

Ocena rozprawy doktorskiej mgr Aleksandry Jasińskiej-Maciążek***Modele oceny szkolnej efektywności nauczania*****Treść rozprawy**

Na rozprawę doktorską mgr Aleksandry Jasińskiej-Maciążek składa się 13 publikacji: 8 zbiorowych z jej znacznym (od 30% do 60%) potwierdzonym udziałem i 5 samodzielnych. Większość publikacji bezpośrednio dotyczy modelowania analiz wyników egzaminu zewnętrznego na różnych szczeblach systemu egzaminacyjnego w ujęciach podłużnych, jako edukacyjnej wartości dodanej (EWD). Pozostałe są także pedagogicznym zastosowaniem **metod statystycznych**, co nadaje całości dorobku cechę niezbędnej spójności.

Analizy statystyczne wymagają dużych prób reprezentujących dobrze zdefiniowane populacje. Autorka rozprawy uzyskała dostęp do zbiorów danych uzyskanych w ponad 150 wylosowanych szkołach podstawowych i gimnazjach, a w każdym z nich ponad 5 tysięcy uczniów. Dane te poddała zaawansowanym analizom korelacyjnym: regresyjnej i (w przypadku badań gotowości szkolnej) czynnikowej. Osiągnięcia uczniów były interpretowane według klasycznej teorii testu i probabilistycznej teorii wyniku zadania (IRT). W zespołach badawczych prowadzących odpowiednie badania odgrywała rolę głównego specjalisty pomiarowego, co jednoznacznie wskazuje na jej wysoki poziom kompetencji metodologicznej.

Tytułowymi „modelami” są formuły statystyczne. Mgr Jasińska preferuje **metody ilościowe w edukacji**: porównuje liczby, a nie obrazy, tworzy klasy ilościowe, a nie typologie¹. Charakterystyczne są wyrażenia „liczenie modelu” (tekst 9, s. 38 i in.) zamiast jego „stosowania” oraz „liczenie rzetelności” (13, s. 46) zamiast jej „szacowania”. Większość raportów z badań raczej sucho przedstawia procedury i szczegółowe wyniki obliczeń, choć zdarzają się i takie, które dobrze je tłumaczą słabiej przygotowanemu odbiorcy (12, artykuł do czasopisma *Edukacja*).

Głównym problemem badań jest wpływ szkoły na osiągnięcia poznawcze uczniów mierzone ich **wartością dodaną** między kolejnymi egzaminami zewnętrznymi. Ponieważ te egzaminy obejmują odmienne zakresy programowe, trzeba było udowodnić, że o ich wynikach decyduje **zmienna ukryta** w postaci ogólnego poziomu rozwoju poznawczego. Jej istnienie potwierdzają analizy trafności kryterialnej wyników egzaminu gimnazjalnego: korelacje +0,64 z ilorazami inteligencji (test Ravena), +0,53 z nauczycielskimi ocenami a matematyki, +0,53 z motywacją do uczenia się (1, s. 74n). Należało też

¹ Wyjątkiem jest typologia gimnazjów: „szkoły sukcesu”, „szkoły wymagające pomocy”, „szkoły wspierające” i „szkoły niewykorzystanych możliwości” (13, s. 45, rys. 5). Dwie ostatnie kategorie są mocno deficytowe ze względu na umiarkowanie wysoką korelację wyników egzaminacyjnych z wartością dodaną przez szkołę.

udowodnić tożsamość tej zmiennej w kolejnych etapach edukacji szkolnej. Próba wypadła pomyślnie, bo współczynniki korelacji uczniowskich wyników międzyprzedmiotowego sprawdzianu po szkole podstawowej i wyników egzaminu gimnazjalnego wyniosły +0,78 dla części matematyczno-przyrodniczej i +0,77 dla części humanistycznej (1, s. 82). Autorka zdaje sobie jednak sprawę z tego, że lepiej byłoby dysponować dwoma wynikami tych samych lub empirycznie zrównanych testów, co pozwoliłoby na oszacowania wartości dodanej bezwzględnej w postaci przyrostu określonych umiejętności uczniów, a nie zmiany ich pozycji w rozkładzie wyników.

Przyrosty osiągnięć uczniów są dodatnio skorelowane z poziomem tych osiągnięć, a współczynnik korelacji tych dwu wskaźników może sięgać wartości +0,65 (8, s. 85). Uczniowie zdolniejsi i bardziej umotywowani nie tylko potrafią więcej, lecz także – statystycznie biorąc – rozwijają się szybciej. To zjawisko nazywane jest „efektem wachlarzowym” (*fan effect*), lecz autorka badań nie używa tej nazwy. Dokłada starań by oszacowanie wpływu szkoły uwolnić od **czynników kontekstowych**, budując modele statystyczne o różnej złożoności, obejmujące wiek, inteligencję i pozycję społeczno-ekonomiczną rodziny ucznia. Te działania przynoszą umiarkowany sukces teoretyczny, bo wariancja średnich wyników egzaminacyjnych szkół stanowi tylko 5 – 10% wariancji ogólnej (5, s. 86), interkorelacja inteligencji uczniów i ich środowiska społecznego z osiągnięciami szkolnymi jest stosunkowo wysoka, a obciążenie wyników różnicami wieku uczniów okazuje się „zaniedbywalnie małe” (12, s. 135). Informacje o osiągnięciach szkół nabierają natomiast znaczenia praktycznego w badaniach segregacji międzyszkolnej uczniów w dużych miastach na podstawie publikowanych „tabel ligowych” wyników egzaminów zewnętrznych. Efekt wachlarzowy egzaminów, wykryty i objaśniony przez prof. Romana Dolatę, powoduje szybki wzrost zróżnicowania średnich osiągnięć szkół. Nasze gimnazja są znacznie bardziej zróżnicowane co do poziomu niż szkoły podstawowe, w dodatku zróżnicowanie systematycznie rośnie, co jest zjawiskiem edukacyjnie szkodliwym (6, s. 176; 13, 44, rys. 4).

