

prof. zw. dr hab. Bolesław Niemierko

Akademia Marynarki Wojennej w Gdyni

Między prawdą a skutecznością – perspektywy oceniania szkolnego

Referat dotyczy prorozwojowej funkcji oceniania szkolnego. Nie spełnia jej należycie ani (a) ocenianie sumujące (*summative assessment*), zbyt wąskie i dokonywane zbyt późno na to, by wspomagać uczenie się, ani (b) ocenianie kształtujące (*formative assessment*), w przewadze uznaniowe, o niepewnym wpływie na umiejętności ucznia.

W literaturze światowej zasygnalizowano potrzebę budowania systemów (c) oceniania orientującego (*interim assessment*) ucznia i nauczyciela o tempie rozwoju określonego rodzaju osiągnięć. Autor referatu proponuje następujące założenia takiego systemu:

I. Wraz z postępem ekonomicznym i społecznym – w tym zwłaszcza komunikacyjnym – wzrastają (a) rozmaitość i (b) zróżnicowanie osiągnięć uczniów.

II. Kształcenie szkolne powinno stwarzać jednakowo korzystne warunki uczenia się dzieciom i młodzieży o różnym tempie rozwoju w poszczególnych dziedzinach.

III. Ocenianie etapowe powinno obejmować (a) osiągnięcia emocjonalno-motywacyjne i (b) osiągnięcia poznawcze, w proporcjach zapewniających najwyższą skuteczność kształcenia.

IV. Zarówno procedury (a) pomiaru sprawdzającego (*criterion-referenced measurement*), jak i (b) pomiaru różnicującego (*norm-referenced measurement*) są przydatne w wewnątrzszkolnym ocenianiu orientującym.

1. Niezborne kategorie?

Teoretycznie rzecz biorąc, prawda, czyli zgodność określonego zdania z rzeczywistością, i skuteczność, czyli zgodność wyniku działania z jego celem, dwa pojęcia umieszczone w tytule referatu nie stanowią ani opozycji, ani, tym bardziej, alternatywy. Przeciwnością prawdy jest „fałsz” albo „kłamstwo”, zaś przeciwieństwem skuteczności jest „niepowodzenie” lub „błąd”.

Prawda jest dla nas fundamentalna. Ma walor logiczny i etyczny bez względu na cele działania jednostki i grupy, zaś jej zatajanie, a tym bardziej – świadome posługiwanie się kłamstwem, jest powszechnie potępiane. Gdy słyszymy, że zasadą polityki jest skuteczność, odczuwamy dreszcz obaw o manipulację faktami, a występy polityków w środkach masowego przekazu niejednokrotnie te obawy potwierdzają. Bywa, że w dążeniu do szybkiego sukcesu mijają się z prawdą.

Prawda i skuteczność są dla nas kategoriami z różnych, niejednakowo cennych „bajek”. Zestawione, przyprawiają o ból głowy, wywołują podwójne widzenie świata, astygmatyzm intelektualny. Pedagogika jest jednak polem zastosowań obydwu podejść. Oto przykład zaczerpnięty z moich wczesnych doświadczeń prelegentkich (Niemierko, 1997, s. 122):

Przed wielu laty poproszono mnie o wykład dla nauczycielek (mówiło się wtedy – wychowawczyń) przedszkola na temat sprawdzania osiągnięć. Dwoiłem się i troiłem, by uzasadnić potrzebę formułowania i stosowania stałych wymagań programowych, ale w żaden sposób nie mogłem swoich słuchaczek przekonać. Dowodziły mi, że przecież zdolności dzieci są różne, że każde dziecko trzeba zachęcać itp.

Zdesperowany, zdobyłem się na następującą anegdotę:

Przypuśćmy, że kierujecie panie koloniami letnimi. Dzieci powinni jeść dużo, by nabrać sił i poprawić zdrowie. Tymczasem jedzą mało, bo są podniecone i rozbiegane. Mierzmy, jak przewidują przepisy, wzrost i ciężar ciała na początku kolonii, a następnie mówimy, że te dzieci, które będą dużo jadły, na pewno staną się większe i silniejsze.

