

**prof. dr hab. Bolesław Niemierko**

SWPS Uniwersytet Humanistycznospołeczny w Sopocie

## **Ocenianie naturalne umiejętności praktycznych**

Ocenianie osiągnięć uczniów bywa przedmiotem ostrych napięć w klasie szkolnej i w mediach oraz solidnych studiów w instytucjach naukowych. Zdaje sobie sprawę z trudności uporządkowania tego pola w czytelny i użyteczny sposób, ale sądzę, że to jest możliwe.

Nie da się spełnić wszystkich uzasadnionych postulatów wobec oceniania szkolnego jednym rodzajem procedur, więc potrzebne jest ich rozróżnienie. Referat zawiera próbę wyodrębnienia i wyważenia roli dwu typów oceniania, które nazwę „naturalnym” i „regulowanym”. Obydwa są niezbędne w systemie kształcenia, ale nie dają się gładko połączyć.

### **Źródła tematyki referatu**

Susan Brookhart, światowy autorytet w zakresie oceniania osiągnięć uczniów, wraz z siedmioma innymi specjalistami tej dziedziny reprezentującymi łącznie sześć amerykańskich uniwersytetów, dokonała systematycznego przeglądu badań nad ocenianiem szkolnym w ciągu ostatnich stu lat (Brookhart i in., 2016). Raporty z tych badań podzieliła na pięć grup:

- 1. Wczesne studia nad rzetelnością ocen nauczycielskich (1912–1933).** Rzetelność ocen okazywała się niska lub bardzo niska, nawet poniżej wartości 0,5, zwłaszcza w grupie wyników najslabszych. Po upływie kilku tygodni nauczyciele inaczej oceniali te same prace pisemne. Pewną poprawę rzetelności oceniania można było uzyskać, wymuszając rozkład ocen zbliżony do krzywej normalnej. Podobne badania przeprowadzono w tych latach i w Polsce (Kryński, 1936; Pieter, 1969). Prowadziły do rezygnacji z traktowania ocen wewnątrzszkolnych jako wyników pomiaru i do przyznania im odrębnego statusu dydaktycznego, a także – w Stanach Zjednoczonych – do porzucenia skali stustopniowej, w której przedział ufności oceny wynosił aż 20–30 punktów, na rzecz skali pięciopunktowej (A, B, C, D, F), stosowanej do dzisiaj.
- 2. Oceny nauczycielskie a wyniki pomiaru osiągnięć poznawczych i pozapoznawczych (1925–2014).** Te badania były prowadzone na dużych i bardzo dużych próbach, z zastosowaniem baterii testów osiągnięć i zaawansowanych metod analiz statystycznych. Podobne badania wykonał w Polsce Instytut Badań Edukacyjnych, wykorzystując wyniki ogólnokrajowych egzaminów zewnętrznych. W Stanach Zjednoczonych średnia ocen szkolnych korelowała z wynikami standaryzowanych baterii testów osiągnięć absolwentów szkół średnich (SAT, ACT) w wysokości 0,50–0,60, co mogło świadczyć o umiarkowanej trafności diagnostycznej obu metod oceniania. Takie współczynniki

korelacji uznawano albo za świadectwo niepełnej zgodności tych metod co do mierzonej właściwości uczniów, albo za dowód nieuniknionego subiektywizmu oceniania nauczycielskiego.

Wielozmiennowe analizy danych wyodrębniły dwa równorzędne czynniki oceny nauczycielskiej: **poznawczy**, obejmujący wiadomości i umiejętności, oraz **pozapoznawczy** (*non-cognitive*), obejmujący motywację, pilność i nawyki pracy uczniów. To rozróżnienie ma silne podstawy we współczesnej psychologii poznawczej (Doliński, 2000). Zaangażowanie i wysiłek uczniów są zarówno warunkiem, jak i wynikiem kształcenia szkolnego, obserwowalnym na co dzień, ale nie w toku egzaminu zewnętrznego, przebiegającego w szczególnej mobilizacji do wysiłku. Ocenianie nauczycielskie jest pedagogicznie bogatsze, co wykazały badania podłużne, dowodząc, że na podstawie ocen nauczycielskich łatwiej można przewidzieć przyszłe sukcesy i porażki uczniów, w tym odpad na wyższych szczeblach kształcenia, niż na podstawie standaryzowanych baterii testów osiągnięć poznawczych.

