

Prof. dr hab. Bolesław Niemierko

Uniwersytet SWPS w Sopocie

Co mi się zdaje, że wiem o diagnostyce edukacyjnej Jesienne refleksje

Okoliczności

W marcu Wydawnictwo Naukowe PWN poprosiło mnie o przygotowanie nowego wydania *Diagnostyki edukacyjnej*. Zanim w lipcu przesłałem do niego 700-stronicowy maszynopis, musiałem przestudiować poprzednią wersję podręcznika i uzupełnić ją tym, co uznałem za najważniejsze w światowym i krajowym rozwoju dziedziny. To dało mi okazję do oceny mojej wiedzy i niewiedzy o diagnostyce edukacyjnej.

Wśród wielu zajmujących materiałów znalazłem inauguracyjne przemówienie, które Mark Reckase wygłosił w amerykańskiej Krajowej Radzie Pomiaru Edukacyjnego (NCME), obejmując stanowisko jej prezesa w 2009 roku. Miało tytuł *Co mi się zdaje, że wiem* (Reckase, 2010), zaczerpnięty z tomu wierszy zaprzyjaźnionego poety (Dana, 1999). Postanowiłem więc spróbować pójść śladem Reckase'a i krytycznie ocenić swoją wiedzę o dziedzinie, w której działam od ponad pół wieku. Pandemia sprzyja retrospekcjom i refleksjom.

Co wiem

Podobnie jak Reckase, wiem na pewno, lepiej niż w poprzednich latach, że **człowiek jest skomplikowany** (*people are complicated*). Jego złożoność można porównać do genomu istot żywych, a nawet do bezkresu Wszechświata. Chodzi tu zarówno o liczbę i różnorodność elementów, memów oraz doświadczeń życiowych, jak i o misterne sieci powiązań między nimi. Żadne badanie psychologiczne ludzi nie może zamknąć się konkluzją, że poznaliśmy je dokładnie.

Władysław Jacek Paluchowski (2001) przyjął następujące twierdzenie jako motto podręcznika diagnostyki psychologicznej:

Każdy człowiek jest pod pewnym względem [a] jak wszyscy inni, [b] jak niektórzy inni ludzie, [c] jak nikt poza nim.

Grupa *a* tych właściwości człowieka interesuje głównie biologów, grupa *c*, fascynująca i często wyolbrzymiana, sugeruje jałowość wszelkich uogólnień, a tylko grupa *b* umożliwia psychologię, socjologię, pedagogikę i resztę nauk humanistyczno-społecznych.

W przytoczoną maksymę włączył się Wiesław Łukaszewski, który w swym dziele *Pięknie się różnić, mądrze być podobnym* postawił następujące pytania (2019, s. 36):

Pod jakimi względami wszyscy ludzie są podobni do siebie? Czy ktoś, gdzieś, kiedyś odważył się to powiedzieć? Pod jakimi względami każdy człowiek jest inny niż pozostali?

Po drugie, kim są ci niektórzy, do których człowiek jest podobny, i ci niektórzy, do których podobny nie jest? Co ich wyróżnia?

Tak oto wewnętrzne skomplikowanie ludzi nie tylko wskazuje pole badań nad człowiekiem, lecz także wyznacza jego granice. Zróżnicowanie zerowe i zróżnicowanie totalne stanowią ramę, w której mieści się nieskończenie wiele ogólnych struktur emocji, przekonań i zachowań, w tym zjawisk edukacyjnych podlegającej diagnozie.

Wiem także, że zajmując się ludźmi, **jestemy skazani na błędy**. To nieunikniony skutek komplikacji przedmiotu poznania. Jak tłumaczy Michael Kane (2010; 2011), uproszczenia, a więc niedokładności diagnoz, są nieuniknione. Różnica między laikiem szafującym swoimi ocenami a uczonym budującym modele rzeczywistości wyraża się w wielkości i świadomości błędów. Laik chętnie wygłasza pochopne sądy, nie dopuszczając myśli, że może się mylić. Uczony zakłada określoną *tolerancję błędu*, która pozwalałaby zachować pożytek z badania.

Profesjonalizm diagnozy to ujmowanie zjawisk w szerokim kontekście, przekraczającym wyobrażenia laików. Polega na zestawianiu wyników wielu pomiarów, z których każdy jest obciążony błędami.