Taak – potwierdziły panie z pełnym zrozumieniem, spoglądając na mnie po raz pierwszy z sympatią. Ja zaś, nabierając pewności siebie, kontynuowałem:

Jedne dzieci, ufając wychowawczyniom oraz z własnego usposobienia, jedzą pilnie i dużo, a inne lekceważą sobie obowiązek i bardziej dbają o zabawę. Na koniec kolonii powtarzamy pomiary. Okazuje się, niestety, że niekoniecznie te posłusznie jedzące urosły i przybrały na wadze. W niektórych przypadkach wyniki są odwrotne do starań, ale wszystkie dzieci są żywo zainteresowane wynikami pomiarów. Co im przekazemy: prawdę o ich rozwoju czy też inne dane – w centymetrach i kilogramach – dobrane według starań?

Oczywiście, że według starań! – odpowiedziały panie, podsumowując w ten sposób mój wykład.

W tej przygodzie, jak to bywa ze zdarzeniami krytycznymi w edukacji, czyli interakcjami pedagogicznymi o doniosłym znaczeniu dla rozwoju systemu kształcenia (Tripp, 1996), w tym przypadku – dla dojrzewiania poglądów jednego autora, jest to wszystko, co składa się na problem oceniania pedagogicznego:

1. Celem działań edukacyjnych jest rozwój wychowanka, rozumiany jako ciąg zmian długotrwałych, korzystnych, spowodowanych jego siłami wewnętrznymi, nie w pełni kontrolowanymi środkami pedagogicznymi (Brzezińska, 2000, rozdz. 2). W szkole i na koloniach letnich możemy – i powinniśmy – wspomagać rozwój dzieci i młodzieży. Nie potrafimy go jednak powodować ani zapewniać, a więc wszelkie obietnice co do tego są mocno ryzykowne.
2. Obok zadatków genetycznych, na rozwój wychowanka, także fizyczny, wpływa tryb życia, regulowany po części motywacją zewnętrzną, uruchamianą wzmocnieniami (Brophy, 2002, s. 25). Do zabawy wystarczy

- motywacja wewnętrzna, uruchamiana przyjemnością udziału w przedsięwzięciu, ale do uczenia się w szkole i do przerywania zabawy na koloniach zwykle nie wystarcza. Sięgnięcie po wzmocnienia pożądanych zachowań bywa więc konieczne.
3. Istotnym źródłem wzmocnień jest dla wychowanka ocena ze strony pedagoga, na którą składa się informacja o wyniku czynności i komentarz do tego wyniku (Niemierko, 2002, s. 187). Informacja powinna być prawdziwa, czyli zgodna z faktami, zaś komentarz reprezentuje system edukacyjny pedagoga. Rzeczywistość wychowawcza nie zawsze układa się po naszej myśli, ale jest wiele sposobów wyrażenia aprobaty lub dezaprobaty zachowań dzieci bez uciekania się do fałszowania lub ukrywania danych.
 4. Wynik rozwoju jednostki tylko w części podlega pomiarowi, czyli uściślonej procedurze skalowania, jak w przypadku wzrostu, ciężaru ciała lub umiejętności rozwiązywania pewnego rodzaju zadań matematycznych. W innej części ten wynik wymaga intuicyjnych oszacowań, na przykład kolonijnego wyczynu, stanu zdrowia dziecka lub postawy ucznia wobec matematyki. „Twarde” dane pełnią funkcję zasadniczą w ocenianiu rozwoju, ale sprowadzanie wyniku do tych danych ogranicza trafność decyzji podejmowanych na jego podstawie.
 5. W praktyce pedagogicznej ocenianie wąsko dydaktyczne, oparte wyłącznie na standardach jako normach rozwoju, ma ograniczone zastosowanie. Odwołanie się do tablic rozwoju fizycznego dzieci w niczym by wychowawczyom kolonijnym nie pomogło. Potrzebne jest ocenianie społeczno-wychowawcze, poszerzone o liczne kryteria kontekstowe (Niemierko, 2002, s. 212). Dzieci powinny zyskać przekonanie, że ich praca nad sobą była co najmniej równie cenna jak to, co pokazuje wzrostomierz i waga osobowa.