3. **Badania poglądów i praktyk nauczycieli w zakresie oceniania** (1993–2016). Sondáže i wywiady pokazały, że aczkolwiek osiągnięcia poznawcze liczą się najbardziej, większość nauczycieli włącza do oceny – w różnych kombinacjach – takie czynniki jak: wysiłek (*effort*), prace domowe, postęp, postawę i zachowanie. Wysiłek jest wyżej ceniony u mniej zdolnych uczniów, w przedmiotach humanistyczno-społecznych i przez nauczycieli mężczyzn, a nauczycielki częściej nagradzają stopniami dobre zachowanie. Różnice indywidualne między pedagogami w zakresie filozofii oceniania i jej praktycznych wdrożeń są duże i trudne do przewidywania (Szyling, 2011).
4. **Ocenianie oparte na standardach wymagań** (2003–2013). W tych badaniach przeważa wywiad z nauczycielami, uczniami i ich rodzicami. Nadzieje na wysoką trafność wewnętrzną takiego oceniania spełniają się tylko częściowo, bo niektórzy uczniowie znacznie przekraczają standardy, a wielu uczniów – przy rozsądnej normie ilościowej, 50% lub wyższej – nie osiąga poziomu uznawanego za podstawowy. Nauczyciele i rodzice uczniów chwalą jednak pomiar sprawdzający, a jego teoria rozwija się także i w Polsce (Niemierko, 1990; Niemierko, 2002).
5. **Ocenianie na wyższych uczelniach** (1972–2009). Na tym szczeblu kształcenia przeprowadzono wiele eksperymentów. Korzystając ze swej autonomii, uczelnie wyższe wprowadzały systemy oceniania bardzo rozmaite między krajami i wewnątrz krajów. Stosowano skale długie, nawet stu-stopniowe, i krótkie, pięciostopniowe. Próbowano oceniać „według krzywej” i, mimo dyskredytacji tego modelu już pół wieku temu, to założenie do dziś pozostało „milczącym partnerem” wielu systemów. Tym większą konsternację budzi żywiołowy wzrost ocen (*grade inflation*) przypisywany postępom psychologii humanistycznej i wpływom politycznym. Sprawdzone, czy ewaluacja kursów przez studentów wymusza wysokie oceny studentów, i ustalono, że przewidywane oceny nisko, w granicach 0,10–0,30, korelują z zapowiadanyimi wynikami takiej ewaluacji.

Podsumowując tę metaanalizę dorobku stulecia, możemy stwierdzić, że badania nad ocenianiem szkolnym odbyły długą drogę od dyskwalifikacji oceniania wewnątrzszkolnego w roli pomiaru osiągnięć uczniów do przyznania mu rangi wielowymiarowego procesu decyzyjnego i szerokiego pola badań psychologiczno-pedagogicznych. Raport Brookhart zamyka następujący akapit:

Żaden plan badań naukowych nie wyeliminuje całkowicie różnic ocen między nauczycielami. Mimo to autorzy tego przeglądu zalecają parę ścieżek postępu. Badanie ocen w szerszym kontekście nauczania i oceniania pomoże skupić uwagę na istotnych źródłach i przyczynach nietrafnych lub nierzetelnych decyzji o ocenie. Badanie sposobów bardziej skutecznej, powszedniej i naturalnej indywidualizacji kształcenia zmniejszy nauczycielskie poczucie presji, by promować uczniów, którzy dokładają starań, ale nie uzyskują oczekiwanego poziomu osiągnięć. Badanie wielowymiarowego pojęcia „sukcesu szkolnego” uświadomi nam, że stopnie mierzą coś znaczącego, czego nie mierzą testy osiągnięć szkolnych. Badanie sposobów pomagania nauczycielom w formułowaniu lub dobieraniu kryteriów, a potem komunikowaniu uczniom i stosowaniu tych kryteriów wobec pracy uczniów, poprawi jakość oceniania. To wszystko wydaje się realne w następnym stuleciu badań nad ocenianiem. To wszystko z pewnością podniesie trafność, rzetelność i bezstronność oceniania szkolnego (Brookhart i in., 2016, s. 836–837).

Tematyka **XXIII Krajowej Konferencji Diagnostyki Edukacyjnej** sprzyja refleksji nad rozwojem teorii oceniania szkolnego. Diagnozowanie umiejętności praktycznych kieruje naszą uwagę na realne kompetencje uczniów, ich wspomaganie oceną bieżącą i poświadczenie oceną końcową. Ani **różnicowanie** osiągnięć w konwencji psychometrycznej, ani ich **sprawdzanie** oparte na standardach edukacyjnych nie zapewniają w pojedynkę w pełni trafnej diagnozy takich umiejętności (Niemierko, 1999). Warto zastanowić się głębiej nad ograniczeniami tych metod.

Na nadużycia testów „papierowych” w diagnostyce umiejętności praktycznych zwrócił mi uwagę dr Henryk Szaleniec, prezes naszego Towarzystwa, prosząc o ustosunkowanie się do hasła „oceniania autentycznego” (*authentic assessment*), głoszonego przez zapalczywych reformatorów egzaminowania, przeciwników zadań wyboru wielokrotnego (Wiggins, 1998). Przy dużym uznaniu dla **autentyzmu uczenia się**, jego naturalności (Niemierko, 2007, rozdz. 4), nie mogłem zgodzić się na potraktowanie testowania pisemnego jako „nieautentycznego” i na ograniczenie zastosowań standaryzowanych baterii testów osiągnięć poznawczych. Tak oto zarysowała się koncepcja tego referatu.

### Umiejętności praktyczne sprawdzane pośrednio

**Umiejętność** jest sprawnością osiągnięcia celu. Jest tym bardziej praktyczna, im bardziej użyteczna w życiu, w edukacji, w pracy zawodowej. Oceniana **bezpośrednio** wymaga sytuacji, w której wszystkie etapy czynności i ich pełny wynik mogą być obserwowane, a oceniana **pośrednio** przebiega w symulacji, uproszczeniu do zjawisk uznanych za istotne. **Wysoka symulacja** jest ograniczeniem czynności do jej intelektualnego odtworzenia, jej postać naturalna lub laboratoryjna jest zastąpiona zadaniem tekstowym lub działaniem na symbolach.