W teorii pomiaru rozróżniamy błędy losowe i systematyczne. *Błędy losowe* możemy kontrolować statystycznie. Staramy się je zmniejszać przez powtarzanie czynności, tak by wyeliminować niekontrolowane wpływy. W najprostszym przypadku mnożymy pytania kwestionariuszowe i zadania testowe, rozbudowując narzędzia. *Błędy systematyczne* są zwykle ukryte (Niemierko, 2012). W edukacji wynikają nie tylko z chwiejności programów kształcenia, lecz także z indywidualnego odbioru każdej sytuacji. Nazwano to *współczynnikiem humanistycznym* badania ludzi i uczyniono postulatem metodologiczno-etycznym nauk społecznych (Znaniński, 1934; Wysocka, 2013, s. 304).

Standaryzacja procedur i narzędzi pomiarowych jest skuteczna w redukcji błędów losowych pomiaru, ale nie błędów systematycznych. Kane napisał, że „prowadzi [ona] do zmniejszenia błędu losowego i zwiększenia błędu systematycznego, [a więc] im więcej standaryzujemy, tym bardziej zmierzamy do zmniejszenia błędu losowego i zwiększenia błędu systematycznego [pomiaru]” (2011, s. 18). Jak to się dzieje? Uściślając warunki, przebieg i analizę wyników diagnozy, zawężamy jej pole i usztywniamy interpretację, zatem więcej aspektów problemowych i personalnych spychamy na margines.

Błędy nie tkwią w żadnym punkcie pola dokonywanej diagnozy. Są cechą naszej interpretacji faktów. Gdy kwitujemy obserwację jedną liczbą lub jednym słowem, ponosimy wielkie ryzyko. Jednak by działać, musimy upraszczać obraz rzeczywistości. Kane (2011, s. 12) stwierdził, że „nasze modele muszą być proste, by dały się stosować (*to be manageable*)”. Autorzy badań miewają z tym problem, gdy pragną nas olśnić finezją warsztatu naukowego.

Działania edukacyjne wnoszą *zmiany systemowe* w jednostkach i grupach społecznych. Z tymi zmianami wiąże się trzecie i ostatnie z moich silnych przekonań. Otóż **głębokie zmiany dokonują się powoli** na każdym poziomie. O trudach przekształcenia – z wewnątrz, a zwłaszcza z zewnątrz – charakteru człowieka wie każdy wychowawca. O długim dojrzewaniu systemu oświatowego kraju upewniłem się co najmniej dwukrotnie:

1. W Stanach Zjednoczonych, jak mogłem to zaobserwować na miejscu, przez dziesiątki lat trwa przechodzenie z systemu edukacji 8 + 4, na system 6 + 6 lub, częściej, 6 + 3 + 3. Te systemy działają w poszczególnych stanach, a niekiedy w sąsiednich inspektoratach. O zmianie decydują: wola obywateli, przekonania działaczy, zasoby kadrowe i względy logistyczne, w tym sieć komunikacyjna okolicy. Pedagodzy uważają to za korzystne.
2. Na spotkaniu ekspertów unijnych w przodującej edukacyjnie Finlandii przekonaliśmy się – z pewnym zaskoczeniem – że każdy z krajów europejskich jest w jakimś stadium reformy oświatowej: przygotowania, wdrażania lub mozolnego utrwalania z nieustannym pokonywaniem wypaczeń. W tych działaniach urzędowi i naukowcy eksperci nie zalecali pośpiechu, pilnie dbając o jak najlepszą sytuację uczniów.

Również Polskie Towarzystwo Diagnostyki Edukacyjnej nastawia się na analizę skutków oświatowej zmiany systemowej. Zmiana dokonana w 2017 roku, o kierunku przeciwnym reformie amerykańskiej i polskiej reformie sprzed 18 lat, stała się przedmiotem kilkunastu referatów na zeszłorocznej konferencji Towarzystwa. Szczególnym polem diagnoz jest sytuacja uczniów (Szmigel i Szaleniec, 2019) i nauczycieli (Kędzierska, 2019). Uzasadniony jest pogląd, że w okresach *tranzycji* dwa systemy działają – w emocjach, poglądach i decyzjach – na zasadzie *hybrydowej*, przeplatając się wpływami (Szyling, 2018).

