyką (Clotfelter i Ladd, 1996). W Karolinie Północnej natomiast na podstawie modeli EWD wyznaczano szkołom z wyprzedzeniem cel w postaci oczekiwanej wielkości przyrostu umiejętności uczniów (Ladd i Walsh, 2002). Szkołom, które uzyskały wyższy niż oczekiwany średni wzrost wyników, przyznawano nagrody finansowe (Goertz, 2005).

Najbardziej popularnym amerykańskim systemem ewaluacyjnym wykorzystującym wskaźniki EWD jest *Tennessee Value-Added Assessment System* (TVAAS) i jego komercyjny odpowiednik *Education Value-Added Assessment System*

(EVAAS), rozwijane od połowy lat 80. pod kierunkiem Williama Sandersa. W tym systemie wskaźniki początkowo publikowano na poziomie okręgów szkolnych, następnie szkół, a w końcu największą wagę zaczęto przykładać do wskaźników efektywności nauczycieli (Wright, Horn i Sanders, 1997). Cechą charakterystyczną modeli w ramach tego systemu jest uwzględnienie większej niż dwa liczby pomiarów osiągnięć szkolnych. Modele nie zawierają dodatkowych informacji o kontekście pracy szkoły, ich znaczenie jest kontrolowane pośrednio przez uwzględnienie historii wyników nauczania (McCaffrey, Lockwood, Koretz i Hamilton, 2003). System ten jest obecnie wykorzystywany w wielu stanach.

W Wielkiej Brytanii rozwój modeli oceny szkolnej efektywności nauczania związany był z rosnącą rolą podejścia quasi-rynkowego oraz traktowaniem rodziców i uczniów jako klientów systemu edukacyjnego. Szansę na wzrost jakości edukacji upatrywano w decentralizacji zarządzania i finansowania oświaty oraz urynkowieniu systemu edukacji przez umożliwienie rodzicom wyboru szkoły. Reformy doprowadziły do powstania systemu egzaminów zewnętrznych, których wyniki są publicznie dostępne (od 1992) i stały się podstawą obliczania wskaźników efektywności nauczania (Stoll i Sammons, 2007). Dodatkowo, na początku lat 90., badania SER i SIR zaczęły wpływać na decyzje polityczne. Na podstawie ich wyników stworzono i opublikowano „kluczowe charakterystyki efektywnych szkół”, a także zamówiono dodatkowe badania czynników efektywnego nauczania różnych grup uczniów (Sammons, Thomas, Mortimore, Owen i Pennell, 1994).

Prace nad metodą EWD na potrzeby angielskiego systemu ewaluacji rozpoczęły się w połowie lat 90. XX w., a na początku XXI w. wskaźniki zostały upublicznione. Początkowo modele były bardzo proste i zawierały tylko wyniki dwóch pomiarów osiągnięć. Wraz z rozwojem systemu zbierania danych oświatowych pojawiła się możliwość włączenia do modeli dodatkowych zmiennych. W 2006 roku opublikowano kontekstowe wskaźniki EWD zawierające wiele charakterystyk z poziomu ucznia i szkoły (Ray, McCormack i Evans, 2009). Modele te były krytykowane m.in. ze względu na ich zbyt złożoną dla odbiorców formę oraz brak zgodności postaci modelu z celami, którym mają służyć wskaźniki (uwzględnienie zmiennych opisujących kontekst szkoły) (Leckie i Goldstein, 2009). Dlatego w 2012 roku wrócono do prostszego modelu, kontrolującego jedynie uprzednie osiągnięcia szkolne.

Badania efektywności nauczania były początkowo prowadzone przede wszystkim w USA i Wielkiej Brytanii. W innych krajach europejskich do wczesnych lat 80. badania koncentrowały się na procesie nauczania, zachowaniu nauczycieli i tym, co się dzieje w klasie. Badania szkolnych uwarunkowań wyników nauczania rozpoczęły się w Europie kontynentalnej w latach 80. Pierwsze projekty były próbą zastosowania metodologii badań amerykańskich. Jednym z nich było holenderskie badanie przeprowadzone przez Vermeulena, w którym przetestowano znaczenie dla wyników nauczania pięciu cech efektywnej szkoły. Niestety okazało się, że zaadoptowane kwestionariusze nie były dobrze dostosowane do sytuacji Holandii, więc nie dostarczyły wiarygodnych miar. Dość szybko jednak badania holenderskie dorównały standardom krajów wiodących (USA, Wielka Brytania). W innych krajach pierwsze projekty badawcze w tradycji SER pojawiły się pod koniec lat 80. i w latach 90. (Creemers, 2007).

Na początku XXI w. metoda EWD zyskała większy rozgłos w krajach europejskich za sprawą prowadzonego przez OECD w latach 2005–2008 projektu *Project on the development value-added models in education systems* (OECD, 2008). Wzięło w nim udział 12 krajów europejskich (w tym Polska). Warto wspomnieć, że jednym z ekspertów projektu był Maciej Jakubowski, członek działającego już wtedy w Polsce zespołu zajmującego się rozwojem metody EWD. W większości krajów udział w projekcie nie przełożył się na wdrożenia. Poza Wielką Brytanią i Polską próby wdrożenia metody EWD podjęto na Słowacji (Kolková i Ivica, 2013) i Węgrzech (Balázsi, 2006).