Ocenianie bezpośrednio umiejętności praktycznych jest droższe, bardziej czasochłonne, trudniejsze organizacyjnie, a bywa, że nie dość bezpieczne dla uczniów samodzielnie wykonujących złożone czynności (Finch, 1991). Z tych przyczyn, a często po prostu dla wygody, bywa zastępowane pośrednim ocenianiem umiejętności praktycznych, w werbalnym kontakcie pisemnym, w tym internetowym. Przypomnijmy zjadliwą krytykę takiego postępowania w anegdocie Roberta Magera (1900) przytoczonej w podręczniku *Pomiar wyników kształcenia* (Niemierko, 1999, s. 111):

Przypuśćmy, że dałeś z siebie wszystko na kursie, by osiągnąć cel, jaki nauczyciel sformułował pierwszego dnia zajęć: *Umieć przejechać na rowerze jednokołowym 100 metrów bez upadku ulicą z krawężnikami*. Gdy już opanowałeś, po wielkich trudach, jazdę na odległość 200 metrów, otrzymujesz – w przykładzie Magera – następujące zadania egzaminacyjne:

1. Zdefiniuj rower jednokołowy.
2. Napisz krótkie wypracowanie na temat historii roweru jednokołowego.
3. Podaj nazwy co najmniej sześciu części roweru jednokołowego.
4. Opisz swój sposób wsiadania na rower jednokołowy.

Jak czułbyś się, gdyby ci kazano czegoś się uczyć, a sprawdzano coś innego? – pyta retorycznie Mager. A oto możliwe „wyjaśnienia” tej sytuacji przez nauczycieli:

*Brak nam sprzętu do sprawdzania osiągnięć praktycznych!*

*Nie mamy dość rowerów jednokołowych dla wszystkich!*

*To jest szkoła, a nie klub sportowy!*

*Nieważne, jak wam idzie jazda – jeśli nie poznacie roweru, nie będziecie go szanować!*

*Zbyt łatwo jest nauczyć się jeździć – potrzebne są także trudniejsze zadania!*

*Gdyby każdy nauczył się jeździć, musiałbym dać wszystkim piątki!*

*W teście muszą być różnorodne zadania, by nie był nudny!*

*Chcę, by uczniowie uczyli się czegoś z moich testów!*

*Kształcę myślenie twórcze i wartości duchowe!*

*Te testy muszą nadawać się do automatycznego punktowania!*

*Uczniowie muszą uczyć się przez odkrywanie!*

Bez względu na to, ile prawdy i fałszu jest w powyższych komentarzach, pozostaje faktem, że nie da się określić, czy ktoś umie jeździć na rowerze bez zaobserwowania, że ten ktoś jeździ na rowerze – stwierdza Mager. W przytoczonych komentarzach nauczycieli można dopatrzeć się jednak także argumentów za rzekomą wyższością zintelektualizowanego oceniania pośredniego nad bezpośrednim ocenianiem konkretnych umiejętności praktycznych.

## Postępy psychologii i dydaktyki

W stuleciu analizowanym przez Brookhart dokonała się zasadnicza zmiana w psychologicznych podstawach dydaktyki, w tym w podstawach oceniania szkolnego. Mieścił się w nim rozkwit **behawioryzmu**, doświadczalnej nauki o zachowaniu z edukacyjnym szczytem w postaci nauczania programowanego, i jego zmierzch wywołany postęпами psychologii humanistycznej i psychologii poznawczej, wykraczających poza obiektywistyczne schematy przyrodznawstwa.

**Psychologia humanistyczna** wiąże psychologię z filozofią, gdyż akcentuje podmiotowość, rozwój osobowości i samorealizację człowieka, a humanistyczne systemy pedagogiczne są oparte na zaspokajaniu potrzeb dzieci i młodzieży. Wraz z **psychologią poznawczą**, skoncentrowaną na badaniu organizacji wiedzy w umyśle człowieka, dają podstawę **konstruktywizmowi**, przyznającemu uczniom prawo do tworzenia własnych znaczeń poznawanej rzeczywistości (Shepard, 2000; Klus-Stańska, 2002). *To uczeń decyduje, które aspekty materiału są dla niego ważne, a które nie* – czytamy w konstruktywistycznych deklaracjach (Soltis i Phillips, 2003, s. 103). Przy takim założeniu racjonalność jednolitych standardów i testowych norm empirycznych mocno słabnie. Uczniowie zyskują prawo do uczenia się i porządkowania wiedzy na swój własny autorski sposób.