To tyle. Mark Reckase napisał (2010, s. 3): „«Co wiem», odnosi się do tego, co wiem na pewno wskutek pracy w obszarze pomiaru edukacyjnego. Po zastanowieniu się, zdałem sobie sprawę, że wiem niewiele (*I realized that I did know very much*)”. Także moją, autora tych wyznań, wiedzę o człowieku, jego błędach i niepodatności na zmiany można uznać za banalną, a nawet za *Banały powtarzane bez końca* (Łukaszewski, 2019, tytuł rozdziału). Rzecz w tym, że o skomplikowaniu człowieka, niepewności sądów o nim i trudności edukacji zbyt często zapominamy. A to hamuje postęp diagnostyki edukacyjnej.

Co mi się zdaje, że wiem

Teraz przejdę do przekonań dalszych od pewności, a przez to dyskusyjnych. Zacznę od kolejnych pytań Wiesława Łukaszewskiego (2019, s. 36):

Po trzecie, co w tym kontekście znaczy słowo „jest” (podobny lub niepodobny). Czy mowa o stanie trwałym (był, jest i będzie), czy raczej o stanie zmiennym (jest teraz i tutaj, ale bywa, że nie jest). A zatem, czy chodzi o cechy, czy raczej o stany?

Po czwarte, czy ci wszyscy, którzy są podobni, są podobni pod tymi samymi względami, a jeśli jedno lub drugie, to pod jakimi?

Te pytania, moim zdaniem, prowadzą do *typologii* rozumianej jako uporządkowanie elementów zbioru przez porównywanie ich cech z cechami elementów uznanych za *typy* (modele). Typologia to uporządkowanie logicznie słabsze niż

klasyfikacja, gdyż nie musi być wyczerpujące i rozłączne. Odgrywa dużą rolę w naukach społecznych, obejmując obszary o wielkiej złożoności i niskiej trafności pomiaru, takie jak środowisko, kultura, zainteresowania, temperament, osobowość, kompetencja człowieka. W każdej z tych dziedzin jednostka może mieć właściwości kilku różnych typów, i w większości przypadków tak bywa.

W nowej wersji podręcznika *Diagnostyka edukacyjna* typologie pojawiają się już na wstępie. To przede wszystkich *typologia uczenia się*, oparta na Wincentego Okonia (1967) teorii wielostronnego kształcenia, a nadto *typologia diagnostów*, zgodna z Czesława Nosala (1990) koncepcją modeli umysłu. Jestem przekonany, że **typologie w diagnostyce edukacyjnej są użyteczne** mimo umowności i ruchomości typów. Zastosowana w ogólnopolskich badaniach typologia uczenia się pozwoliła na zakwalifikowanie około 2/3 gimnazjalistów jako uczących się głównie przez (α) *przyswajanie* (30%), (β) *działanie* (15%), (γ) *odkrywanie* (40%), i (δ) *przeżywanie* (15%). W szkołach bardziej selekcyjnych i na wyższych poziomach kształcenia rośnie przyswajanie gotowego materiału kosztem odkrywania. W sopockiej uczelni (SWPS) stwierdziłem, że wśród studentów psychologii mamy (α) *proceduralistów* (20%), (β) *ryzykantów* (30%), (γ) *teoretyków* (20%) oraz (δ) *intuicjonistów* (30%) diagnozowania uczenia się. Na wyższych latach studiów szybko rośnie grupa proceduralistów, kosztem – co zrozumiałe – grupy ryzykantów, pochopnie dokonujących diagnoz (Niemierko, 2015).

Dane te nie są ściśle, postawy badanych są zmienne, a mimo to diagnoza typologiczna prowadzi do cennych wniosków. Najważniejszy z nich dotyczy przesadnej *intelektualizacji* kształcenia, wymuszającego przyswajanie wzorów rozumowania przez uczniów gotowych do uczenia się innymi drogami. Zasoby emocjonalne i zdolności praktyczne młodzieży są zbyt mało wykorzystywane w szkołach ogólnokształcących w większości krajów europejskich.