## Badania efektywności nauczania w Polsce

Omawiając polski kontekst badań nad efektywnością nauczania, zarysuję nieco szersze tło, by lepiej pokazać uwarunkowania prowadzące do rozwoju tradycji badawczej *school effectiveness research* w naszym kraju. W polskich pracach można znaleźć trzy podejścia do badania efektywności nauczania. Pierwsze koncentruje się na ocenie wyników nauczania, drugie na kontroli procesu (a w szczególności metod) nauczania, a trzecie na badaniu „wkładu” szkoły lub nauczyciela w wyniki nauczania. Granice między tymi podejściami nie zawsze są ostro zarysowane, a osiągnięcia badań prowadzonych w jednym duchu wpływają na rozwój pozostałych podejść. Dlatego warto przedstawić ważniejsze fakty i prace wpisujące się w każdy z tych nurtów. Nie można pominąć także uwarunkowań politycznych i regulacji prawnych, które umożliwiły realizację projektów związanych z oceną efektywności nauczania. Kontekst badań nad efektywnością nauczania w Polsce zostanie omówiony w porządku chronologicznym.

Okoń (1998b) podaje, że pierwsze badania osiągnięć szkolnych w Polsce miały miejsce około stu lat temu. Prowadzili je Kazimierz Twardowski na początku XX wieku oraz Władysław Witwicki w latach 1925–1927. W 1928 roku Maria Grzywak-Kaczyńska przeprowadziła pierwsze, duże jak na tamte czasy, badanie osiągnięć szkolnych uczniów klas I–III warszawskich szkół podstawowych, a ich wyniki omówiła w kilku książkach (Grzywak-Kaczyńska, 1933, 1935). W wydanej już po wojnie książce, wykorzystującej wyniki tych samych badań, omówiła korzyści dla szkoły ze stosowania testów wiadomości (Grzywak-Kaczyńska, 1960). Zaprezentowała także wyniki analiz pokazujących międzyszkolne i międzyoddziałowe różnice poziomu osiągnięć szkolnych uczniów. Zwróciła uwagę na to, że są one uwarunkowane zarówno czynnikami szkolnymi, jak i niezależnymi od szkoły. Spostrzeżenia te mogłyby stać się przyczynkiem do rozwoju badań nad szkolną efektywnością nauczania, gdyż właśnie z takich obserwacji się one wywodzą. Jednak tak się nie stało.

W latach 1945–1950 Mieczysław Choynowski, mający niezwykle zasługi dla rozwoju polskiej psychometrii, podjął badania nad obiektywizacją metod selekcji kandydatów na studia wyższe (Manturzevska, 2002). W mniej więcej tym samym czasie badania osiągnięć prowadził Marian Falski (1949), a po roku 50. zrealizowano kilka projektów pod kierunkiem Okonia (Okoń, 1951, 1952), które doprowadziły do udziału Polski w 1960 roku w międzynarodowych badaniach osiągnięć szkolnych 13-latków *IEA Pilot Twelve-Country Study* (Foshay, Thorndike, Hotyat, Pidgeon i Walker, 1962).

Wydawało się więc, że rozwój teorii pomiaru w polskiej pedagogice jest na dobrej drodze. Niestety po 1968 roku spadło zainteresowanie naukowymi badaniami wyników kształcenia (Okoń, 1998b). W 1970 roku Polska, pod kierunkiem Okonia, uczestniczyła jeszcze w przygotowaniach i badaniach terenowych w ramach *IEA Six Subject Survey*, jednak władze polityczne nie wyraziły zgody na przekazanie wyników do międzynarodowego konsorcjum, zaprzeczając tym samym możliwość skorzystania z tego przedsięwzięcia.

W tym samym mniej więcej okresie prowadzone były badania nad skutecznością różnych metod nauczania. Zintensyfikowały się one po 1955 roku, kiedy to w związku z podjętą w kraju ostrą krytyką stosowanych zbyt często metod podających coraz częściej w dyskusjach naukowych pojawiała się problematyka efektywności kształcenia (Denek, 1980, s. 32). Kupisiewicz (1960) prowadził np. badania nad efektywnością metod nauczania przedmiotów przyrodniczych, wykorzystując metodę eksperymentu naturalnego i traktując przyrost wiadomości jako wskaźnik efektywności. We wrześniu 1968 roku odbyło się w Sopocie II Ogólnopolskie Sympozjum Ekonomiki Kształcenia, na którym podkreślano potrzebę podjęcia badań w zakresie efektywności procesu dydaktycznego. Metody takich badań rozwijał w swoich pracach m.in. Kazimierz Denek (Denek, 1968, 1971, 1980), również opierając się na eksperymencie naturalnym. Słabym punktem tego typu badań był brak precyzyjnego określenia celów i kryteriów efektywnego kształcenia, na co wskazywał sam autor.

Barbara Żechowska (1982) natomiast koncentrowała się na tendencjach metodologicznych w badaniach efektywności nauczycieli. Przedstawiła model badania efektywności nauczycieli, będący próbą uogólnienia podejść stosowanych przez różnych badaczy. Uwzględniła w nim wiele czynników z poziomu nauczyciela (psychologicznych, pedagogicznych i społecznych), interakcje między nauczycielem i uczniami, a jako wskaźnik efektywnego działania nauczyciela przyjęła stopień realizacji celów i zadań pedagogicznych, czyli zmiany w wiadomościach, umiejętnościach i postawach uczniów. Zwróciła również uwagę na odmienność podejść amerykańskiego i polskiego w badaniach efektywności nauczycieli w tym czasie. Amerykanie szukali czynników efektywności nauczyciela i sposobów ich mierzenia (rozwijająca się już w tym czasie tradycja badań *school effectiveness research*), polscy badacze starali się natomiast zrozumieć, od czego zależy efektywne działanie. Mimo że badania nad efektywnym działaniem lub skutecznością różnych metod dydaktycznych były prowadzone w innej tradycji niż badania skoncentrowane na „wkładzie” szkoły w wyniki kształcenia, to znaczącą ich zasługą jest zwrócenie uwagi na wartość naukowego poznania efektywności oddziaływań szkoły i nauczycieli.