Skuteczność dydaktyczna pozostająca w pełnej prawdzie, bez nadużycia pozycji nauczyciela-wychowawcy, jest nie lada wyzwaniem pedagogicznym. Nic zatem dziwnego, że wielu dydaktyków albo chce wyeliminować pomiar dydaktyczny, jako nosiciela niepełnej prawdy o rozwoju ucznia, albo kwestionuje skuteczność jako właściwość kształcenia. Nowszą jest ta druga tendencja. „Paradygmat skuteczności” uważa się w niej za oręż pedagogiki instrumentalnej, spływającej kształcenie szkolne do realizacji „oczekiwań władzy, elit i dominujących grup społecznych” (Krause, 2009, s. 92).

Czy wyrzeczenie się skuteczności uzdrowiłoby systemy kształcenia? Odpowiedź na to pytanie wymaga konsultacji z ekonomistami i politykami, po uprzednim przekonaniu ich, że pomysł jest poważny. Ekonomisci są mocno zainteresowani zwiększaniem kapitału niematerialnego społeczeństw, jaki tworzą kompetencje nabywane w toku nauki szkolnej i w pracy zawodowej (Wałukiewicz, 2009), a politycy postulują systemy odpowiedzialności edukacyjnej (*educational accountability systems*), w których instytucje edukacyjne i ich kadry byłyby konsekwentnie rozliczane z wyników kształcenia (Koretz i Hamilton, 2006). Nic zatem nie zapowiada zdjęcia z barków pedagogów obowiązku osiągnięcia celów kształcenia.

2. Ocenianie kształtujące osiągnięcia uczniów

Impulsem do podjęcia problematyki oceniania na XVI Konferencji Diagnostyki Edukacyjnej był monograficzny numer czasopisma *Educational Measurement: Issues and Practice* (2009a) poświęcony trafności różnych rodzajów oceniania szkolnego, a w szczególności słowo wstępne do tego numeru napisane przez Susan Brookhart, redaktorkę naczelną czasopisma, autorkę szeroko znanego dzieła „Ocenianie szkolne” (2004).

Na początek cofniemy się do historii oceniania kształtującego.

Termin „ewaluacja kształtująca” (*formative evaluation*), wkrótce przekształcony na „ocenianie kształtujące” (*formative assessment*), wprowadził do teorii pedagogicznej Michael Scriven (1967) jako wartościowanie programu szkolnego w toku jego powstawania. Wartościowanie gotowego programu szkolnego zostało nazwane przez Scrivena „ewaluacją sumującą” (*summative evaluation*).

Pojęcie oceniania kształcenia rozwijało się i pogłębiało przez lata, by przybrać współcześnie, według Susan Brookhart (2009a, s. 1), następującą postać:

Ocenianie kształtujące to (I) informacja o procesie uczenia się, którą (II) nauczyciele mogą wykorzystać do decyzji dydaktycznych, a (III) uczniowie mogą wykorzystać do podniesienia osiągnięć, (IV) motywująca uczniów.

Definicja utrwała sekwencję wkładu kolejnych autorów w to pojęcie:

(I) Michael Scriven (1963) określił „kształtowanie” jako wykorzystywanie informacji do ulepszania działań, w odróżnieniu od podejmowania decyzji o ich końcowym wyniku.

(II) Benjamin Bloom (Bloom i in. 1971) przeniósł uwagę z budowania programów na proces kształcenia, akcentując to, co nauczyciel ma zrobić z uzyskaną informacją.

(III) David Sadler (1989) przeniósł akcent z tego, co nauczyciel powinien zrobić, na to, co uczeń (pierwotnie – student uczelni wyższej) ma zrobić z tą informacją.

(IV) Brytyjczycy, Paul Black i David Wiliam (1998), wnieśli do koncepcji oceniania kształtującego przekonanie, że „zaangażowanie ucznia wymagane do oceniania kształtującego obejmuje nie tylko zjawiska poznawcze, lecz także zjawiska motywacyjne” (Brookhart, 2009a, s. 2). Ich stanowisko poparło wielu innych autorów.

