

Arleta POREBA-KONOPCZYŃSKA  
Ośrodek Doskonalenia Nauczycieli  
Zielona Góra

## MOJE REFLEKSJE NAD OCENIANIEM NA PODSTAWIE UKŁADU RZETELNOŚCI I TRAFNOŚCI WYNIKÓW SPRAWDZANIA ZAPROPONOWANEGO PRZEZ PETERA T. KNIGHTA

Problem rzetelności i trafności jest kluczowym zagadnieniem w procesie sprawdzania i oceniania osiągnięć uczniów. H. Muszyński uznaje za niezbędne w diagnozie edukacyjnej, aby jej odbiorca „miał zaufanie do tego, że jej wyniki nie powstały w sposób przypadkowy, [...] by [...] oddawała rzeczywistość te stany rzeczy, o których mowa, [...] spełniała [...] warunek komplementarności [...] i dokładności [...]” oraz aby „jej konstatacje były niezależne od postaw i oczekiwań tych, którzy jej dokonują” (Muszyński 1998, 22).

Według B. Niemiarki „rzetelność pomiaru jest zgodnością wyników, uzyskanych przez tych samych badanych w testowaniu dokonanym w różnym czasie wersjami równoległymi testu” (Niemiarko 1975, 21), informuje więc, na ile uzyskane wyniki są miarodajne. Problem rozbieżności w ocenianiu jest podnoszony przez wielu autorów. Jego analizą w związku z systemem egzaminowania we Francji zajmowali się G. Niozet i J. P. Caverni (Niozet, Caverni 1988). Ujawnili oni rażące uchybienia w tej dziedzinie oraz przeprowadzili rozległe studia nad możliwościami usuwania źródeł rozbieżności. W Polsce problematyka porównywalności ocen stanowi również przedmiot zainteresowania wielu autorów zarówno przedwojennych, jak i współczesnych (Szuman 1935; Racinowski 1966, 30–33; Pieter 1973, 94–95; Karolczak-Biernacka 1996; Setlak 1998). Prace nad uzyskaniem wysokiej porównywalności oceniania nasiliły się w związku z przygotowaniem do utworzenia systemu egzaminów zewnętrznych (Przychodzień 1997; Szaleniec, Szmigel 2001).

Trafność najogólniej można zdefiniować jako „możliwość zasadnego wnioskowania o osiągnięciach uczniów i podejmowania skutecznych działań na podstawie wyników tego pomiaru” (Niemiarko 1999, 173). Pojęcie oraz typologia trafności zostały zaproponowane w 1954 r. przez Technical Recommendation for Psychological Test and Diagnostic Techniques, przy czym wyróżniono cztery kategorie trafności: wewnętrzną, diagnostyczną, prognostyczną i teoretyczną (Magnusson 1981, 189). W kolejnych latach rozumienie pojęcia trafności podlegało ewolucji, wzbogacono je o możliwości stosowania testu, aspekt ekonomiczny testowania, użyteczność (Niemiarko 1975, 172; Niemiarko 1990, 159–163). Najnowsze ujęcie prezentuje S. Messick, rozszerza on znaczenie pojęcia trafności o funkcje aksjologiczne i społeczne oraz przedstawia typologię trafności w postaci macierzy postępującej (Messick 2000). Ze względu na doniosłość zagadnienia problemy dotyczące trafności oceniania stanowią przedmiot ożywionych dyskusji i analiz (Niemiarko 1995, Hasemann 1996, Zalewski 1998).

Między trafnością i rzetelnością wyników egzaminowania zgodnie z klasyczną teorią pomiaru dydaktycznego występuje zależność hierarchiczna (Niemierko 1990, 119). Rzetelne wyniki wzbogacone o stosowność pomiaru pozwalają uzyskać wysoką jego trafność. Rzetelność jest więc niezbędnym składnikiem trafności, ta z kolei jest wartością wyższego rzędu. Życie jednak przynosi więcej rozwiązań. Nie w każdym przypadku wysoka trafność warunkowania jest wysoką rzetelnością. Nie zawsze też należy dążyć do podwyższenia trafności pomiaru kosztem rzetelności jego wyników (Niemierko 1999, 169–170).

Ciekawe ujęcie zależności między trafnością i rzetelnością wyników oceniania zaproponował Peter T. Knight, który zrezygnował z hierarchizowania na rzecz analizy zależności (Knight, 2000, 44, por. tab. I w niniejszym referacie).