Duże zasługi dla przejścia w Polsce od badań osiągnięć szkolnych do diagnozy edukacyjnej, a także dla rozwoju i zastosowania teorii pomiaru na gruncie pedagogiki ma Bolesław Niemierko, którego dorobek naukowy po 1975 roku jest poświęcony przede wszystkim tej problematyce. Nie sposób przytoczyć tu wszystkich przeprowadzonych przez niego badań i opublikowanych prac. Ich zwięzłe podsumowanie można znaleźć w referacie Okonia (1998a). Należy jednak odnotować, że Niemierko zgromadził wokół siebie środowisko naukowe zajmujące się pomiarem dydaktycznym i stworzył podwaliny pod powstanie

diagnostyki edukacyjnej – nowej w Polsce subdyscypliny naukowej, zajmującej się przebiegiem i wynikami uczenia się (Niemierko, 1990). W ogólnym modelu uczenia się, z którego korzystał, uwzględnił kontekst, wejście, proces i produkt oraz relacje między tymi pojęciami, co koresponduje z założeniami modeli oceny efektywności nauczania. Z inicjatywy Niemierki od 1993 roku organizowane są co roku Konferencje Diagnostyki Edukacyjnej umożliwiające wymianę myśli naukowej, rozwój dziedziny, a także dialog między środowiskiem naukowym i praktykami systemu oświaty, a od 2004 roku istnieje Polskie Towarzystwo Diagnostyki Edukacyjnej, którego głównym celem jest wspomaganie rozwoju wiedzy naukowej o wewnątrzszkolnym i zewnętrznym systemie oceniania, uwarunkowaniach osiągnięć uczniów i procesie kształcenia, aby ulepszyć nauczanie.

Niestety badania z zakresu diagnostyki edukacyjnej w niewielkim stopniu przebiegały się w naukowych kręgach pedagogicznych. Krzysztof Kruszewski, diagnozując ten stan rzeczy, stwierdził, że prace te były zbyt hermetyczne, jawiły się jako niepotrzebne i nieużyteczne oraz dodał: „Badania te i towarzysząca im teoria są zbyt trudne dla przeciętnie wykształconego badacza i praktyka pedagogicznego, by mogli je przyjąć lub stoczyć polemikę. Wolą więc nie zauważać” (Kruszewski, 2005, s. 131). Przyczyniło się do tego zapewne odłączenie pedagogiki od psychologii i psychometrii oraz zbyt zredukowane kształcenie metodologiczne i statystyczne studentów pedagogiki.

Zainteresowanie pomiarem jakości nauczania nie było tylko domeną naukowców. W latach 1995–1997 zrealizowano w Polsce program TERM-IAE<sup>5</sup> kierowany do dyrektorów szkół, wizytatorów, przedstawicieli placówek doskonalenia nauczycieli oraz pracowników samorządowych związanych z edukacją. Celem modułu IV „Mierzenie jakości pracy szkoły” było dostarczenie szkole pełnej, rzetelnej informacji o jej mocnych i słabych stronach, efektach działalności w porównaniu z innymi szkołami podobnego typu, w podobnym środowisku społeczno-ekonomicznym (Bednarek i Lempa, 2005). Przeprowadzono wiele szkoleń, wydano liczne publikacje pokazujące założenia programu i przykładowe narzędzia pomiaru, głównie ankiety badające opinie. Autorzy polskich materiałów podkreślali jednak, że mierzenie jakości szkoły, zwłaszcza osiągniętych efektów, jest trudne z powodu braku standardów edukacyjnych, kryteriów oceny szkoły, systemu egzaminów państwowych i standaryzowanych testów osiągnięć (Dzierzgowska i Wlazło, 2000). Program ten był raczej próbą zastosowania w edukacji koncepcji zarządzania Total Quality Management (TQM). Nie przywiązywano w nim zbyt dużej wagi do metodologii badań, nie zadbano również o to, by wypracować sposób mierzenia efektywności nauczania. W konsekwencji działania sprowadziły się do stosowania licznych ankiet i pomiar jakości pracy szkoły zaczął być utożsamiany z badaniem ankietowym.

Prawdziwym punktem zwrotnym w rozwoju metod pomiaru szkolnej efektywności nauczania było powstanie w Polsce systemu egzaminów zewnętrznych. Jego zaczątkiem był prowadzony w latach 90. w kilku województwach eksperyment wałbrzyski. Polegał on na zastąpieniu lokalnych egzaminów wstępnych do szkół ponadpodstawowych jednakowymi standaryzowanymi testami osiągnięć

---

<sup>5</sup> Był to program finansowany z programu Phare mającego na celu pomoc materialną państwom kandydującym do Wspólnot Europejskich.



przeprowadzanymi na zakończenie – wtedy ósmioletniej – szkoły podstawowej. W tym samym czasie rozpoczęły się w Kwidzynie, inicjowane przez władze miasta, badania monitorujące osiągnięcia szkolne uczniów. W 1997 roku opublikowano książkę pokazującą wartość wykorzystywania monitorowania osiągnięć szkolnych na poziomie szkoły, lokalnym i ogólnokrajowym (Dolata, Murawska, Putkiewicz i Żytko, 1997). Wykorzystywała ona wyniki pierwszej diagnozy w Kwidzynie i zrealizowanego w 1995 roku ogólnopolskiego badania efektywności edukacji początkowej. Już w tej publikacji pokazano przykład prostych kontekstowych modeli oceny efektywności nauczania zbudowanych na podstawie informacji o wykształceniu rodziców oraz wyników testów osiągnięć szkolnych. Badania w Kwidzynie były kontynuowane w kolejnych latach, obejmując różne poziomy edukacji (Dolata, Murawska i Putkiewicz, 2001).