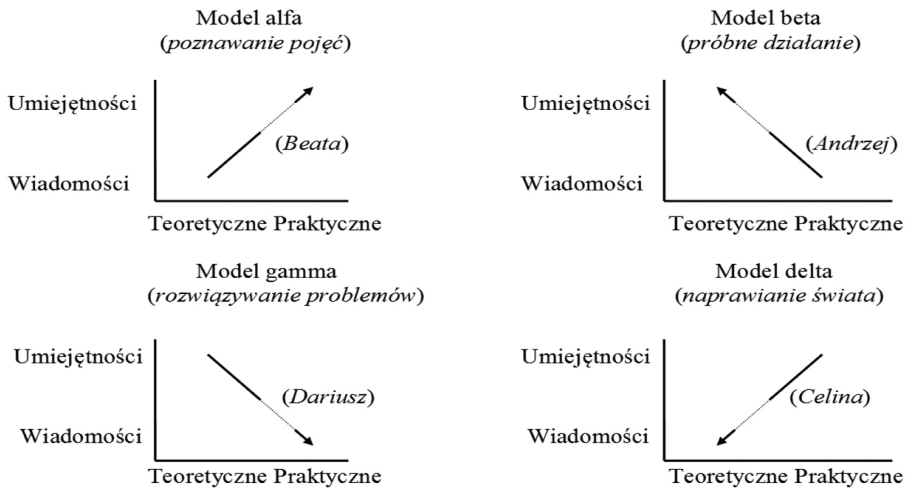
O dominacji takiego podejścia świadczy fakt, iż obecny przewodniczący Krajowej Rady Pomiaru Edukacyjnego (NCME), wiodącego w USA pedagogicznego stowarzyszenia pomiarowego, Mark Wilson, poświęcił pięć kolejnych listów do członków stowarzyszenia ocenianiu nauczycielskiemu (*classroom assessment*), oddolnemu (*bottom up*), a nie odgórnemu (*top-down*), rejestrującemu „trajektorie uczenia się”, budującemu „mapy drogowe” (*road maps*) postępu uczniów i uwalniającemu szkoły od *błędnego trójkąta, w którym plany nauczycieli co do uczenia się ich uczniów są ściśnięte między programem kształcenia a ocenianiem sumującym prowadzonym dla celów ewaluacyjnych*.

Za postępami psychologii podąża – znacznie wolniej i nie bez zahamowań – współczesna dydaktyka. Indywidualizacja kształcenia od dawna jest postulatem w naszym systemie pedagogicznym (Lewowicki, 1977), ale do pełnej indywidualizacji oceniania szkolnego mamy w Polsce jeszcze dość daleko, a wprowadzenie testowych egzaminów zewnętrznych jeszcze bardziej nas od niej oddaliło.

Z szansę na uporządkowanie teorii oceniania szkolnego daje nam **teoria kształcenia wielostronnego**, sformułowana przed półwieczem przez Wincentego Okonia (Okoń, 1967). Obejmuje ona cztery **drogi uczenia się**, wywiedzione z dorobku wcześniejszych pokoleń pedagogów. Oto one:

1. *Przez przyswajanie* – tą drogą opanowujemy głównie opisy rzeczywistości (np. budowy wszechświata i Ziemi, przebiegu procesów fizycznych, wyglądu i trybu życia organizmów żywych, doniosłych wydarzeń społecznych, treści dzieł nauki i kultury),
2. *Przez odkrywanie* – tą drogą opanowujemy głównie wyjaśnienia prawidłowości zjawisk (występowania, porządku, przyczynowości, praw współdziałania),
3. *Przez przeżywanie* – tą drogą opanowujemy głównie ocenianie wartości (znaczenia przyrody i społeczeństwa, dzieł nauki i kultury, własnej i cudzej pracy),
4. *Przez działanie* – tą drogą opanowujemy głównie umiejętności praktyczne (obserwacji i przewidywania zjawisk, czytania i pisanie, interpretacji matematycznej, pracy ręcznej i posługiwania się narzędziami, planowania i wykonywania działań).

W zgodzie z teorią kształcenia wielostronnego wyróżniono cztery **modele uczenia się**, przedstawione na rysunku 1 (Niemierko, 2009, s. 16–22):



Rysunek 1. Cztery modele uczenia się

Modele *alfa* i *beta*, czerpiące z wiadomości, bardziej nadają się do zastosowania w zobiektywizowanych egzaminach, podczas gdy modele *gamma* i *delta*, akcentujące umiejętności, prowadzą ucznia do oryginalnych struktur wiedzy, niedających się przewidzieć w programach i testach, lecz dających się ocenić intuicyjnie. To, co wiedzą *Beata* i *Andrzej*, może być wyszczególnione **analitycznie**, ale to, co umieją *Celina* i *Dariusz*, trzeba ująć **holistycznie**, całościowo.

Uczenie się oparte na wiadomościach przynosi wyniki ilościowe, podlegające sumowaniu i uśrednianiu oraz porównaniom ważnym dla administracji oświatowej. Uczenie się oparte na umiejętnościach owocuje jakością, której dalsze kształtowanie jest celem ewaluacji. To rozróżnienie opiera się na fundamentalnej typologii Michaela Scrivena, który przeciwstawił **ewaluację sumującą**, dokonywaną na podstawie końcowych wyników kształcenia, **ewaluacji kształtującej** (*formative*), dokonywaną na bieżąco w toku kształcenia (Scriven, 1967). Typologię Scrivena przeniesiono wkrótce do oceniania osiągnięć uczniów (*achievement assessment*), gdzie zrobiła ogromną karierę, owocując licznymi badaniami, podręcznikami i kursami dla nauczycieli (Stróżyński, 2011).