TABELA 1

	Wysoka rzetelność (szczególnie ważna w ocenianiu doniosłym, np. dyplomowaniu)	Niska rzetelność (typowa dla wartościowania złożonych działań lub prac artystycznych)
Wysoka trafność (ważna, gdy chodzi o czyjeś uprawnienia do fachowej praktyki)	<p>Ideał oceny szkolnej. Trafna, rzetelna ocena złożonych wyników (kompetencji, umiejętności, rozumienia i wiedzy; jest kosztowna)</p>	<p>W ocenianiu powszednim, gdy występuje względnie otwarty dialog między oceniającymi i ocenianymi, ograniczona rzetelność nie ma tak wielkiego znaczenia</p>
Niska trafność (niepożądana cecha oceny, lecz często niewątpliwa własność oceny uzyskanej z testu)	<p>Te oceny mogą być użyteczne, gdy są właściwie wykorzystane przez osoby znające ich ograni- czenia; wyniki testów inteligencji są koronnym przykładem</p>	<p>W potoku życia codziennego nieuchronnie dokonujemy oceniania o niskiej rzetelności i niskiej trafności. To nie znaczy, że ich pragniemy</p>

(wg P. T. Knight, Lancaster University, UK, *The Value of a Programme-wide Approach to Assessment*, „Assessment & Evaluation Higher Education” 2000, No 3)

Takie ujęcie daje większe możliwości interpretacyjne w odniesieniu do bogatej rzeczywistości edukacyjnej. Ma to szczególne znaczenie wówczas, gdy rozpatrujemy ocenianie z punktu widzenia systemu egzaminów doniosłych oraz egzaminów powszednich. Mimo że oba systemy powiązane są ideą przewodnią, wyrażoną w preambule „Podstawy programowej” (*Podstawa programowa kształcenia ogólnego*), to cele egzaminowania oraz możliwości organizowania sytuacji egzaminacyjnych są różne. Układ zależności zaproponowany przez Petera T. Knighta pozwala się przyjrzeć różnym sytuacjom egzaminacyjnym oraz dokonać ich analizy, zbliża też rozważania teoretyczne do „życia szkoły”, jego problemów i niedostatków. Moje refleksje dotyczące relacji trafności i rzetelności oceniania w ujęciu Petera T. Knighta ujęłam w formie diagramu logicznego.

TABELA 2

Relacje trafności i rzetelności w odniesieniu do egzaminów doniosłych i powszednich

	Rzetelność wysoka	Rzetelność umiarkowana	Rzetelność niska
Trafność wysoka	Perły profesjonalizmu pomiarowego	Nauczyciel — oświecony profesjonalista	Nauczyciel obdarzony intuicją pedagogiczną
Trafność umiarkowana	Egzaminy doniosłe	Nauczyciel — rzemieślnik stosujący pomiar dydaktyczny	Codzienna praktyka szkolna
Trafność niska	Pola minowe pomiaru	Pola minowe pomiaru	Pola minowe pomiaru

W tabeli zaciemniałam pola, na które zwracam szczególną uwagę. „Pola minowe pomiaru” to obszary zakazane. Odrzucam, jako niemoralną, możliwość stosowania w praktyce sprawdzania osiągnięć uczniów charakteryzującego się niską trafnością. Jeśli bowiem działania edukacyjne z definicji mają charakter celowy (Niemierko 2000, 5), zgodny z przyjętymi ideami wychowania oraz wizją człowieka wykształconego, sprawdzanie osiągnięć ucznia dotyczyć powinno spełnienia stawianych mu oczekiwań. Niska trafność zarówno treści nauczania, jak i sprawdzania osiągnięć prowadzi u młodzieży do zaburzenia istniejących schematów myślowych i wprowadzenia w to miejsce „chaosu”. Uczniowie w efekcie dysponują podwójnym repertuarem wiedzy: wyuczoną w szkole na pamięć wiedzą bierną oraz wiedzą zdroworoządkową, stosowaną na co dzień, często stanowiącą zespół mitów. Innym pospolitym zjawiskiem w naszych szkołach związanym z nietrafnymi sprawdzianami jest nabywanie przez uczniów postawy wyuczonej bezradności.