Na powstanie systemu egzaminów zewnętrznych, oprócz inicjatyw naukowych, miały oczywiście wpływ także kolejne reformy systemu oświaty. Zmiany po przełomie 1989 roku prowadziły do decentralizacji zarządzania i finansowania szkolnictwa. Stopniowo zwiększano autonomię szkół i nauczycieli, także w zakresie wyboru programu nauczania czy podręczników, a prowadzenie szkół powierzano jednostkom samorządu terytorialnego. Po zwiększeniu różnorodności programów nauczania ustala możliwość oceniania pracy nauczyciela i szkoły na podstawie realizacji programu. Konieczne stało się przeniesienie uwagi z tego, czy zrealizowano program, na wyniki nauczania (Sławiński, 1996). W 1998 roku resort edukacji pod kierownictwem Mirosława Handkego, nowelizując ustawę o systemie oświaty, rozpoczął wprowadzanie nowego ustroju szkolnego. Natomiast reforma administracyjna, która weszła w życie z początkiem 1999 roku, zmieniła sposób finansowania oświaty i sprawowania nad nią nadzoru. Samorządy terytorialne są odpowiedzialne za prowadzenie szkół, a do obowiązków państwa należą zapewnienia im środków na ten cel (w formie oświatowej subwencji ogólnej). Państwo wyznacza także ogólne cele i treści kształcenia opisane w podstawach programowych i standardach wymagań w zakresie kształcenia ogólnego.

Zmiany te zrodziły potrzebę wprowadzenia zewnętrznego pomiaru wyników nauczania, który miał zapobiec zbyt daleko idącemu różnicowaniu się polskiej oświaty (Dolata i Szaleniec, 2012). Na początku 1999 roku powołano Centralną Komisję Egzaminacyjną i okręgowe komisje egzaminacyjne, których głównym zadaniem było zbudowanie systemu egzaminów zewnętrznych. Było to bez wątpienia kluczowe wydarzenie dla zmieniającego się w Polsce sposobu myślenia o nadzorze w oświacie, polegającego na przesunięciu zainteresowania z treści i procesu nauczania na wyniki kształcenia. W zreformowanym systemie oświaty pojawiły się dwa mechanizmy kontrolne: nadzór sprawowany przez kuratoria polegający na zewnętrznej ewaluacji działalności edukacyjnej szkół i placówek oświatowych oraz pomiar efektów pracy szkoły z wykorzystaniem egzaminów zewnętrznych.

Kolejne rozporządzenia MEN zmieniały zasady sprawowania nadzoru pedagogicznego. Rozporządzenie z dnia 13 sierpnia 1999 roku wprowadziło do zadań kuratoriów oświaty mierzenie jakości pracy szkół, będące nawiązaniem do programu TERM. Podstawowym sposobem sprawowania nadzoru pedagogicznego stało się zatem zorganizowane i systematyczne analizowanie oraz ocenianie

stopnia spełniania przez szkołę wymagań wynikających z jej zadań, z uwzględnieniem opinii uczniów, rodziców (Dz.U. 1999 nr 67 poz. 759). W 2003 roku rozporządzenie uzupełniono o standardy oceny jakości pracy szkół i placówek. Niestety podejmowane działania były postrzegane przez szkoły jako ankietomania i bezcelowy biurokratyzm (Bednarek i Lempa, 2005).

W 2008 roku rozpoczął się projekt „Program wzmocnienia efektywności systemu nadzoru pedagogicznego i oceny jakości pracy szkoły”<sup>6</sup> (Nadzór Pedagogiczny – System Ewaluacji Oświaty; NPSEO), który miał na celu zmodernizowanie nadzoru pedagogicznego i oceny jakości pracy szkoły. W *Rozporządzeniu MEN z dnia 7 października 2009 roku w sprawie nadzoru pedagogicznego* (Dz.U. nr 168 poz. 1324) sformułowano wymagania państwa wobec szkół i placówek, wyznaczające cele ich działań. Nowy model nadzoru pedagogicznego wyróżnił trzy formy jego sprawowania: ewaluację działalności edukacyjnej szkół i placówek, kontrolę przestrzegania prawa oraz wspomaganie pracy szkół i placówek oraz nauczycieli w zakresie ich działalności statutowej. Wspieranie wdrożenia nowego modelu nadzoru pedagogicznego było możliwe dzięki kolejnym dwóm etapom projektu NPSEO. Jednym z jego celów było stworzenie modelu badania jakości pracy szkoły przez operacjonalizację wymagań państwa wobec szkół oraz zbudowanie odpowiednich wskaźników, procedur i narzędzi badania<sup>7</sup>. Niestety kolejny raz sprowadziło się to głównie do zbierania opinii uczestników systemu edukacji za pomocą ankiet i kwestionariuszy wywiadów. W ramach tych działań nie zbudowano wskaźników wykorzystujących wyniki egzaminów zewnętrznych, choć odwoływano się do istniejących już wskaźników (Matuszczak i Wasilewska, 2015).