Wielostronność uczenia się prowadzi do wniosku, że **pomiar w edukacji powinien mieć charakter sprawdzający** (*criterion-referenced*), a nie różnicujący (*norm-referenced*). To znaczy, że osiągnięcie pewnej kolekcji celów kształcenia, a nie wyprzedzanie rówieśników na wyznaczonej trasie jest w nim pierwszoplanowe. Pierwszoplanowe, ale nie w pełni metodologicznie odrębne. Tę odrębność głoszą w epoce zmięchu behawioryzmu i narodzin psychologii poznawczej, fascynując się nauczaniem programowanym i technologicznym postępowaniem kształcenia (Glaser, 1963; Linn, 1994; Millman, 1994). Operacjonalizacja celów kształcenia okazała się nie dość zobiektywizowana, mimo to przetrwał pogląd, że pomiar sprawdzający mieści się w rygorach metodologii nauk społecznych (*Standards...*, 2014, s. 96 i 183–201).

W naszym kraju pomiar sprawdzający jest chlebem codziennym pedagoga. Książki na jego temat (Niemierko, 1990, 1991, 1999, 2002) miały sporo wydań, odbyło się wiele studiów podyplomowych z tej dziedziny, a Centralna Komisja Egzaminacyjna i jej okręgowe oddziały spopularyzowały testy sprawdzające

Wprowadzanie pomiaru sprawdzającego do programu kształcenia nauczycieli rozpoczęto na Wydziale Chemii Uniwersytetu Łódzkiego (Zakrzewski i Wypych-Stasiewicz, 2017). Analityczność chemii, a szerzej – ścisłość przedmiotów przyrodniczych wydaje się sprzyjać zainteresowaniu pomiarem

sprawdzającym. Mankamentem pozostaje niezwykle niska norma ilościowa (30% punktów) stosowana do zaliczania większości przedmiotów szkolnych. To kompromis między ambicjami dydaktyków (fasadą systemu edukacyjnego) a stanem osiągnięć większości uczniów.

Może trzeba zrezygnować z jednolitych podstaw programowych i zróżnicować wymagania? Przez wiele lat byłem pewny, że każdą treść kształcenia da się uporządkować według skali stopni szkolnych. Tak powstała teoria *kształcenia wielostopniowego* (Niemierko, 1990), dobrze znana większości członków PTDE. Już straciłem tę pewność, ale założenie, że **osiągnięcia uczniów powinny być rozróżniane jakościowo** wciąż należy do mojego pedagogicznego credo. Rejestrowanie liczby rozwiązanych zadań i punktów przyznanych według pewnych kryteriów pełni funkcję podrzędną, pomocniczą. Nie powinno być podstawą konstrukcji skali ocen szkolnych. Przekładanie wyników pomiaru różnicującego, z natury homogenicznego, nawet tak statystycznie wyrafinowanego jak w teorii wyniku zadania (IRT; Szalaniec, 2010), na oceny szkolne jest uproszczeniem, na które nie mogę się zgodzić.

Po ćwierćwieczu doświadczeń z *testem sprawdzającym wielostopniowym* (TSW) wyłoniły się dwa jego ograniczenia:

1. Tylko niektóre zakresy treści kształcenia mają *strukturę hierarchiczną*. Już na starcie teorii pomiaru sprawdzającego zauważył to Anthony Nitko (1980). Cztery działania arytmetyczne bez posługiwania się kalkulatorem stają się coraz trudniejsze w miarę zwiększania zakresu liczbowego tych działań, ale teoria zbiorów nie jest łatwiejsza niż oparta na niej algebra elementarna. Dydaktycy chcieliby uczynić naukową strukturę przedmiotu podstawą skali ocen, a to się nieraz kłóci z hierarchią uczenia się (*learning hierarchy*; Resnick, 1973).
2. Jeśli uczeń może zdobywać wiedzę własną drogą, jak to wynika z typologii uczenia się, to ocenianie jego postępów musi to uwzględniać. I znów, u zarania pomiaru sprawdzającego zauważono, że niektórzy poprawnie stosują prawo fizyczne (np. prawo Ohma), nie znając jego treści (Hnilickova, 1970). Dla jednych uczniów postęp to poprawne zastosowanie posiadanych wiadomości, a dla innych – teoretyczne uzasadnienie opanowanych już umiejętności. Na horyzoncie mamy więc indywidualizację skali ocen szkolnych, co będzie szokiem dla zwolenników tradycyjnego systemu klasowo-przedmiotowo-lekcyjnego.

Co mi się zdaje

W przestrzeni swobodnych myśli, gdzie uczucia mieszają się nagminnie z logiką, wciąż nawiedzają mnie idee diagnostyki edukacyjnej. Te idee dotyczą niedalekiej przyszłości.