Brookhart przyznaje, że inaczej zbudowana, nie w układzie historycznie narastających warstw pojęcia, definicja amerykańskiej Rady Zarządzających Szkołami Publicznymi (CCSSO) jest zgrabniejsza, „bardziej elegancka”. Brzmi ona:

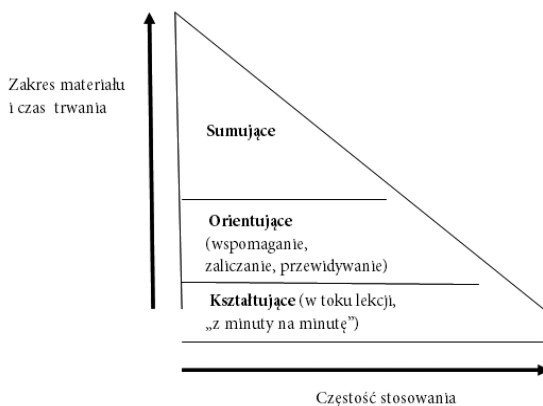
Ocenianie kształtujące jest procesem wykorzystywanym przez nauczycieli i uczniów w toku kształcenia, dostarczającym sprzężenia zwrotnego, by skierować bieżące nauczanie i uczenie się na podniesienie osiągnięć w zakresie zamierzonych wyników kształcenia.

Tak rozumiane ocenianie zrobiło wielką karierę w dydaktyce, na świecie i w Polsce (Stróżyński, w druku). Nauczyciele jednogłośnie opowiedzieli się za ocenianiem kształtującym (inne wersje terminu: formatywnym, konstruktywnym, wspierającym, bieżącym), pozostawiając, z żalem lub bez żalu, ocenianie sumujące (sumatywne, konkluzywne, całościowe, końcowe) wyspecjalizowanym agencjom egzaminacyjno-testowym. Można to zinterpretować jako dowód nastawienia nauczycieli na skuteczność działania, na wspomaganie procesu uczenia się przy niepełnej znajomości osiągnięć ucznia, a nie na pełną (całą, ostateczną) prawdę o jego osiągnięciach poznawczych, jakiej należałoby się spodziewać po egzaminie końcowym.

Inna, głębsza interpretacja ujawnionej preferencji nauczycieli dotyczy zakresu informacji o uczniu: w egzaminie końcowym, zwykle zewnętrznym i anonimowym, ważne jest tylko rozwiązanie zadań, a więc jest to ocenianie wąsko dydaktyczne, podczas gdy nauczyciel wie wiele o przeszłości, teraźniejszości i prawdopodobnej przyszłości ucznia, a więc – bardziej lub mniej jawnie – korzysta z tej wiedzy, nadając ocenianiu charakter społeczno-wychowawczy.

3. Trzy rodzaje oceniania szkolnego

Gdy dwa rodzaje oceniania szkolnego stały się wyraźnie odrębne (Niemierko, 2007, s. 302-308), między nimi wytworzyła się luka. Tę lukę ma obecnie wypełnić ocenianie orientujące, rozumiane jako wartościowanie wyników uczenia się ze względu na cele, do których osiągnięcia jednostka lub grupa zbliża się różnymi drogami i w różnym tempie. Jego usytuowanie przedstawia następujący diagram (Perie i in., 2009, s. 7; zamieszczony także na okładce numeru specjalnego *Educational Measurement: Issues and Practice*):



Rys. 1. Rodzaje oceniania szkolnego

Ocenianie sumujące jest stosowane najrzadziej, z najstaranniejszym przygotowaniem, zwykle jako dyplomowe, warunkujące ukończenie szkoły. Jego treść jest najszersza pod względem materiału kształcenia, ale najuboższa w elementy charakterystyki emocjonalno-motywacyjnej ucznia. Dostarcza wiele

informacji o osiągnięciach poznawczych absolwenta szkoły (klasy, kursu), ale niewiele informacji o jego potrzebach edukacyjnych, stylu uczenia się, możliwości wspomagania rozwoju, dojrzałości społecznej.