Niestety, rzeczywistość szkolna w dziedzinie oceniania jawi się dość ponuro. B. Niemierko porównuje ją do „dzielnicy slamsów” (1996, 13). Treści nauczania w programach szkolnych często nie są dostosowane do stadium rozwojowego uczniów. Nauczyciele nie informują uczniów o stosowanych wymaganiach edukacyjnych. Sprawdzanie często nie ma związku z nauczanymi treściami. Mocno też zakorzenił się w naszej szkole pogląd, że sprawdzanie osiągnięć to „łapanie złodzieja”. Dobry uczeń to niekoniecznie ten, który rozumie procesy, potrafi korzystać ze swojej wiedzy i samodzielnie ją rozwijać. Nazbyt często jest to osoba, która „nie da się złapać”. Podręcznik szkolny i nierzadko akademicki zna na pamięć wraz z przypisami, zna też odpowiedzi na wszystkie pytania w tak zwanych zbiorach testów oraz przekazywane „pocztą pantoflową” po odbytych egzaminach wstępnych czy olimpiadach przedmiotowych. Uspakajamy się, zwracając uwagę, że między

ocenianiem nauczycielskim a wynikami egzaminów na wyższą uczelnię występuje duża zbieżność. Czy zbieżność ta ma jednak związek z wysoką trafnością? Obawiam się, że jest raczej związana ze wspólnym popełnianiem tych samych błędów. Jaką bowiem informację o osiągnięciach ucznia mamy, gdy udzieli on odpowiedzi na pytanie:

„Zwierzęciem, którego żołądek może być wynicowany na zewnątrz jest:

- A. wypławek
- B. rozgwiazda
- C. mątwą
- D. jeżowiec
- E. nereida

(Testy z biologii dla licealistów i kandydatów na wyższe uczelnie)<sup>33</sup>

Jedynym sensem zaprezentowanego pytania jest wybranie spośród zdających tych osób, które posiadły pewną szczegółową, niewiele znaczącą dla rozumienia procesów i zależności biologicznych informację. Testy wstępne na uczelni wyższe oraz testy na olimpiadach przedmiotowych obfitują w podobne pytania, które „trafnie” pozwalają wyodrębnić tych uczniów, którzy dotarli do możliwie wielu szczegółów. Konstruktorzy testów nie martwią się o to, jakie znaczenie dla przyszłych losów młodego człowieka posiada taka wiedza. Bycie bardzo dobrym uczniem nie ma z reguły związku z uzyskiwaniem w przyszłości sukcesów.

Stosując nietrafne egzaminy powszednie czy doniosłe nie możemy się też spodziewać, że młodzież posiada cechy, o które zabiegają pedagodzy. Niewiele tu dadzą deklaracje i zapisy w dokumentach państwowych.

W tabeli zacieniowałam także pole, które nazwałam „perły profesjonalizmu pomiarowego”. W tym przypadku powód wyodrębnienia jest inny. Umiejętność pogodzenia wysokiej trafności i wysokiej rzetelności wyników sprawdzania to cel niedościgły. Jednakże uważam, że główne wysiłki organizatorów edukacji ucznia — od nauczyciela po konstruktorów standardów i testów egzaminacyjnych — powinny się skupić na tym właśnie celu. Im bardziej trafne i rzetelne są wyniki sprawdzania, tym większą wartość prezentują dla wszystkich osób, które z nich korzystają. Uczeń ma dzięki nim szansę kierowania własną nauką oraz właściwego zaplanowania swojej przyszłości, gdyż poznaje swoje możliwości oraz rodzaje popełnianych błędów. Nauczyciele mogą doskonalić własny warsztat metodyczny, gdyż takie wyniki stanowią wartościową podstawę ewaluacji ich pracy. Uczelnie wyższe mogą kształcić fachowców, gdyż egzaminy wstępne dokonują właściwej selekcji kandydatów, a wyniki egzaminów końcowych informują o posiadaniu przez studenta cech pożądanых w danym zawodzie. Zakłady pracy natomiast mogą dobrać kadrę posiadającą oczekiwane kompetencje, gdyż wyniki egzaminów są gwarancją ich posiadania.