Dane z systemu egzaminacyjnego przyczyniły się jednak do gwałtownego rozwoju badań edukacyjnych i umożliwiły rozwój metodologii szacowania modeli edukacyjnej wartości dodanej (EWD) w Polsce. Wykorzystanie tej metody stało się możliwe w 2005 roku. Wtedy gimnazja ukończył pierwszy rocznik absolwentów, dla których dysponowaliśmy wynikami egzaminu gimnazjalnego oraz sprawdzianu w VI klasie. W 2006 roku ukazały się pierwsze polskie publikacje pokazujące możliwości metody EWD (Dolata, 2006a, 2006b). Najważniejszym osiągnięciem tego okresu było uświadomienie decydentom politycznym możliwości tkwiących w podłużnie analizowanych wynikach egzaminacyjnych. Dzięki temu powołano przy CKE zespół badawczy ds. rozwoju metody EWD, kierowany przez Romana Dolatę (projekt realizowany w latach 2005–2006 w ramach EFS). Opracowano wtedy pierwsze modele EWD dla gimnazjów oraz specjalny arkusz kalkulacyjny (Kalkulator EWD) umożliwiający szkołom prowadzenie analiz wskaźników efektywności nauczania w ramach ewaluacji wewnętrznej. Przeprowadzono także szkolenia użytkowników metody.

W latach 2007–2013 zrealizowano projekt „Badania dotyczące rozwoju metodologii szacowania wskaźnika edukacyjnej wartości dodanej”, w ramach którego wypracowano m.in. doskonalsze sposoby szacowania wskaźników, wykorzystujące wielopoziomowe modele mieszanych efektów (Raudenbush

<sup>6</sup> Był to projekt systemowy III Priorytetu Programu Operacyjnego Kapitał Ludzki „Wysoka jakość systemu oświaty”, realizowany w trzech etapach w latach 2008–2015 przez Ośrodek Rozwoju Edukacji w partnerstwie z Uniwersytetem Jagiellońskim i Erą Ewaluacji.

<sup>7</sup> Informacje o projekcie można znaleźć na stronie: [www.nadzorpedagogiczny.edu.pl](http://www.nadzorpedagogiczny.edu.pl).

i Bryk, 2002) oraz skalowane metodą IRT wyniki egzaminacyjne (De Ayala, 2009). W 2007 roku ukazała się pierwsza polska monografia naukowa dotycząca metody EWD (Dolata, Jakubowski, Pokropek, Rappe i Stożek, 2007). Co roku liczone modele EWD zawierające dane jednej kohorty absolwentów (tzw. jednoroczne) na potrzeby ewaluacji wewnętrznej, a także od 2009 roku zaczęto publikować na stronie projektu wskaźniki z modeli zawierających dane z trzech kolejnych sesji egzaminacyjnych (tzw. trzyletnie) – najpierw dla gimnazjów, a później dla liceów i techników. W 2009 roku rozpoczęła się realizacja trzech badań podłużnych, mających dostarczyć informacji potrzebnych do rozwoju metody, w tym oceny jej trafności. Badanie w szkołach pogimnazjalnych, na zlecenie CKE, zostało zrealizowane przez IFiS PAN (Karwowski, 2013). Badania w szkołach podstawowych i gimnazjach były realizowane przez zespół EWD (Dolata i in., 2013, 2014, 2015). Przeprowadzono również wiele działań szkoleniowych i upowszechniających, które umożliwiały konfrontację wypracowanej metodologii oceny szkolnej efektywności nauczania z opinią praktyków.

W latach 2013–2015 kontynuowano prace nad metodą EWD w ramach projektu „Rozwój metody edukacyjnej wartości dodanej na potrzeby wzmocnienia ewaluacyjnej funkcji egzaminów zewnętrznych”<sup>8</sup>. Przeprowadzono kolejne etapy badania podłużnego w szkołach podstawowych (Dolata i in., 2015) oraz badania wykorzystania metody EWD w ewaluacji nauczania (Jasińska, Humenny i Humenny, 2013; Kędracka, Matuszczak, Rappe i Stożek, 2013; Mischczuk i Karmiński, 2015). Opracowano także modele EWD dla szkół podstawowych, które uczestniczyły w Ogólnopolskim Badaniu Umiejętności Trzecioklasistów.

Program Operacyjny Kapitał Ludzki umożliwił prowadzenie także innych ważnych projektów naukowych poświęconych badaniu efektywności kształcenia. W latach 2010–2015 został w Polsce zrealizowany przez Instytut Badań Edukacyjnych projekt systemowy „Badanie jakości i efektywności edukacji oraz instytucjonalizacja zaplecza badawczego” (Entuzjaści Edukacji)<sup>9</sup>. Jednym z prowadzonych w jego ramach projektów było, realizowane pod naukową opieką Romana Dolaty, ogólnopolskie badanie podłużne „Szkolnych uwarunkowań efektywności kształcenia” (SUEK), mające na celu poznanie szkolnych i pozaszkolnych determinantów wyników i efektywności nauczania w szkołach podstawowych. W badaniu tym, podobnie jak w projekcie EWD, efektywność nauczania była definiowana i badana jako „wkład” szkoły w osiągnięcia szkolne uczniów (Dolata, 2014).

Na rozwój metodologii badania szkolnej efektywności nauczania w Polsce miały także wpływ międzynarodowe projekty badania osiągnięć szkolnych, w których Polska w ostatnich latach regularnie uczestniczy. Od 2000 roku bierze udział co trzy lata w badaniach PISA organizowanych przez OECD, a od 2006 i od 2011 w badaniach PIRLS i TIMSS realizowanych co pięć lat przez IEA. Stosuje

<sup>8</sup> Oba projekty (2007–2013, 2013–2015) były realizowane początkowo w CKE, później w IBE, ze środków Europejskiego Funduszu Społecznego w ramach działania 3.2. Rozwój systemu egzaminów zewnętrznych priorytetu III. Wysoka jakość systemu oświaty Programu Operacyjnego Kapitał Ludzki.

