

Wprowadzenie

Pojęcie „trafności pomiaru” osiągnięć obejmuje współcześnie nie tylko teoretyczne uzasadnienie mierzonych właściwości uczniów i użyteczność wniosków, jakie możemy oprzeć na wynikach pomiaru, lecz także całokształt „konsekwencji”, jakie pomiar wnosi w życie ucznia, pracę nauczyciela, działalność szkoły i funkcjonowanie systemu oświaty. To pole bezkresne i ogromny ciężar odpowiedzialności. Każdy przyczynek do problematyki trafności witany jest z radością, już to jako zwiastun postępu lub przynajmniej nadążania za krajami przodującymi w diagnostyce edukacyjnej, już to jako sygnał trudności – pojęciowych, technicznych, organizacyjnych, komunikacyjnych, badawczych i wdrożeniowych – które, jeśli nie dzisiaj, to wkrótce – musimy pokonać.

W części teoretycznej publikacji zamieszczamy teksty, które przynoszą do kraju nową wiedzę o problematyce trafności pomiaru, porządkują pojęcia oraz dają przykłady modeli i metod badawczych. Otwiera ją obszerna rozprawa prof. Bolesława Niemierki *Jaki pomiar dydaktyczny jest nam potrzebny?* Przedstawia on przyczyny i etapy wyodrębniania się pomiaru dydaktycznego z psychometrii, różnorodne modele i dyskusje aktualne dla obecnego stanu diagnostyki edukacyjnej w Polsce, a wreszcie emocjonalne i poznawcze wyzwania, przed jakimi stoi nasz system egzaminacyjny. Autor nawołuje do badania „wielu cech” i do stosowania „wielu miar” w diagnostyce edukacyjnej.

W oryginalnym raporcie badawczym *Egzamin 2003: rzetelność i trafność testu matematyczno-przyrodniczego* prof. Krzysztof Konarzewski pokazuje, jak trudno jest potwierdzić strukturę wymagań egzaminacyjnych za pomocą analizy wyników zadań testowych. Ponieważ zadania z różnych przedmiotów i z różnych zakresów materiału

systematycznie różnią się formą i trudnością, analiza czynnikowa najpierw wyodrębniła po prostu zadania trudne – otwarte (średnia łatwość „czystych” nosicieli czynnika: 0,32), potem łatwe – zamknięte (0,69), potem bardzo trudne – otwarte (0,24), a na koniec bardzo łatwe (0,94). Tak to prawa rządzące współczynnikiem korelacji zmienionych zero-jedynkowych (ograniczenie wartości przez proporcje brzegowe) zdominowały strukturę wyników analizy. Należy wszakże wątpić, czy zastosowanie silniejszych i lepiej wyważonych miar osiągnięć objętych standardami potwierdziłoby ich strukturę w analizie czynnikowej, bo standardy reprezentują układ celów kształcenia, a rozwiązania zadań – układ wyników pomiaru. Hierarchie programowe i hierarchie osiągnięć uczniów, jak dowodzą badania amerykańskie (np. Resnicków), nie pokrywają się.

Prof. Peter Krope i prof. Wilhelm Wolze z Uniwersytetu w Kilonii przedstawili opracowanie *Język potoczny a język naukowy. Trafność testu z pozycji konstruktywizmu*, które w interesujący sposób wspiera końcowe zdania poprzedniego akapitu. Otóż uczniowie (w Niemczech i w Polsce, bo badania mają charakter porównawczy), a także niektórzy dorośli, posługują się pojęciami przednaukowymi (*life worldly knowledge*) w rozumowaniach dotyczących otaczającego świata. Ich myślenie, wnioskowanie i rozwiązania zadań są inne niż podręcznikowe. Swe badania autorzy wiążą z pojęciem trafności teoretycznej testu.

Piotr Skorupiński przedstawił *Genezę pojęcia trafności teoretycznej* we wzorowej filologicznie rozprawie, sięgając do bardzo wielu, mało znanych w Polsce źródeł. Doceniając odwagę Samuela Messicka, który otworzył to pojęcie na świat przewidywanych konsekwencji pomiaru, nie powinniśmy zapominać o fundamentalnych koncepcjach decyzyjnych Lee Cronbacha, który potrafił połączyć tradycje korelacyjne trafności ze zrozumieniem znaczenia sieci pojęciowej („nomologicznej”), mieszczącej sens mierzonych właściwości. Dzięki tej rozprawie poznajemy wiele dziejowych postaci dyskursu trafności pomiaru, ich dzieł, poglądów, a nawet charakterystycznych wyrażań. Warto ją przestudiować, choćby tylko jako przykład dobrej roboty naukowej.

Michał Daszkiewicz poświęcił swoje ambitne studium *Znaczenie i zawartość „argumentu” oraz jego rola w ustalaniu trafności i komunikowaniu wyników egzaminu* nakreśleniu bogactwa treści pojęcia „dyskursu” (ang. *argument*) oraz doniosłości funkcji argumentacyjnej języka w świetle klasyfikacji Karla Poppera. To studium otwiera drogę badaniom nad przebiegiem (pisemnego lub ustnego) dyskursu niezbędnego dla ustalania trafności konsekwencyjnej egzaminów.

Próby Marka Krynińskiego, przedstawione w artykule *Klasyczne i probabilistyczne miary jakości zadania testowego – nowe możliwości*, zbliżają nas do systematycznego stosowania probabilistycznych modeli wyniku zadania testowego w diagnostyce edukacyjnej. Autor zwięźle nakreślił potrzebę wykroczenia poza klasyczną teorię testu oraz podał przykłady analizy zadań z rejonowego konkursu informatycznego w województwie pomorskim. Kolejnym krokiem postępu w tej dziedzinie będą krajowe warsztaty zastosowań modeli probabilistycznych, poprowadzone jako specjalizacja w ramach

sesji końcowej VI Podyplomowego Studium Pomiaru Dydaktycznego i Oceniania w Gdańsku (na początku lutego 2004 r.).

Interesujące, bo proste w zamyśle badania przeprowadziła dr Elżbieta Kowalik. W raporcie *Dwa rodzaje egzaminów w opiniach uczniów, nauczycieli oraz dyrektorów szkół* pokazuje, że świadomość potrzeby zarówno egzaminu powszedniego, jak i doniosłego jest już stosunkowo wysoka, choć dyrektorzy szkół znacznie wyprzedzają pod tym względem nauczycieli i uczniów. Do znajomości wymagań programowych natomiast uczniowie przyznają się znacznie mniej chętnie, niż sądzić by można z oświadczeń ich nauczycieli.

Grażyna Szyling zatytułowała swój raport *Być może będzie w przyszłości ozdobą polskiej piłki nożnej? Wybrane aspekty trafności oceny końcowej osiągnięć ucznia*. Badanie dotyczy „szarej strefy” oceniania szkolnego, której reguły są inne niż oficjalne, dość zagmatwane, mało znane. Wzorem amerykańskich badań Susan Brookhart podejmujemy w kraju próby rozpoznania „drugiego układu wymagań”. Tu przedstawione sprawozdanie jest przykładem starannej analizy materiału ankietowego i docieklowości w interesujących nas kwestiach.

W części szczegółowej publikacji mieszczą się dwie serie prac:

I. „Wokół trafności egzaminowania” (sesje tematyczne IAB – IIAB)

II. „Od motywacji ucznia do etyki egzaminatora” (sesje tematyczne IIIAB – IVAB).

Wszystkim Czytelnikom tego wydawnictwa życzymy interesujących, niezbyt męczących a pożytecznych lektur.

Redaktorzy tomu