Konstelacjami pojęć określających cele i metody oceniania szkolnego zajął się Anthony Nitko, przygotowując wytyczne systemowo-edukacyjne dla krajów rozwijających się (Nitko, 2004; Nitko, 2008). Zauważył, że w tych krajach:

Pedagodzy wszystkich szkół są wzburzeni faktem, że na pewnych szczeblach drabiny szkolnej polega się na pojedynczym „doniosłym” (*high stakes*) egzaminie, ignorując osiągnięcia wykazywane przez uczniów w toku wielu lat nauki w klasie szkolnej.

Szeroki zakres i bogactwo treści zreformowanych programów nie obchodzą nauczycieli, którzy na własną rękę zwiężają program do tych zagadnień, które zapewne zjawiają się na egzaminie (Nitko, 1998, s. 84).

Wytyczne nie sugerowały likwidacji egzaminów państwowych i nie podważały zaufania do ich wyniku. Ich celem było zapewnienie równowagi między wewnętrznymi i zewnętrznymi wobec szkoły źródłami informacji o osiągnięciach uczniów. Te informacje mogą zasadniczo się różnić, prowadzić do odmiennych strategii i decyzji ważnych dla ucznia, szkoły, społeczeństwa i gospodarki kraju.

## Dwa paradygmaty oceniania

Silny kontrast dwu podejść, a zarazem ich wewnętrzna spójność, pozwalają na wprowadzenie do teorii oceniania dwu pojęć:

1. **Ocenianie naturalne** w edukacji jest tworzeniem obrazu osiągnięć ucznia w celu wspierania jego rozwoju. Jest znane od zarania dziejów edukacji, poparte współczesną psychologią i dydaktyką.
2. **Ocenianie regulowane** w edukacji jest określaniem poziomu osiągnięć ucznia według norm umożliwiających porównania. Jest pochodną **psychometrii**, metodologii pomiaru psychologicznych różnic indywidualnych, stosowaną w edukacji masowej.

Założenia dwu podejść, tworzące odrębne **paradygmaty naukowe**, czyli zbiory założeń, pojęć i metod wybranej dziedziny przedstawia tabela 1. W tabeli wyróżniono tłustym drukiem te właściwości, które stanowią o niezbędności danego typu oceniania w systemie edukacyjnym.

Tabela 1. Paradygmaty oceniania naturalnego i regulowanego

Założenia	O c e n i a n i e	
	naturalne	regulowane
A. metodologiczne 1. Strategia badawcza 2. Kontrola procedur 3. Obiektywizm wyniku	<b>1. Holistyczna</b> , rozpoznawanie struktur 2. Kryterialna, w przewadze intuicyjna 3. Niski, uwarunkowany przekonaniem egzaminatora	1. Analityczna, rozpoznawanie elementów <b>2. Empiryczna</b> , w przewadze statystyczna <b>3. Wysoki</b> , niezależny od przekonań egzaminatora
B. psychologiczne 1. Osobowość ucznia 2. Procesy motywacyjne 3. Procesy poznawcze	1. Humanizm, wzór oparty na potrzebach ucznia 2. Nastawienie <b>rozwojowe</b> 3. Konstruktivism, podmiotowe budowanie wiedzy	1. Funkcjonalizm, wzór oparty na potrzebach społecznych 2. Nastawienie przystosowawcze 3. Behawioryzm, utrwalanie reakcji na bodźce
C. dydaktyczne 1. Treść uczenia się 2. Modele uczenia się 3. Typ ewaluacji	<b>1. Umiejętności twórcze</b> i strategię poznawczą na pierwszym planie 2. Przez odkrywanie i przeżywanie <b>3. Kształtująca</b> , wytyczająca wzrost osiągnięć	1. Znajomość faktów i umiejętności odtwórcze na pierwszym planie 2. Przez przyswajanie i działanie <b>3. Sumująca</b> , rejestrująca stan osiągnięć

Założenia	O c e n i e	
	naturalne	regulowane
D. pomiarowe		
1. Układ odniesienia	1. Program kształcenia w ujęciu autorskim, <b>sprawdzanie</b>	1. Osiągnięcia uczniów w populacji generalnej, <b>różnicowanie</b>
2. Typowe formy zadań	2. Próba pracy, nisko symulowane, rozszerzonej odpowiedzi	2. Wyboru wielokrotnego, krótkiej odpowiedzi, wysoko symulowane
3. Organizator oceniania	3. Nauczyciel przedmiotu, pracodawca	3. Komisja państwowa, instytut naukowy, agencja oświatowa
E. interpretacyjne		
1. Rzetelność wyniku	1. Umiarkowana lub niska,	<b>1. Wysoka</b> , uzyskana przez standaryzację
2. Trafność wyniku	2. Umiarkowana, poszerzony zakres treści	2. Umiarkowana, zwężony zakres treści
3. Użytkownicy wyniku	3. Uczeń i nauczyciel, rodzic, pracodawca	3. Uczeń, nauczyciel, szkoła wyższego szczebla, nadzór pedagogiczny