<sup>9</sup> Zrealizowany ze środków Europejskiego Funduszu Społecznego w ramach Programu Operacyjnego Kapitał Ludzki, Priorytet III: Wysoka jakość systemu oświaty, Poddziałanie 3.1.1. Tworzenie warunków i narzędzi do monitorowania, ewaluacji i badań systemu oświaty.

się w nich podobną strategię badania efektywności (uwzględnienie cech kontekstowych w analizie wyników testów), jednak pytania dotyczą efektywności systemów edukacyjnych. Co być może ważniejsze, projekty te wyznaczają również standardy metodologiczne i analityczne edukacyjnych badań ilościowych.

## Uwagi końcowe

Artykuł przedstawia wydarzenia i prace najważniejsze do rozwoju metod oceniania efektywności nauczania w Polsce i na świecie. W ostatnich latach obserwowaliśmy w Polsce dynamiczny rozwój naukowych badań wyników i efektywności nauczania. Być może należałoby stwierdzić bez fałszywej skromności, że w pewnych aspektach – jak na przykład implementacja metody EWD w krajowym systemie oświaty – dorównaliśmy w tym czasie krajom wiodącym.

Opisane modele oceny efektywności nauczania skoncentrowane na „wkładzie” szkoły w osiągnięcia uczniów są charakterystyczne dla pomiaru różnicującego – pozwalają interpretować wynik na tle określonej grupy odniesienia. Bez wątplenia wartościowe z naukowego i praktycznego punktu widzenia, choć także niezwykle trudne, byłyby badania nad możliwością zbudowania metod oceny szkolnej efektywności nauczania nawiązujących do pomiaru sprawdzającego (Glaser, 1963; Niemierko, 1990), których wyniki mogłyby być interpretowane w odniesieniu do ustalonych kryteriów efektywnego nauczania.

Warto także zaznaczyć, że tekst ten dotyczy metod oceny efektywności nauczania, więc z konieczności dominuje w nim koncentracja na szkole, która jest jednostką analizy w tak zdefiniowanym problemie badawczym. Może to pozostawiać mylne wrażenie, że diagnoza prowadzona na poziomie indywidualnym jest mniej ważna. Dlatego na zakończenie warto podkreślić, że badania skoncentrowane na uczniu, szkole (lub klasie) czy systemie oświaty powinny być traktowane jako komplementarne, bo każda z tych perspektyw wnosi swą istotną wiedzę na temat uwarunkowań wyników nauczania i uczenia się.

## Bibliografia

- Balázi, I., National Assessment of Basic Competencies in Hungary. Center for Evaluation Studies, 2006, [www.iaea.info/documents/paper\\_1162a1d92d.pdf](http://www.iaea.info/documents/paper_1162a1d92d.pdf) [dostęp: 30.06.2017].
- Bednarek, K. i Lempa, C., *Utracone/Utracone mierzenie* [w:] Niemierko, B. i Szyling, G. (red.), *Holistyczne i analityczne metody diagnostyki edukacyjnej. Perspektywy informatyczne egzaminów szkolnych*, Fundacja Rozwoju UG, Gdańsk 2005, s. 197–204.
- Brookover, W.B., Beady, C., Flood, P., Schweitzer, J. i Wisenbaker, J., *School social systems and student achievement: schools can make a difference*, Praeger, New York 1979.
- Clotfelter, C.T. i Ladd, H.F., *Recognizing and rewarding success in public schools* [w:] Ladd, H.F. (red.), *Holding schools accountable: performance-based reform in education*, Brookings Institution, Washington D.C 1996.
- Coleman, J.S., Campbell, E.Q., Hobson, C.J., Mc Partland, J., Mood, A., Weinfeld, F.D. i York, R.L., *Equality of Educational Opportunity*, U. S. Government Printing Office, Washington, D. C 1966.