Ocenianie orientujące ma średni zakres materiału, jest dokonywane zwykle ogólnie dostępnymi narzędziami, nie wykluczając testów nauczycielskich. Etapowe, odmierzane modułami programowymi, działaniami, semestrami, realizowane siłami nauczyciela lub szkoły, pozwala na powiązanie wyników testowania z obserwacją pracy uczniów w wybranym okresie. Może być wykorzystane do podejmowania decyzji programowych i metodycznych na szczeblu klasy, szkoły, a nawet regionu, ale także do samodzielnego monitorowania rozwoju. Gdy pilny uczeń samorzutnie rozwiązuje zadania umieszczone na końcu kolejnego rozdziału podręcznika (*end-of-chapter tests*; *end-of-unit tests*) lub dostępne w Internecie, by sprawdzić, czy nadąży za treścią kształcenia i czy może już pójść naprzód, to jest właśnie, zdaniem autorów diagramu, ocenianie orientujące.

Najwięcej krytycznych uwag znajdujemy w referowanym numerze czasopisma pod adresem oceniania kształtującego. Ma ono charakter bieżący, nierozdzielny z procesem kształcenia, często improwizowany przez nauczyciela. Nawet „pięciosekundowe ocenianie” (Perie i in., 2009, s. 6) może być konsekwencyjnie trafne, gdy wspomogło uczenie się. Rzecz jednak w tym, że termin „kształtujące” częściej wyraża intencję nauczyciela niż fakt regulacji uczenia się przez ocenianie (Black i Wiliam, 1998). Brak regularnych badań nad skutecznością oceniania uważanego za kształtujące, a testowanie nauczycieli pokazuje, że planowanie ćwiczeń eliminujących błędy popełniane przez uczniów, na przykład z powodu niezrozumienia reguł matematycznych, jest najsłabszym elementem warsztatu metodycznego nauczycieli amerykańskich (Heritage i in., 2009). Tymczasem bardziej skuteczne, czyli „kształtujące”, niż chwalenie lub ganieńie, kierujące uwagę uczniów na nich samych i na ich zdolności, jest diagnozowanie trudności i naprowadzanie uczniów na właściwe reguły wykonywania określonych czynności (Shepard, 2009, s. 33). Z tego powodu niektórzy z autorów monograficznego numeru czasopisma skłonni są nawet porzucić termin „ocenianie kształtujące” na rzecz mniej fasadowej nazwy ocenianie na potrzeby uczenia się (*assessment for learning*; Brookhart, 2009; Nichols i in., 2009). „Ocenianie kształtujące to w równym stopniu tematyka uczenia się jak oceniania” – powtarza Susan Brookhart w swoim wprowadzeniu (2009b).

4. Dwie strategie oceniania szkolnego

„Nowo wynalezione ocenianie orientujące” (*newly-invented interim assessment*), jak z przekąsem określa je Lorrie Shepard (2009, s. 32), i „nadanie ocenianiu kształtującemu statusu celebryty” (tamże, s. 36) zmieniają sytuację pomiaru dydaktycznego w szkole amerykańskiej. Zgodnie z postulatami federalnego programu „Żadne dziecko nie zostaje w tyle” (No Child..., 2002), akcent przesuwają się z uzyskiwania informacji i jej interpretacji na wykorzystanie informacji do celu podnoszenia skuteczności kształcenia, w tym zwłaszcza do udzielania pomocy słabszym uczniom, „zostającym w tyle”. Wzrasta rola teorii uczenia się i dydaktyk przedmiotowych (Shepard, 2009).

„Wynalezienie”, czy raczej odkrycie przez specjalistów pomiaru dydaktycznego (a) istotnej roli wewnątrzszkolnych przeglądów wyników uczenia się w postaci okresowych sprawdzianów, klasówek, kartkówek i zastosowań ogólnodostępnych testów osiągnięć szkolnych i (b) dydaktycznej doniosłości nauczycielskiej diagnostyki nieformalnej, dokonywanej w działaniu, „z minuty na minutę” – stwarza szczególną okazję do przedyskutowania teoretycznych podstaw oceniania z pozycji wielu udziałowców kształcenia szkolnego.

Udziałowcy kształcenia (*educational stakeholders*) to osoby świadomie inwestujące swój czas i wysiłek w jego przebieg: uczniowie, nauczyciele, rodzice uczniów, administracja oświatowa, doradcy naukowci, działacze społeczni, ekonomiści i politycy różnych szczebli. Ich oczekiwania i doświadczenia w zakresie oceniania szkolnego zbiera tab. 1., sporządzona na podstawie różnych lektur i wieloletnich obserwacji autora tego referatu.