Pierwsza z nich to przekonanie, że **wzrasta rola samoregulacji ucznia**. Rewolucja cyfrowa, umożliwiającą dostęp do bezmiaru informacji i tworzenie dowolnej sytuacji wirtualnej, wyzwoliła samodzielność jednostek. Umiejętność wyboru, przekształcania i oceny własnych czynności staje się motorem ich rozwoju (Ledzińska, 2000).

Nauczycielom niełatwo jest wspierać samoregulację uczniów, bo wydaje się ograniczać ich władzę i często bywa przez młodzież nadużywana. Także rodzice mogą pragnąć „mocnego” pedagoga, który wzięłby odpowiedzialność za uczenie się ich dzieci. Wśród powolnych zmian systemowych edukacji ta może trwać najdłużej, tym bardziej należy ją monitorować.

Jeśli tak, to zapewne **rosnąć będzie rola oceniania** – w szkole i w sieciowym systemie wspomagania uczenia się. *Ocenianie* rozumiem jako dostarczanie informacji potrzebnej do rozwoju, nie zaś jako rozgrywanie władzy przez zwierzchnika. Zakładam ciągły postęp demokratyzacji wychowania, redukcji jego autorytarnego składnika. To także potrwa długo, zwłaszcza w konserwatywnych krajach, bo globalizacja kultur jest skrajnie przewlekłym procesem.

Zdradzę tu jeszcze jedną, najodważniejszą nadzieję, a mianowicie że **ludzie stają się coraz bardziej etyczni**, w dużej mierze dzięki edukacji. Dość trudno w to uwierzyć, gdy wszędobylskie media donoszą nam o lawinie wyrafinowanych przestępstw, a w tym o strzelaninach w szkołach przodującego gospodarczo kraju. Jednak liczba wojen maleje, prawa mniejszości rosną, religie miarkują zaborczość, słabi zyskują opiekę. Nawroty egoizmu w skali od jednostki do narodu są rzadsze i coraz mocniej nas rażą.

Diagnostyka edukacyjna ma udział w tych korzystnych przemianach. Jej normy moralne są dobrze ugruntowane (Groenwald, 2001; Groenwald, 2011), a uczniowie pilnie baczą na etykę oceniania osiągnięć (Groenwald, 2004). Wierzę, że rozwój diagnostyki podniesie humanistyczne walory edukacji. Skłonni są to przyznać nawet filolodzy języka ojczystego, najskuteczniejsi w obronie duchowych wartości kształcenia przed inwazją chłodnej technologii.

Przytoczę wydarzenia z przełomu wieków, których ślady napotkałem, pracując nad nowym wydaniem *Diagnostyki*. Przed wprowadzeniem systemu egzaminów testowych do szkół ogólnokształcących ukazał się monograficzny numer miesięcznika *Polonistyka* poświęcony tej tematyce i zatytułowany *Testy, testy*. Znalazło się w nim miejsce dla (1) sceptycyzmu Tadeusza Patrzalka (1998a), który ostrzegał przed „zamerykanizowaniem mózgu narodu testami” (cytat z *Twórczości*) i przed „elegancją przesłuchania policyjnego”, oraz dla jego rozważań *O strukturze testu czytania ze zrozumieniem*, w których aprobeuje test złożony z zadań zamkniętych, jeśli jest oparty na tekście użytkowym, a nie literackim (Patrzalek, 1998b), (2) dla mojego wprowadzenia do pomiaru dydaktycznego (Niemierko, 1998) i (3) dla pogłębionego studium Bożeny Chrzastowskiej, którego założeniem był **hybrydyczny** charakter przedmiotu szkolnego *język polski*, obejmującego „treści różnych dyscyplin humanistycznych i różnych dziedzin artystycznych” w różnym stopniu podatnych na obiektywizację (Chrzastowska, 1998).