- Creemers, B.P.M., *Educational Effectiveness and Improvement: The Development of the Field in Mainland Europe* [w:] Townsend T. (red.), *International Handbook of School Effectiveness and Improvement*, Springer, Netherlands 2007, s. 223–242.
- De Ayala, R.J., *The theory and practice of item response theory*, Guilford Press, New York 2009.
- Denek, K., *Kryteria i metody oceny efektywności nauczania programowanego*. Zaprezentowano na II Ogólnopolskim Sympozjum Ekonomiki Kształcenia, Sopot 1968.
- Denek, K., *Efektywność nauczania programowanego w szkole wyższej*, Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza, Poznań 1971.
- Denek, K., *Pomiar efektywności kształcenia w szkole wyższej*, PWN, Warszawa 1980
- Dolata, R., *Edukacyjna wartość dodana w komunikowaniu wyników egzaminów zewnętrznych*, „Egzamin. Biuletyn Badawczy CKE” 2006a nr 8, s. 9–20.
- Dolata, R., *Efektywność nauczania w gimnazjach miasta X. Analiza edukacyjnej wartości dodanej*, „Egzamin. Biuletyn Badawczy CKE” 2006b nr 8, s. 28–37.
- Dolata, R., Hawrot, A., Humenny, G., Jasińska, A., Koniewski, M., Majkut, P. i Żółtak, T., *Trafność metody edukacyjnej wartości dodanej dla gimnazjów*, Instytut Badań Edukacyjnych, Warszawa 2013.
- Dolata, R., Hawrot, A., Humenny, G., Jasińska-Maciążek, A., Koniewski, M. i Majkut, P., *Kontekstowy model efektywności nauczania po pierwszym etapie edukacyjnym*. Instytut Badań Edukacyjnych, Warszawa 2014.
- Dolata, R., Hawrot, A., Humenny, G., Jasińska-Maciążek, A., Koniewski, M., Majkut, P., Blukacz, M., Grygiel, P., Otręba-Szklarczyk, A., *(Ko)warianty efektywności nauczania: wyniki badania w klasach IV–VI*, Instytut Badań Edukacyjnych, Warszawa 2015.
- Dolata, R., Jakubowski, M., Pokropek, A., Rappe, A. i Stożek, E., *Edukacyjna wartość dodana jako metoda oceny efektywności nauczania: na podstawie wyników egzaminów zewnętrznych*, (red. Dolata, R.), Centralna Komisja Egzaminacyjna, Warszawa 2007.
- Dolata, R., Murawska, B. i Putkiewicz, E., *Monitorowanie osiągnięć szkolnych jako metoda wspierania lokalnego środowiska edukacyjnego*, Żak, Warszawa 2001.
- Dolata, R., Murawska, B., Putkiewicz, E. i Żytko, M., *Monitorowanie osiągnięć szkolnych jako metoda doskonalenia edukacji: zarys metody oraz przykłady zastosowań w edukacji początkowej*, Żak, Warszawa 1997.
- Dolata, R. i Szaleniec, H., *Funkcje krajowych egzaminów w systemie edukacji*. „Polityka Społeczna” 2012, nr tematyczny 1, s. 37–41.
- Dzierzgowska, I. i Wlazło, S., *Mierzenie jakości pracy szkoły podstawowej: poradnik*, Wydawnictwo Centralnego Ośrodka Doskonalenia Nauczycieli, Warszawa 2000.
- EACEA, *Ogólnokrajowe egzaminowanie uczniów w Europie: cele, organizacja i wykorzystanie wyników* (tłum. M. Kowalczyk), Fundacja Rozwoju Systemu Edukacji, Warszawa 2010.
- Edmonds, R., *Effective Schools for the Urban Poor*, “Educational Leadership” 1979, nr 37(1), s. 15–27.
- Falski, M., *Uwagi wstępne do badań wyników nauczania*, „Nowa Szkoła” 1949, nr 7.
- Fenstermacher, G.D. i Soltis, J. F., *Style nauczania* (tłum. Kruszewski, K.), Wydawnictwa Szkolne i Pedagogiczne, Warszawa 1998.
- Foshay, A.W., Thorndike, R.L., Hotyat, F., Pidgeon, D.A. i Walker, D.A., *Educational achievement of thirteen-year-olds in twelve countries*, UNESCO Institute for Education, Hamburg 1962.
- Gardner, D.P. i in., *A Nation At Risk: The Imperative For Educational Reform. An Open Letter to the American People. A Report to the Nation and the Secretary of Education*, 1983.
- Glaser, R., *Instructional Technology and the Measurement of Learning Outcomes: Some Questions*, “American Psychologist” 1963, nr 18, s. 519–521.
- Goertz, M.E., *Implementing the No Child Left Behind Act: Challenges for the States*, “Peabody Journal of Education” 2005, nr 80(2), 73–89.

- Grzywak-Kaczyńska, M., *Testy i normy dla użytku szkół powszechnych*, Nasza Księgarnia, Warszawa 1933.
- Grzywak-Kaczyńska, M., *Powodzenie szkolne a inteligencja*, Nasza Księgarnia, Warszawa 1935.
- Grzywak-Kaczyńska, M., *Testy w szkole*, Państwowe Zakłady Wydawnictw Szkolnych, Warszawa 1960.
- Guthrie, J.W. i Springer, M.G., *A Nation at Risk Revisited: Did „Wrong” Reasoning Result in „Right” Results? At What Cost?*, „Peabody Journal of Education” 2004, nr 79(1), s. 7–35.
- Hanushek, E. A., *Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data*, „American Economic Review” 1971.
- Heckman, J. i Neal, D., *Coleman’s Contributions to Education: Theory, Research Styles and Empirical Research* [w:] Clark, J. (red.), *James S. Coleman* Routledge, New York 1996, s. 81–102.
- Jasińska, A., Humenny, G. i Humenny, L., *Wykorzystanie przez szkoły metody edukacyjnej wartości dodanej. Wyniki badania ilościowego*, Instytut Badań Edukacyjnych, Warszawa 2013.
- Jenks, C., *Inequality: A Reassessment of the Effect of Family and Schooling in America*, Basic Books, New York 1972.
- Kędracka, E., Matuszczak, K., Rappe, A. i Stożek, E., *Badanie na temat wykorzystania edukacyjnej wartości dodanej (EWD) przez szkoły ponadgimnazjalne*, Instytut Badań Edukacyjnych, Warszawa 2013.
- Kochan, S., *Evolution of School Performance Research in the USA: From School Effectiveness to School Accountability and Back* [w:] Townsend, T. (red.), *International Handbook of School Effectiveness and Improvement*, Springer, Netherlands 2007, s. 485–502.
- Kolková, M. i Ivica, L., *Čo odhaduje tzv. pridaná hodnota školy?* [w:] Pichaničová, I., Pokorný, M. i Ruman, J. (red.), *Hodnotenie kvality vzdelávania – súčasny stav a perspektívy. Zborník príspevkov z konferencie*, Národný ústav certifikovaných meraní vzdelávania, Bratislava 2013.
- Kruszewski, K., *Test wiadomości dobrego i złego dziesięć lat później* [w:] Groenwald, M., Szyling G. i Daszkiewicz M. (red.), *Diagnostyka edukacyjna. Niemierkowskie spotkania i inspiracje*, Fundacja Rozwoju Uniwersytetu Gdańskiego, Gdańsk 2005, s. 125–134.
- Kupisiewicz, C., *O efektywności nauczania problemowego: z badań nad metodami nauczania przedmiotów matematyczno-przyrodniczych*, PWN, Warszawa 1960.
- Kupisiewicz, C., *Nauczanie programowane*, Państwowe Zakłady Wydawnictw Szkolnych, Warszawa 1966.
- Ladd, H.F. i Walsh, R.P., *Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right*, „Economics of Education Review” 2002, nr 21(1), s. 1–17.
- Leckie, G. i Goldstein, H., *The Limitations of Using School League Tables to Inform School Choice*, „Journal of the Royal Statistical Society. Series A (Statistics in Society)” 2009, nr 172(4), s. 835–851.
- Manturzevska, M., *Mieczysław Choynowski – twórca polskiej psychometrii 1 XI 1909–1 X 2001*, „Przegląd Psychologiczny” 2002, nr 45(3), s. 275–288.
- Matuszczak, K. i Wasilewska, O., *Wyniki egzaminów zewnętrznych w pracy szkoły* [w:] *Raport o stanie edukacji 2014. Egzaminy zewnętrzne w polityce i praktyce edukacyjnej*, Instytut Badań Edukacyjnych, Warszawa 2015.
- McCaffrey, D.F., Lockwood, J.R., Koretz, D.M. i Hamilton, L.S., *Evaluating value-added models for teacher accountability*, CA: RAND, Santa Monica 2003.
- Miszczuk, E. i Karmiński, M., *Wykorzystanie przez szkoły metody edukacyjnej wartości dodanej. Wyniki badania ilościowego*, Instytut Badań Edukacyjnych, Warszawa 2015.
- Murnane, R.J., *Impact of School Resources on the Learning of Inner City Children*. Harper Collins Distribution Services, Cambridge 1975.