Trudność pogodzenia wysokiej trafności i rzetelności wyników sprawdzania wiąże się z wewnętrzną sprzecznością między tymi miarami. Sprowadzenie sytuacji egzaminacyjnej do rozwiązywania krótkich, najlepiej zamkniętych zadań z poziomu wiedzy, sprzyja uzyskiwaniu wysokich rzetelności, lecz gubimy w tym momencie to, co jest najcenniejsze w uczeniu — jego twórczość; nie mamy wglądu w jego schematy poznawcze, nie potrafimy stwierdzić, czy potrafi on ze swojej wiedzy korzystać. Z drugiej strony, gdy sprawdzamy umiejętności złożone lub oceniamy efekty pracy twórczej uczniów, otrzymujemy wyniki charakteryzujące się niską rzetelnością. Działania, jakie zmuszony jest podejmować kon-

struktor testów, Peter T. Knight nazwał „handlem pomiędzy rzetelnością i trafnością” (Knight 2000, 43). W zależności od celu, jaki nam przyświeca, wybieramy raz trafność kosztem rzetelności, innym razem odwrotnie. Takie sytuacje prezentują w pozostałych okienkach tabeli, traktując je jako występujące lub możliwe do osiągnięcia w polskim szkolnictwie.

Szczególna dbałość o wysoką rzetelność wyników kosztem ich trafności charakteryzuje egzaminy doniosłe. Jest to związane z ich funkcją selekcyjną. Wyniki egzaminu gimnazjalnego określają horyzont możliwości wyboru przez ucznia szkoły ponadgimnazjalnej, a wyniki matury będą decydowały o przyjęciu na studia. Podkreśla się także diagnostyczną funkcję systemu egzaminów zewnętrznych. Będą one źródłem informacji dotyczących monitorowania postępów uczniów oraz jakości pracy szkół (MEN 1999; CKE 2000). Tych funkcji nie mogą pełnić egzaminy, których wyniki dają chwiejny obraz osiągnięć. Z tego powodu prowadzone są szeroko zakrojone prace na rzecz podniesienia porównywalności oceniania (Szalencic, Szmigel 2000; Brożek 2001). Do takich działań należą przede wszystkim: opracowywanie i udostępnianie z odpowiednim wyprzedzeniem standardów edukacyjnych, szkolenie szerokiej rzeszy egzaminatorów w zakresie sprawdzania kryterialnego, określenie ścisłych procedur i standardów organizowania sytuacji egzaminacyjnych. Trafność egzaminów doniosłych będzie jednak ograniczona. Zapisy w informatorach egzaminacyjnych i sylabusach dowodzą, że rezygnuje się ze sprawdzania niektórych umiejętności, uznanych za kluczowe dla rozwoju ucznia i ujętych w „Podstawie programowej”. Do takich umiejętności należą:

1. planowanie, organizowanie i ocenianie własnego uczenia się, przyjmowanie odpowiedzialności za własną naukę;
2. efektywne współdziałanie w zespole, budowanie więzi międzyludzkich;
3. poszukiwanie informacji, posługiwanie się technologią informacyjną;
4. rozwój osobistych zainteresowań;
5. przyswojenie sobie metod i technik negocjacyjnego rozwiązywania konfliktów i problemów społecznych.

Rezygnacja ze sprawdzania tych kompetencji mocno ogranicza użyteczność wyników. Planowana forma egzaminów jest także wynikiem kompromisu. Poza egzaminem maturalnym z języka polskiego i języków obcych wszystkie pozostałe będą egzaminami pisemnymi, w większości opartymi na zadaniach krótkiej odpowiedzi lub zadaniach zamkniętych. Taka sytuacja nie jest sprzyjająca dla wielu uczniów, którzy mają problemy z werbalizowaniem swoich myśli lub dla uczniów bardzo dobrych, dla których pytania rozszerzonej odpowiedzi wydają się bardziej odpowiednie. Sprawdzanie zadań zamkniętych lub zadań krótkiej odpowiedzi nie nastrocza większych problemów interpretacyjnych, co sprzyja podwyższaniu rzetelności. Zadania takie najczęściej jednak nie są odpowiednie do sprawdzania umiejętności złożonych czy strategii poznawczych uczniów.

Zgodnie z nową filozofią oceniania egzaminy powszednie mają przede wszystkim służyć wspieraniu rozwoju ucznia. Osłabia się ich funkcję selekcyjną na rzecz funkcji motywującej i monitorującej jego rozwój. Peter T. Knight twierdzi: „Prowadzi to do możliwości bardziej twórczego myślenia o ocenianiu tam, gdzie rzetelne stopnie nie są potrzebne. [...] wiele problemów zanika, [...] gdy na celu mamy usprawnienie uczenia się, a wszystkie pozostałe kwestie schodzą na plan dalszy”.