**Ocenianie naturalne**, dokonywane w procesie kształcenia, akcentuje to, co jest różne i niespodziewane w uczeniu się i w osiągnięciach uczniów. Jego trafność, uzyskiwana zwykle kosztem rzetelności, polega na wykrywaniu „mocnych stron” każdego ucznia. Prowadzi do indywidualizacji kształcenia, twórczości programowej i metodycznej. Wzmacnia współdziałanie uczniów z nauczycielami, wytwarza więzi emocjonalne i bodźce motywacyjne sprzyjające rozwojowi (Dweck, 2013). Te niewątpliwe walory pedagogiczne oceniania naturalnego stanowią o jego swoistym immunitecie: chronią go przed ingerencją z zewnątrz, ujmowaniem w sztywne ramy w celu dokonywania porównań. Zarazem jednak pojawia się zagrożenie trafności ocen niezadawalającą kompetencją przedmiotową oceniającego i jego skłonnością do emocjonalnych reakcji na zachowania ucznia.

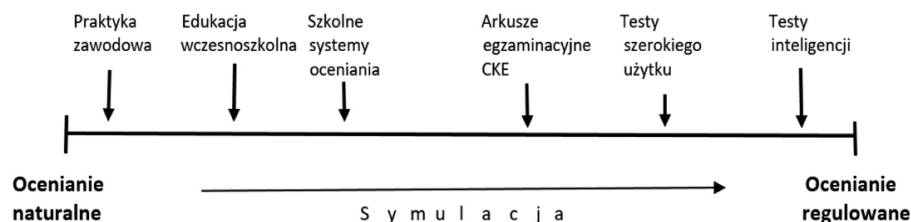
**Ocenianie regulowane**, dokonywane w wyznaczonym terminie, akcentuje to co wspólne i przewidziane programem kształcenia. Jego wzorem są nauki ścisłe, obfitujące w procedury badawcze. Dzięki obiektywizmowi dobrze służy uzasadnianiu decyzji wobec uczniów, nauczycieli i placówek oświatowych, ale łatwo przepadza się w uciążliwą biurokrację i nadregulację. Rzetelność takiego oceniania jest znacznie wyższa niż jego trafność, sprowadzona do szczegółowo zaplanowanego układu czynności ucznia i egzaminatora. Szeroki odbiór informacji o wynikach oceniania regulowanego obfituje, niestety, w nadinterpretację (Szaleniec, 2004) i nie sprzyja spontanicznej **samoregulacji** ucznia, nauczyciela i szkoły (Ledzińska i Czerniawska, 2011, rozdz. 3). Może wywołać zwężenie programu kształcenia do przewidywanej treści i formy egzaminu (Nitko, 1968).

Błędem byłoby traktowanie dwu typów oceniania szkolnego jako wykluczających się kategorii. Ocenianie naturalne zyskuje na formułowaniu **kryteriów**, czyli zasad wartościowania osiągnięć, i nawet, choć to ryzykowne, wyznaczaniu ich względnej wagi. Ocenianie regulowane zyskuje na symulowaniu rzeczywistych (*real-life*) sytuacji i wydarzeń oraz pośrednim, nawet tylko werbalnym, odtwarzaniu umiejętności. Rozsądne jest zatem traktowanie tych dwu procedur oceniania jako **typów idealnych**, niespotykanych na co dzień w czystej formie.



O potrzebie mieszanych modeli oceniania szkolnego świadczy koncepcja **oceniania orientującego** (*interim assessment*), etapowo sumującego, rozumianego jako wartościowanie wyników uczenia się ze względu na cele, do których jednostka lub grupa zmierza różnymi drogami i w różnym tempie (Perie i in., 2009; Niemierko, 2010). Ocenianie orientujące wiąże sumowanie z kształtowaniem osiągnięć i według jego zwolenników zachowuje zalety obydwu rodzajów ewaluacji. Dodać można, że – podobnie jak wiele innych pojęciowych hybryd – jest bliskie tradycjom metodycznym szkolnictwa.

Z dr. Henrykiem Szaleńcem uzgodniliśmy, że typologię oceniania szkolnego będziemy przedstawiać jako **kontinuum** rozpięte między całkowitą naturalnością i pełną regulacją, którego elementami są wszystkie kombinacje tych dwu właściwości uporządkowane według rosnącej symulacji naturalnych procesów i zjawisk. Ilustracją tej koncepcji jest rysunek 2.