- Niemierko, B., *Pomiar sprawdzający w dydaktyce*, PWN, Warszawa 1990.
- OECD, *Measuring improvements in learning outcomes: best practices to assess the value-added of schools*, Organisation for Economic Co-operation and Development, Paris 2008.
- Okoń, W. (red.), *Badanie wyników nauczania w szkołach ogólnokształcących*, Nasza Księgarnia, Warszawa 1951.
- Okoń, W. (red.), *Wyniki nauczania w szkołach ogólnokształcących*, Państwowe Zakłady Wydawnictw Szkolnych, Warszawa 1952.
- Okoń, W., *Wielostronne uczenie się a problem aktywności uczniów*, „Nowa Szkoła” 1965, nr 7/8, s. 23–29.
- Okoń, W., *Od badania osiągnięć do diagnozy edukacyjnej* [w:] Niemierko, B. i Kowalik, E. (red.), *Perspektywy diagnostyki edukacyjnej*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk 1998a, s. 280–285.
- Okoń, W., *Wprowadzenie do dydaktyki ogólnej* (wyd. 4), Wydawnictwo Akademickie „Żak”, Warszawa 1998b.
- Raudenbush, S.W. i Bryk, A.S., *Hierarchical Linear Models: Applications and Data Analysis Methods*, SAGE, US 2002.
- Ray, A., McCormack, T. i Evans, H., *Value Added in English Schools*, “Education Finance and Policy” 2009, nr 4(4), s. 415–438.
- Rutter, M., Maughan, B., Mortimore, P. i Ouston, J., *Fifteen thousand hours: secondary schools and their effects on children*, Harvard University Press, Cambridge, Mas. 1979.
- Sammons, P., Thomas, S.M., Mortimore, P., Owen, C. i Pennell, H., *Assessing School Effectiveness: Developing Measures to put School Performance in Context*, Office for Standards in Education, 1994, s. 87.
- Sitek, M., *Miejsce egzaminów zewnętrznych w systemach edukacyjnych i polityce edukacyjnej* [w:] Dolata, R. i Sitek, M. (red.), *Raport o stanie edukacji 2014. Egzamininy zewnętrzne w polityce i praktyce edukacyjnej*, Instytut Badań Edukacyjnych, Warszawa 2015.
- Skinner, B.F., *The Technology of Teaching*, Copley Publishing Group, Acton, Mas. 1968.
- Sławiński, S., *Reforma szkolna w III Rzeczypospolitej*, Wydawnictwa Szkolne i Pedagogiczne, Warszawa 1996.
- Stoll, L. i Sammons, P., *Growing Together: School Effectiveness and School Improvement in the UK* [w:] Townsend, T. (red.), *International Handbook of School Effectiveness and Improvement*, Springer, Netherlands 2007, s. 207–222.
- Teddle, C. i Stringfield, S., *A History of School Effectiveness and Improvement Research in the USA Focusing on the Past Quarter Century* [w:] Townsend, T. (red.), *International Handbook of School Effectiveness and Improvement*, Springer, Netherlands 2007, s. 131–166.
- Webster, W.J. i Mendro, R.L., *The Dallas Value-Added Accountability System* [w:] Millman J. (red.), *Grading teachers, grading schools: is student achievement a valid evaluation measure?*, Corwin Press, Thousand Oaks, California 1997, s. 81–99.
- Wright, P.S., Horn, S.P. i Sanders, W.L., *Teacher and classroom context effects on student achievement: Implications for teacher evaluation*, “Journal of Personnel Evaluation in Education” 1997, nr 11, 57–67.
- Żechowska, B., *Efektywność pracy nauczyciela: wyznaczniki, tendencje, problemy: studium porównawcze*, Wydawnictwo Uniwersytetu Śląskiego, Katowice 1982.