

prof. dr hab. Bolesław Niemierko
Akademia Marynarki Wojennej w Gdyni

Dyskusja referatu dr. Macieja Jakubowskiego „Międzynarodowe badania umiejętności uczniów a badania w Polsce”

Trudno przecenić wartość, jaką ma dla uczestników konferencji, przegląd zdobyczy metodologicznych Programu Międzynarodowej Oceny Umiejętności Uczniów (PISA), dokonany przez „naszego człowieka” w pionie edukacyjno-badawczym Organizacji Współpracy Ekonomicznej i Rozwoju (OECD), a ukierunkowany na ulepszenie krajowych badań osiągnięć uczniów, w tym – egzaminów zewnętrznych.

Przypomnę, że udział Polski w badaniach Międzynarodowego Stowarzyszenia Badań Osiągnięć Pedagogicznych (IEA) na przełomie lat sześćdziesiątych i siedemdziesiątych zaowocował szerokim wdrożeniem tak fundamentalnych pojęć i procedur pomiarowych jak taksonomia celów kształcenia (*Blooma*, potem *ABC*), analiza zadań (w tym archaiczny już dziś wskaźnik D_{25}^1), skali staninowej, szacowania rzetelności i, szerzej, zastosowań klasycznej teorii testu. Obecnie, jak przekonał nas autor referatu, mamy okazję do wzorowania się na zastosowaniach znacznie bardziej zaawansowanych modeli, spośród których za najcenniejsze uważam cztery następujące:

1. Szeroki **zakres programowy oceniania** (*assessment framework*), wybiegający poza program kształcenia ku umiejętnościom nabywanym poza szkołą, w drodze indywidualnego rozwoju poznawczego w poszerzonym środowisku edukacyjnym.
2. **Macierzowy dobór próby** treści i populacji (*matrix sampling*), polegający na zróżnicowaniu **zeszytów testowych** (*test booklets*), pozwalającym na bardziej gęste pokrycie zadaniami rozległego zakresu programowego oceniania. Najprostszym modelem takiego doboru jest tworzenie wersji równoległych testu z zadaniami kotwiczącymi te wersje.
3. Wiązanie **pomiaru różnicującego** z **pomiarem sprawdzającym wielostopniowym** przez wyznaczanie poziomów umiejętności w **skali akademickiej** (*CEEB*) oraz określanie ich jako pewnej liczby (np. trzech lub pięciu) **standardów edukacyjnych**, stanowiących „kategorie jakościowe” umiejętności. Tymczasem w Polsce nawet dwa poziomy matury (podstawowy i rozszerzający) nie tworzą dotychczas

¹ Ten wskaźnik, jak najstarsi „pomiarowcy” zapewne pamiętają, obliczało się w ołówku z sum liczebności grup skrajnych (ćwiartek rozkładu) w niewielkiej próbie uczniów. A teraz potrafimy szacować parametr a modeli probabilistycznych i próby standaryzacyjne urosły do paru tysięcy uczniów!

wspólnej skali, która pozwalałaby na przewidywanie osiągnięć niższego poziomu z osiągnięć wyższego poziomu, to jest na zaliczanie niższego poziomu na podstawie pozytywnego wyniku sprawdzania wyższego poziomu.

4. Wprowadzenie **wartości spodziewanych** (*plausible values*) jako swoistych *norm grupowych* (środowiskowych, lokalnych), na które są orientowane wyniki indywidualne. To podejście, ryzykowne w ocenianiu wewnątrzszkolnym ze względu na zagrożenie stereotypizacją, może znaleźć zastosowanie w ewaluacji jakości pracy szkół działających w różnych środowiskach i regionach.

Tyle w skrócie o głównych naukach płynących z głęboko inspirującego referatu dr. Jakubowskiego. Obejmował on także wiele innych aktualnych zagadnień – niekoniecznie nazwanych terminami stosowanymi w krajowej diagnostyce edukacyjnej – takich jak **stałość wyniku pomiaru** (*test score stability*), **zrównywanie wyników** (*equating*), **skalowanie pionowe** (*vertical scaling*) i tworzenie **banków zadań** (*item banks*) zaopatrzonych w parametry **teorii wyniku zadania** (*IRT*) oraz interpretacje **znaczenia treściowego** (*content meaning*) każdego zadania. To nie są nowe pojęcia, ale daleko nam do systematycznych wdrożeń odpowiednich procedur. Wypada mi wyrazić nadzieję, że – jak to było z elementami klasycznej teorii testu – potrafimy wkrótce opanować ten arsenał pomiarowy, a potem zastosować go w naszym systemie wielkich badań pedagogicznych i doniosłych egzaminów.