Rysunek 2. Kontinuum składników oceniania szkolnego i hipotetyczne pozycje sześciu form diagnozy edukacyjnej

Na końcach wyznaczonego odcinka, bliskie typów idealnych są praktyki zawodowe, pełne życia i tylko z grubsza ustrukturuwane regulaminowo oraz badanie ucznia klasycznym testem inteligencji logiczno-matematycznej (Ravena), dokonujące się w świecie symboli i figur geometrycznych. Warto jednak zaznaczyć, że postęp w psychometrii dokonuje się w kierunku inteligencji wielorakich, bliskich szkolnych przedmiotów kształcenia (Gardner, 2002), a więc oceniania naturalnego. Edukacja wczesnoszkolna i szkolne systemy oceniania są nasycone obserwacją zachowania ucznia z dużą rolą aspektu emocjonalno-motywacyjnego. Arkusze egzaminacyjne CKE, bywa że kilkudziesięciostronicowe, są coraz bogatsze w wierną symulację wydarzeń społecznych i eksperymentów przyrodniczych, co skutkuje ogromnym obciążeniem egzaminatorów, ale podnosi naturalność. Wydawcy popularnych testów szerokiego użytku muszą dbać o sprawność ich stosowania przez nie-specjalistów, więc chętniej stosują zamknięte formy zadań.

Przed problemem wyboru typu oceniania często staje nauczyciel szkolny i akademicki. By zachęcić uczestników Konferencji do podjęcia tego problemu, podam przykład z przedmiotu *Psychologiczne podstawy edukacji*, prowadzonego na Uniwersytecie SWPS dla studentów niestacjonarnych zainteresowanych pracą psychologa szkolnego. W jego sylabusie czytamy:

Przedmiot ma ułatwić wykorzystanie wiedzy psychologicznej w działalności diagnostycznej i doradczej w obecnych i przyszłych systemach edukacyjnych.

Sformułowanie celu wskazuje na **umiejętności praktyczne** studenta, jego zdolność do rozpoznawania potrzeb rozwojowych uczniów oraz doradzania uczniom, rodzicom i nauczycielom na podstawie swojej wiedzy psychologicznej. W połowie tegorocznego kursu tego przedmiotu przedstawiłem studentom – na zasadzie ewaluacji kształtującej – następujące zadania:

W renomowanym liceum prywatnym młodzież zainteresowała się ekonomią i powiedziała o tym dyrektorowi, profesorowi pedagogiki. Bardzo dobrze – ucieszył się dyrektor – znam kilku profesorów ekonomii, którzy chętnie przyjdą do was z prelekcjami. Ależ nie – zaprotestowali uczniowie – my chcemy poznać paru młodych przedsiębiorców, którym powiodło się w interesach!

**Zadanie 1.** Jak wytłumaczysz tę różnicę zdań? Sięgnij do swojej wiedzy psychologicznej i pedagogicznej.

**Zadanie 2.** Jak, twoim zdaniem, powinien zachować się dyrektor liceum? Zaprojektuj i uzasadnij plan jego działania w tej sprawie.

Studenci podawali różnorodne rozwiązania tych zadań, w większości popierając inicjatywę licealistów, ale czyniąc z przedsiębiorców jedynie asystentów profesorów ekonomii. Wyróżniły się grupy osób mających doświadczenie biznesowe, doświadczenie edukacyjno-doradcze, a także znajomość liceów niepublicznych. Najlepsze prace nawiązywały do wybranych kierunków psychologii i do prawidłowości pedagogicznych objętych wykładem. Były, moim zdaniem, w pełni twórcze.

Wydarzenie było prawdziwe, opowiedziane mi przez dyrektora szkoły, a więc uważam je za równie naturalne jak te, które są przedstawiane w arkuszach CKE. Zadania byłyby jeszcze bliższe pracy psychologa, gdyby studenci brali udział w działaniach na terenie szkoły, rozmawiali z młodzieżą i wspólnie projektowali nowy przedmiot lub cykl zajęć dla zainteresowanych ekonomią.

Czy można było zastąpić pisemną pracę twórczą studenta zbiorem zadań zamkniętych lub krótkiej odpowiedzi na temat funkcji psychologa szkolnego, autorytarnych systemów edukacji, pokoleń X – Y – Z, modeli uczenia się, roli motywacji do rozwoju i podobnych? Nie sądzę, choć byłoby to bardziej zobjektywizowane i wygodniejsze dla wykładowcy.

### Podsumowanie referatu

Oto podsumowanie rozważań zawartych w tym referacie:

1. Współczesne badania nad ocenianiem szkolnym pokazują jego wielowymiarowość.
2. **Ocenianie naturalne** wspiera rozwój ucznia według jego własnych dyspozycji.
3. **Ocenianie regulowane** obejmuje osiągnięcia uczniów oczekiwane programem nauczania.
4. Behawioryzm uzasadnia ocenianie regulowane, a konstruktywizm – ocenianie naturalne.
5. Regulacja podnosi obiektywizm, a naturalność polepsza odbiór ocen szkolnych.

6. Obydwa typy oceniania, naturalne i regulowane, są niezbędne w systemie dydaktycznym.
7. Umiejętności praktyczne ucznia nie mogą być ocenione w sposób ściśle uregulowany.
8. Potrzebujemy analiz i badań porównawczych nad trafnością dwu typów oceniania w poszczególnych przedmiotach szkolnych i na różnych szczeblach systemu kształcenia.