Dyskusja

Analizy statystyczne wykonywane przez mgr Jasińską-Maciążek na poziomie szkoły były tak rozległe i absorbujące, że z jej pola widzenia zniknął pojedynczy uczeń. Tymczasem w psychologii rozwojowej i pedagogice dokonują się współcześnie głębokie zmiany ku **personalizacji** procesu kształcenia i oceniania.

Schyłek behawioryzmu i rozkwit psychologii poznawczej, a w jej łonie – konstruktywizmu, zaowocowały osłabieniem idei wymuszania postępów testowaniem („*test – teach – test*” to ensure *learning*) na rzecz konstruowania wiadomości i ich rozumienia przez uczniów w danym kontekście społecznym.² W pomiarze dydaktycznym akcent przeniósł się na **ocenianie kształtujące** (*formative assessment*), wymagające – w przeciwieństwie do oceniania sumującego – indywidualnych planów

² L.A. Shepard (2000) The Role of Assessment in a Learning Culture. *Educational Researcher*, 27, No. 7, pp. 4-14.

rozwoju uczniów. Analizy statystyczne przez wiele lat dominujące w edukacyjnych zastosowaniach psychometrii ustępują teraz trafniejszym pedagogicznie procedurom wewnątrzszkolnym, wnoszącym wielowymiarowe wspomaganie pracy ucznia.³ O dominacji tego podejścia świadczy fakt, iż obecny przewodniczący Krajowej Rady Pomiaru Edukacyjnego (NCME), wiodącego w USA pedagogicznego stowarzyszenia pomiarowego, Mark Wilson, poświęcił pięć kolejnych listów do członków stowarzyszenia ocenianiu nauczycielskiemu (*classroom assessment*), oddolnemu (*bottom up*), a nie odgórnemu (*top-down*), rejestrującemu „trajektorie uczenia się”, budującemu „mapy drogowe” (*road maps*) postępu uczniów i uwalniającemu szkoły od „błédnego trójkąta, w którym plany nauczycieli co do uczenia się ich uczniów są ściśnięte między programem kształcenia a ocenianiem sumującym prowadzonym dla celów ewaluacyjnych”.⁴

Mgr Jasińska, wyspecjalizowana w analizach statystycznych, nie interesuje się zmianami paradygmatów pomiarowych na świecie i w Polsce. W recenzowanych tekstach, nawet w koronnym „Trafność testów egzaminacyjnych” (1), nie znajdziemy informacji o koncepcjach Michaela Scrivena (ocenianie kształtujące), Samuela Messicka (trafność konsekwencyjna), Michaela Kane’a (trafność argumentacyjna), Roberta Glasera (pomiar sprawdzający) ani o polskiej linii teoretyków pomiaru: psychologów (Mieczysław Choynowski – Jerzy Brzeziński – Elżbieta Hornowska) i pedagogów (Bolesław Niemierko – pomiar wielostopniowy). W kluczowej kwestii trafności wewnętrznej (programowej) dwu testów Autorka rozprawy przyjmuje ryzykowne dydaktycznie założenie o istnieniu cechy ukrytej (1, s. 69) i odsyła nas do Centralnej Komisji Egzaminacyjnej, stwierdzając że „w zakresie trafności jedynym g w a r a n t e m (podkr. BN) stałości mierzonego przez egzaminy konstruktu jest dość szczegółowy plan testu” (13, s. 42).

Uproszczenia niezbędne do uzasadnienia analiz statystycznych są dla pedagoga rażące. W „Prezentacji pracy doktorskiej”, po odwołaniu się do prof. Wincentego Okonia, czytamy:

Tak więc, mimo iż o wynikach nauczania czy kształcenia myślimy mając na uwadze perspektywę szkoły, a osiągnięciach szkolnych – perspektywę ucznia, w obu przypadkach mamy do czynienia z efektami oddziaływań nauczycieli i innych pracowników szkoły i w takim znaczeniu posługuję się tymi pojęciami. W dalszych analizach ograniczam się do wyników nauczania, które traktuję jako podzbiór szerszego pojęcia wyników kształcenia, podzbiór ograniczający się do wyników pracy szkoły w zakresie przyswojonych przez uczniów wiadomości i umiejętności. Gdy w jakimś tekście używam szerszego pojęcia efektów kształcenia, to zwykle mam na myśli ten właśnie węższy podzbiór. Definicje operacyjne pojęć „wyniki nauczania” i „osiągnięcia szkolne” sprowadzam do wyników standaryzowanych testów wiadomości i umiejętności, mając świadomość zawężania przez to ich znaczenia. Jest to konieczne, aby móc zająć się tą częścią wyników nauczania, które poddają się pomiarowi w ograniczonym czasie badań. (1, s. 10)