Wśród nauczycieli, podobnie jak w innych zawodach, można wyodrębnić grupę „artystów”, obdarzonych intuicją pedagogiczną, oraz grupę „rzemieślników”. Twórcze myślenie o ocenianiu dotyczy przede wszystkim tej pierwszej grupy. Intuicja pedagogiczna nauczyciela jest, moim zdaniem, gwarantem uzyskiwania wysokich trafności. Zdolność dotykania istoty problemów, świadomość roli własnego przedmiotu w kształtowaniu stosunku ucznia do świata i siebie samego pozwalają spojrzeć szerzej na procesy nauczania i oceniania, a przez to usytuować je w kontekście przyszłych potrzeb ucznia. Obraz osiągnięć staje się zintegrowany z wiedzą i umiejętnościami zdobywanymi na innych przedmiotach, a także spójny z doświadczeniami życia codziennego. Nauczyciel stwarza uczniowi różnorodne możliwości działania, pozwala mu odkryć własne zdolności, a także ograniczenia, aby tym skuteczniej stawiać im czoło. Sprawdzanie osiągnięć uczniów obejmujące nie tylko wypowiedzi ustne i pisemne, ale także różne formy samodzielnej pracy, pracę zespołową, wytwory działalności uczniowskiej oraz różne formy inicjatyw (np. organizacja debaty, teatrów, projektów pracy w środowisku) powoduje, że mają szansę ujawnienia swoich umiejętności uczniowie o różnych typach inteligencji. Daje to możliwość odkrywania talentów, motywuje do podejmowania wysiłku uczenia się. Sytuacje oceniania nie muszą wiązać się ze stresem, mogą natomiast wyzwalać aktywność uczniów, a także umożliwiać rejestrowanie rozwoju tych kompetencji, których ocenianie pominięte jest w egzaminach doniosłych.

Wzbogacenie wiedzy nauczycielskiej o dziedzinę pomiaru dydaktycznego powoduje, że ocenianie nabiera cech działań planowych. Planowanie procesu oceniania znacznie zwiększa informacyjność oceny, a także sprzyja podnoszeniu rzetelności wyników poprzez:

- zapoznanie uczniów na wstępie etapu edukacyjnego ze stawianymi im wymaganiami i kryteriami oceniania;
- stałość wymagań;
- planowanie zajęć tak, aby uczniowie mieli możliwość osiągania zamierzonych celów;
- sprawdzanie głównie tych umiejętności, które uznajemy za istotne;
- dostarczanie wartościowej informacji zwrotnej o osiągnięciach.

Stosowanie procedur pomiaru dydaktycznego w ocenianiu pełni ważne funkcje wychowawcze. Pozwala uczniom brać odpowiedzialność za własne uczenie się, planować je na odpowiednim dla siebie poziomie. Nauczyciele obdarzeni intuicją pedagogiczną, stosując pomiar dydaktyczny w swojej pracy, stają się „oświeconymi profesjonalistami”.

Codzienna praktyka szkolna oparta jest na ocenianiu intuicyjnym, pozbawionym elementów planowania, hołdującym wiedzy faktograficznej, nie doceniającym twórczości uczniów. Nauczyciele „kurczowo trzymają się” programów i podręczników, z których wiele — mimo zatwierdzenia przez MEN do użytku szkolnego — odbiega w swej treści od zapisów „Podstawy programowej”. W swojej codziennej pracy nauczyciele często nie mają świadomości, w jaki sposób uwarunkowania psychologiczne, stanowiące konsekwencję skądinąd normalnego funkcjonowania naszego umysłu (takie jak uleganie stereotypom czy korzystanie z pseudodanych), deformują ich oceny (Wojciszke 1996, Górska 2001). Wszystko to sprawia, że ocenianie szkolne z reguły charakteryzuje się niską rzetelnością i trafnością. Są to sytuacje nieuniknione, często spowodowane pośpiechem lub niedosta-

teczną wiedzą nauczyciela. Nie znaczy to, że należy się z nimi godzić. Moralnym obowiązkiem nauczycieli jest podejmowanie starań o podnoszenie jakości oceniania.

Podsumowując stwierdzam, że relacje między rzetelnością i trafnością mają uwarunkowania szersze niż proponowana przez klasyczną teorię pomiaru dydaktycznego hierarchiczna zależność. Przemyślenia takich autorów, jak S. Messick i P. T. Knight, stanowią pomost w kierunku innych interpretacji, z pewnością bardziej satysfakcjonujących.