\* \* \*

W referacie uporządkowałem kwestie znane pedagogom od dawna. Zakończę temat cytatem z podręcznika *Pomiar dydaktyczny dla nauczyciela w klasie*, wydane w Stanach Zjednoczonych pół wieku temu:

Stopnie są przyznawane na podstawie jednego lub więcej z następujących rozważań: (1) osiągnięcia każdego ucznia w stosunku do **reszty klasy**, (2) osiągnięcia każdego ucznia w stosunku do **jego zdolności**, (3) osiągnięcia każdego ucznia w stosunku do przyjętego **standardu osiągnięć**. Zauważcie, że w każdym przypadku osiągnięcia są oceniane przez porównanie z czymś innym. Te podstawy nie wykluczają się nawzajem, a przesadne poleganie na którymkolwiek z nich może prowadzić do śmiesznych (ridiculous) praktyk (Smith i Adams, 1966, s. 195, wyróż. – BN).

... i do szkodliwych praktyk – powinniśmy, moim zdaniem, dodać.

## Bibliografia

- Brookhart S.M., T.R. Guskey, A.J. Bowers, J.H. McMillan, J.K. Smith, L.F. Smith, M.T. Stevens, M.E. Welsh (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.
- Doliński D. (2000) Emocje, poznanie i zachowanie. W: J. Strelau (red.) *Psychologia. Podręcznik akademicki. T. 2. Psychologia ogólna*. Gdańsk: GWP.
- Dweck C.S. (2013) *Nowa psychologia sukcesu*. Warszawa: Muza.
- Gardner H. (2002) *Inteligencje wielorakie. Teoria w praktyce*. Warszawa: Media Rodzina.
- Klus-Stańska D. (2002) *Konstruowanie wiedzy w szkole*. Olsztyn: Wyd. Uniwersytetu Warmińsko-Mazurskiego.
- Kryński M. (1936) *Ocenianie pisemnych prac szkolnych*. Warszawa: Nasza Księgarnia.
- Ledzińska M, Czerniawska E. (2011) *Psychologia poznawcza. Ujęcie poznawcze*. Warszawa: WN PWN.
- Lewowicki T. (1977) *Indywidualizacja kształcenia. Dydaktyka różnicowa*. Warszawa: PWN.
- Okoń W. (1967) *Podstawy wykształcenia ogólnego*. Warszawa: Nasza Księgarnia.
- Niemierko B. (1990) *Pomiar sprawdzający w dydaktyce. Teoria i zastosowania*. Warszawa: PWN.
- Niemierko B. (1999) *Pomiar wyników kształcenia*. Warszawa: WSiP.
- Niemierko B. (2002) *Ocenianie szkolne bez tajemnic*. Warszawa: WSiP.

- Niemierko B. (2007) *Kształcenie szkolne. Podręcznik skutecznej dydaktyki*. Warszawa: WAiP.
- Niemierko B. (2009) *Diagnostyka edukacyjna. Podręcznik akademicki*. Warszawa: WN PWN.
- Niemierko B. (2010) Między prawdą a skutecznością – perspektywy oceniania szkolnego (w:) B. Niemierko, M.K. Szmigel (red.) *Teraźniejszość i przyszłość oceniania szkolnego*. Kraków: PTDE.
- Nitko A.J. (1994) *Curriculum-Based Criterion-Referenced Continuous Assessment: A Framework for the Concepts and Procedures of Using Continuous Assessment for Formative and Summative Evaluation of Student Learning*. ERIC, ED 377 199.
- Nitko A.J. (1998) Model egzaminów państwowych opartych na programie nauczania, sprawdzających i różnicujących, **przeznaczonych do dyplomowania i selekcji uczniów** (w:) B. Niemierko, E. Kowalik (red.) *Perspektywy diagnostyki edukacyjnej*. Gdańsk: Wyd. Uniwersytetu Gdańskiego.
- Perie M., S. Marion, B. Gong (2009) Moving toward a Comprehensive Assessment System: A Sramework for Considering Interim Assessment System. *Educational Measurement: Issues and Practice. Special Issue: The Validity of Formative and Interim Assessment*, No. 3.
- Pieter J. (1969) *Egzamin obiektywny*. Warszawa: Nasza Księgarnia.
- Scriven M. (1967) The Methodology of Evaluation. W: R.E. Stake (red.) *Perspectives of Curriculum Evaluation*. Monograph Series on Evaluation, no. 1. Chicago: Rand McNally.
- Shepard L.A. (2000) The Role of Assessment in a Learning Culture. *Educational Researcher*, 27, No. 7, pp. 4–14.
- Smith F.M., S. Adams (1966) *Educational Measurement for the Classroom Teacher*. New York: Harper.
- Soltis J.F., D.C. Phillips (2003) *Podstawy wiedzy o nauczaniu*. Gdańsk: GWP.
- Stróżyński K. (red.) (2011) *Ocenianie kształtujące po polsku. Kurs dla doradców metodycznych. Scenariusze zajęć*. Warszawa: CODN.
- Szaleniec H. (2004) *Jak komunikować uczniom wyniki egzaminów*. Warszawa: WSiP.
- Szyling G. (2011) *Nauczycielskie praktyki oceniania poza standardami*. Kraków: Impuls.
- Wiggins G.P. (1998) *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco: Jossey-Buss Publishers.
- Wilson M. (2016) From the President. *NCME Newsletter*, 24, No 4, p. 2.