Gdy system egzaminów zewnętrznych był już bardzo blisko, inny z redaktorów *Polonistyki*, Stanisław Bortnowski, zauważył, że „tezy Niemierki i jego wyznawców [zbyt] mocno zaciążyły na zmieniającej się szkole”, gdyż ci pedagodzy pragną: „[1] zniszczyć tradycję dydaktyczną (przez wrogość wobec erudycji oraz pasję mierzenia i podliczania wszystkiego), [2] narzucić sprawdzian złożony z zadań krótkiej odpowiedzi, [3] rysować [niepotrzebne] grafy, [4] dzielić

każdy element treści na podstawowy, rozszerzający lub dopełniający”. Redakcja *Polonistyki* poprosiła mnie o ustosunkowanie się do tych zarzutów, co potraktowałem jako okazję do kolejnej dłuższej wypowiedzi (Niemierko, 1998).

Po roku stał się cud. Redaktor Bortnowski przyznał, że jego *dystans* do pomiaru edukacyjnego *zmaliał* wskutek moich wyjaśnień, a pojednawczy ton moich wyjaśnień określił jako *zdumiewający* (Bortnowski, 2001). Zaskoczenie było obopólne! W naszym kraju, w którym **tolerancja**, czyli wyrozumiałość dla odmiennych poglądów, jest w polityce i obyczajach ewenementem, a medialne konfrontacje niemal zawsze zwiększają, zamiast zmniejszać dystans między przekonaniami dyskutantów, traktuję to jako sukces nie tyle osobisty, ile spowodowany walorem etycznym diagnostyki. Wciąż wyobrażam sobie szkołę, w której ocenianie osiągnięć zbliża, a nie oddala uczniów i nauczycieli, humanizuje ich, a nie antagonizuje...

Na koniec

Przeciwnie niż Reckase, który po wynurzeniu się z odmętów statystycznej teorii zadania testowego (IRT) odczuł potrzebę zadeklarowania edukacyjnego optymizmu, ja – przedstawiciel narodu znanego z narzekania (Wojciszke i Rotkiewicz, 2018) – uważam, że wystarczy realizm. Wieloletni trud działaczy Polskiego Towarzystwa Diagnostyki Edukacyjnej, skutecznie dowodzonego przez Henryka Szaleńca i Marię Krystynę Szmigel, a wspieranego udziałem teoretyków w corocznych konferencjach, owocuje rozwojem tej dyscypliny. Jej ducha można wyrazić łacińską maksymą:

Słońce świeci dla wszystkich.

Tę myśl pozwoliłem sobie uczynić mottem książki, której wznowienie dało mi okazję do jesiennych refleksji.

Bibliografia

- Bortnowski, S. (2000) Polonista wobec książki Bolesława Niemierki „Między oceną szkolną a dydaktyką. Bliżej dydaktyki”. *Polonistyka*, 5.
- Bortnowski, S. (2001) Zdumiewający przykład tolerancji. *Polonistyka*, 2.
- Chrzastowska, B. (1998) O humanistyczny wymiar egzaminu. *Polonistyka*, 4.
- Dana, R. P. (1991) *What I think I know: New and selected poems*. Chicago: Another Chicago Press.
- Glaser, R. (1963) Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 18.
- Groenwald, M. (2001) Powinności oceniających. W: B. Niemierko i M.K. Szmigel (red.) *Teoria i praktyka oceniania zewnętrznego*. Kraków: PANDIT.
- Groenwald, M. (2004) Meandry sprawiedliwego oceniania. W: B. Niemierko i H. Szaleńca (red.) *Standardy wymagań i normy testowe w diagnostyce edukacyjnej*. Kraków: PTDE.
- Groenwald, M. (2011) *Etyczne aspekty egzaminów szkolnych*. Gdańsk: Wydawnictwo UG.