Tak zredukowane pojęcia skutkują **ograniczoną przydatnością metody EWD** dla szkół, nauczycieli i uczniów. Uzyskiwane wskaźniki punktowe (uwaga: nie „definicje operacyjne”, bo nie obej-

³ S.M. Brookhart et al (2016) A Century of Grading Research: Meaning and Value in the Most Common Educational Measure. *Review of Educational Research*, 86, No. 4, pp. 803-848.

⁴ M. Wilson (2016) From the President. *NCME Newsletter*, 24, No 4, p. 2.

mują całej treści pojęć) dają szkołom zachętę do spotęgowania lub kontynuacji wysiłków, ale ich efekt ma być przede wszystkim testowy. „Szkołna efektywność nauczania” jest sprowadzona do egzaminów zewnętrznych, a „procedury ich tworzenia są przestrzegane dzięki wiedzy i doświadczeniu pracowników centralnej i okręgowych komisji egzaminacyjnych, którzy odpowiadają za ich jakość” (1, s. 71). To podejście wydaje się odpowiadać „ewaluatorom zewnętrznym” (nadzorowi pedagogicznemu), skłonny do centralizmu, uzyskującym w EWD gotową miarę jakości pracy szkoły (9, s. 39).

Co jednak począć w sytuacji, gdy reforma polskiego systemu kształcenia ogólnego „wygasa” ideę scentralizowanego egzaminowania uczniów z wyjątkiem państwowej matury? Rozwiązaniem jest szeroka produkcja standaryzowanych testów szerokiego użytku zaopatrzonych w **normy wartości dodanej**, udostępnianych na rynku edukacyjnym. Dorobek metodologiczny centralnej i okręgowych komisji egzaminacyjnych oraz Instytutu Badań Edukacyjnych nie powinien być zmarnowany. Bardzo cenna jest przestroga przed segregacją szkolną wskutek przeceniania roli egzaminu zewnętrznego.

Konkluzja

Przedstawione do recenzji teksty spełniają, w mojej ocenie, wymagania stawiane rozprawom doktorskim. Zawierają oryginalne dla naszego kraju rozwiązanie problemu optymalizacji modelu ewaluacyjnego jakości pracy szkoły na podstawie edukacyjnej wartości dodanej. Nie budzą wątpliwości co do kompetencji Autorki w zakresie metodologii szerokich badań lustracyjnych (*large-scale surveys*). Trudniej ocenić jej wiedzę dydaktyczną ze względu na ograniczenie tematyki rozprawy do prawidłowości występujących między szkołami, podczas gdy uczniowie są traktowani zbiorowo. Ponieważ edukacyjna wartość dodana szkoły jest miarą tendencji centralnej wzrostu osiągnięć poszczególnych uczniów, sądzę że wkrótce zainicjowane będą w Instytucie Badań Edukacyjnych spersonalizowane badania tych zjawisk, nawiązujące do dorobku rodzimej dydaktyki.⁵ To pomoże nauczycielom w „profesjonalnym planowaniu własnej pracy”, postulowanym w rozprawie (2, s. 113), bo obecnie „nauczyciele oceniają swoich uczniów relatywnie i nie potrafią zakotwiczyć skali w uniwersalnym dla wszystkich punkcie odniesienia (4, s. 424).

Wnoszę o dopuszczenie mgr Aleksandry Jasińskiej-Maciążek do dalszych etapów przewodu doktorskiego. Poczynione w tej recenzji uwagi krytyczne miały na celu ukierunkowanie dalszego pomysłnego rozwoju jej kompetencji naukowej oraz poszerzenie zadań zespołów badawczych, w jakich przyjdzie jej pracować.

B. Niemcewicz

⁵ Ubocznym efektem powściągliwości Autorki tekstów w korzystaniu z dzieł polskojęzycznych są trudności terminologiczne. Na przykład (1, s. 84) czytamy o postulacie, „by wyniki pomiarów wykorzystanych do modelowania miały wystarczająco dużo wartości”, a wystarczyło napisać, że skala powinna być dostatecznie **długa** (na kolejnej stronie jest tajemnicza wzmianka o „szerszych skalach”...) oraz że „wzrost o jedną jednostkę na skali oznacza równoważny przyrost wiadomości i umiejętności w każdym zakresie skali”, a wystarczyło napisać, że **przedziały skali** są równe pod względem wiadomości i umiejętności. Niekiedy pomysł jest zabawny, bo **efekt pułapu** (*ceiling effect*) został nazwany „efektem sufitowym”, co sugeruje istnienie „sufitu osiągnięć ucznia”.