## LITERATURA

- Brożek A., 2001, *Zadania i struktura zewnętrznego systemu oceniania*, [w:] K. Wenta (red.), *Pomiar edukacyjny jako kompetencje pedagogiczne*, Szczecin.
- Centralna Komisja Egzaminacyjna, 2000, *Ewaluacja upowszechniania informacji o tworzonim systemie oceniania zewnętrznego*, Warszawa (materiały seminaryjne).
- Górska L., 2001, *Stereotypy zachowań nauczycieli i „ukryty program” szkoły jako przeszkody w formułowaniu trafnej diagnozy edukacyjnej*, [w:] K. Wenta (red.), *Pomiar edukacyjny jako kompetencje pedagogiczne*, Szczecin.
- Hasemann K., 1996, *Rola oceniania osiągnięć szkolnych w edukacji*, [w:] „Edukacja” 4.
- Karolczak-Biernacka B., 1996, *Zagadnienie oceniania szkolnego*, [w:] „Edukacja” 4.
- Knight P. T., 2000, *Wartość ogólnoprogramowego podejścia do oceniania*, [w:] B. Niemierko (red.), *Pedagogika, pomiar dydaktyczny i ocenianie szkolne*, skrypt dla uczestników II Podyplomowego Studium Ewaluacji Dydaktycznej na Uniwersytecie Gdańskim, Gdańsk.
- Magnusson D., 1981, *Wprowadzenie do teorii testów*, Warszawa.
- Messick S., 2000, *Trafność*, [w:] B. Niemierko (red.), *Pedagogika, pomiar dydaktyczny i ocenianie szkolne*, skrypt dla uczestników II Podyplomowego Studium Ewaluacji Dydaktycznej na Uniwersytecie Gdańskim, Gdańsk.
- Ministerstwo Edukacji Narodowej, 1999, *O ocenianiu*, Warszawa.
- Muszyński H., 1998, *Diagnoza edukacyjna i jej miejsce w praktyce i teorii pedagogicznej*, [w:] B. Niemierko (red.), *Perspektywy diagnostyki edukacyjnej*, Gdańsk.
- Niemierko B., 1975, *Testy osiągnięć szkolnych. Podstawowe pojęcia i techniki obliczeniowe*, Warszawa.
- Niemierko B., 1990, *Pomiar sprawdzający w dydaktyce*, Warszawa.
- Niemierko B., 1995, *Matura jako pomiar dydaktyczny. Przyczynek do zespolonego sprawozdania z wizyty studyjnej w Londynie oraz projektu „Nowa Matura”*, Gdańsk (maszynopis).
- Niemierko B., 1996, *Między oceną szkolną a dydaktyką bliżej dydaktyki*, Warszawa.
- Niemierko B., 1999, *Pomiar wyników kształcenia*, Warszawa.
- Noizet G., Caverni J. P., 1988, *Psychologiczne aspekty oceniania osiągnięć szkolnych*, Warszawa.
- Pieter J., 1973, *Egzaminy obiektywne*, Warszawa.
- Podstawa programowa kształcenia ogólnego*, rozporządzenie MEN z 15 II 1999 r.
- Przychodzień B., 1997, *Kryteria ocenianie pisemnych prac maturalnych w STOLEM*, „Gdański Miesięcznik Oświatowy” 5.
- Racinowski S., 1966, *Problemy oceny szkolnej*, Warszawa.
- Setlak D., 1998, *Ocenianie szkolne — dylematy teorii i praktyki*, „Doradca. Biuletyn Oświatowy” 10.
- Szaleniec H., Szmigel K., 2001, *Z prac nad porównywalnością wyników oceniania zewnętrznego*, [w:] K. Wenta (red.), *Pomiar edukacyjny jako kompetencje pedagogiczne*, Szczecin.
- Szaleniec H., Szmigel K., 2001, *Egzaminy zewnętrzne, podnoszenie kompetencji nauczycieli w zakresie oceniania zewnętrznego*, Kraków.

- Szuman S., 1973, *Testy wiadomości*, „Encyklopedia Wychowania” I, 1935, za: J. Pieter, *Egzamin obiektywny*, Warszawa.
- Wojciszke B., 1996, *Korzystanie z danych i pseudodanych przy ocenianiu ludzi*, [w:] *Materiał wykładów sesji zimowej*, Gdańsk.
- Zalewski S., 1988, *Ocenianie, egzaminowanie i pomiar jakości pracy szkoły w kontekście reformy*, [w:] „Doradca. Biuletyn Oświatowy” 10.