- Hnilickova, J. (1970), *Vyzkum formalismu ve znalosti fyzikalniho zakona*. Praha: Academia.
- Kane, M. (2010) *Errors of measurement, theory, and public policy*. Princeton: ETS.
- Kane, M. (2011) The errors of our ways. *Journal of Educational Measurement*, 1.
- Kędzierska, H. (2019) Nauczyciele gimnazjum – w poszukiwaniu nowej tożsamości. W: B. Niemierko i M.K. Szmigel (red.) *Wspomaganie rozwoju kompetencji diagnostycznych nauczycieli*. Kraków: PTDE.
- Ledzińska, M. (2000) Uczenie się wykraczające poza warunkowanie. W: J. Strelau (red.) *Psychologia. Tom 2. Psychologia ogólna*. Gdańsk: GWP.
- Linn, R.L. (1994) Criterion-referenced measurement: A valuable perspective clouded by surplus meaning. *Educational Measurement: Issues and Practice*, 4.
- Łukaszewski, W. (2019) *Pięknie się różnić, mądrze być podobnym. O pożytkach z dostrzegania podobieństwa między ludźmi*. Sopot: Smak Słowa.
- Millman, J. (1994) Criterion-referenced testing 30 years later: promise broken, promise kept. *Educational Measurement: Issues and Practice*, 4.
- Niemierko, B. (1990) *Pomiar sprawdzający w dydaktyce. Teoria i zastosowania*. Warszawa: PWN.
- Niemierko, B. (1991) *Między oceną szkolną a dydaktyką*. Warszawa: WSiP.
- Niemierko, B. (1998). *Warsztat pomiarowy współczesnego polonisty. Polonistyka*, 4.
- Niemierko, B. (1999) *Pomiar wyników kształcenia*. Warszawa: WSiP.
- Niemierko, B. (2000) Polonistyczne pożytki z pomiaru. Odpowiedź na wątpliwości p. Stanisława Bortnowskiego. *Polonistyka*, 5.
- Niemierko, B. (2002) *Ocenianie szkolne bez tajemnic*. Warszawa: WSiP.
- Niemierko, B. (2012) Jawne i ukryte błędy pomiaru dydaktycznego. W: B. Niemierko i M.K. Szmigel (red.) *Regionalne i lokalne diagnozy edukacyjne*. Kraków: PTDE.
- Niemierko, B. (2015) Czy psychologowie nam pomogą w diagnostyce edukacyjnej. Badanie ich postaw w tej roli. W: B. Niemierko, M.K. Szmigel (red.) *Zastosowania diagnozy edukacyjnej*. Kraków: PTDE.
- Nitko, A.J. (1980) Distinguishing the many varieties of criterion-referenced tests. *Review of Educational Research*, 3.
- Nosal, C. (1990) *Psychologiczne modele umysłu*. Warszawa: PWN.
- Okoń, W. (1967) *Podstawy wykształcenia ogólnego*. Warszawa: Nasza Księgarnia.
- Paluchowski, W. J. (2001) *Diagnoza psychologiczna. Podejście ilościowe i jakościowe*. Warszawa: Scholar.
- Patrzalek, T. (1998a) Na wszystko test – dobry nie jest. *Polonistyka*, 4.
- Patrzalek, T. (1998b) O strukturze testu czytania ze zrozumieniem. *Polonistyka*, 4.
- Reckase, M. D. (2010) NCME 2009 presidential address: What I think I know. *Educational Measurement: Issues and Practice*, 3.
- Resnick, L.B. (1973) Hierarchies in children's learning. A symposium. *Instructional Science*, 3.
- Standards for educational and psychological tests and manuals* (1996). Washington: American Psychological Association.
- Szaleniec, H. (red.) (2010) *Teoria wyniku zadania IRT. Zastosowania polskim systemie egzaminów zewnętrznych*. Warszawa: CKE.
- Szmigel, M.K., Szaleniec, H. (2019). Uczniowie na progu transformacji szkoły. W: B. Niemierko i M.K. Szmigel (red.) *Wspomaganie rozwoju kompetencji diagnostycznych nauczycieli*. Kraków: PTDE.
- Szyling, G. (2018) Ku różnorodności w ocenianiu szkolnym. W: B. Niemierko i M.K. Szmigel (red.) *Wspomaganie rozwoju kompetencji diagnostycznych nauczycieli*. Kraków: PTDE.
- Wojciszke, B., Rotkiewicz, M. (2018). *Homo nie całkiem sapiens. O automatyzmach myślenia, nadętych politykach, narzekaniu Polaków i pułapkach moralności*. Sopot: Smak Słowa.
- Wysocka, E. (2013). *Diagnostyka pedagogiczna. Nowe obszary i rozwiązania*. Kraków: Impuls.

- Zakrzewski, R., Wypych-Stasiewicz, A. (2017) Zwiększenie kompetencji studentów Wydziału Chemii Uniwersytetu Łódzkiego w zakresie sprawdzania arkuszy egzaminu maturalnego z chemii. W: B. Niemierko i M.K. Szmigel (red.) *Diagnozowanie umiejętności praktycznych w toku kształcenia i egzaminowania*. Kraków: PTDE.
- Znaniecki F. (1934) *The method of sociology*. New York: Farrer & Rinehart.