Tabela 1. Oczekiwania udziałowców kształcenia wobec oceniania szkolnego i ich typowe doświadczenia w tym zakresie

Udziałowcy	Oczekiwania	Doświadczenia
Uczniowie	Oceny powinny być „sprawiedliwe”, dokładnie zgodne ze stanem wiedzy ucznia, a zarazem mobilizujące do uczenia się.	Nauczyciele często wykorzystują ocenianie do zapewnienia dyscypliny w toku zajęć i kierują się w nim ogólną opinią o uczniu.
Nauczyciele	Oceny mają dokumentować poziom wiedzy ucznia i zachęcać go do systematycznej pracy na poziomie indywidualnych uzdolnień.	Większość uczniów orientuje uczenie się na uzyskanie wybranych ocen szkolnych, często najniższych ocen pozytywnych.
Rodzice	Szkoła odpowiada za poziom osiągnięć ucznia, a rodzice, orientując się według ocen, mogą pomóc w regulowaniu uczenia się.	Uczniowie skarżą się na niesprawiedliwe traktowanie przez nauczyciela, uprzedzenia wobec jednych i faworyzowanie innych.
Administracja	Uczniowie powinni uzyskiwać wysokie wyniki uczenia się, a w żadnym wypadku nie niższe niż wyznaczone standardami edukacyjnymi.	Wielu uczniów w normie psychologicznej uzyskuje wyniki niższe niż przewidziane właściwymi programami kształcenia.
Doradcy	Ocenianie powinno mieć wszystkie niezbędne cechy pomiaru dydaktycznego: bezstronność, dokładność, rzetelność, trafność i obiektywizm.	Większość decyzji o treści i metodach działania edukacyjnego jest podejmowana intuicyjnie, nie na podstawie pomiaru.
Działacze	System edukacyjny ma zapewnić wszystkim dzieciom pełny dostęp do oświaty i wyrównać ich warunki startu szkolnego.	Rozwój intelektualny dziecka podlega prawu efektu wachlarzowego, a więc różnice poziomu osiągnięć rosną z wiekiem.
Ekonomiści	Edukacją jest częścią systemu ekonomicznego kraju, wysoko rentowną w pespektywie średniookresowej i długookresowej.	Zysk z inwestycji edukacyjnych w postaci wzrostu osiągnięć uczniów jest trudno mierzalny i mocno odroczony.
Politycy	Właściwe wykształcenie młodzieży zapewni ciągłość kultury narodowej, rozwój ekonomiczny i postęp cywilizacyjny kraju.	Upowszechnianie wykształcenia i zapobieganie marginalizacji społecznej wymaga obniżenia poziomu wymagań.

Tabela 1. zawiera postulaty udziałowców kształcenia oraz powszechnie znane fakty temperujące te oczekiwania. Umożliwia interpretację tych postulatów i faktów w kategoriach strategii P, polegającej na dążeniu do „prawdy”, do w pełni zobiektywizowanego obrazu osiągnięć uczniów, oraz strategii S, polegającej na „skuteczności”, na motywowaniu do uczenia się przez akcentowanie wybranej informacji o tych osiągnięciach. Strategia P premiuje pewność, niezależność od zmiennych i zróżnicowanych potrzeb odbiorców informacji, a strategia S premiuje elastyczność, posługiwanie się informacją do określonego celu przez jej selekcję i porządkowanie w pewien sposób. A oto co wynika z takiej interpretacji:

1. Strategię P, dającą przewagę pomiarowi, wydają się preferować (1) ekonomiści (ekonometry), psychologowie (psychometry), socjologowie, doradcy naukowcy szkoły, zwłaszcza w krajach anglosaskich, oraz – co może być dla nas zaskoczeniem – (2) uczniowie i ich rodzice. Ci pierwsi kierują się potrzebą ścisłego poznawania faktów, by móc ustalać prawidłowości, budować ogólne teorie i posługiwać się tymi teoriami do przewidywania rzeczywistości. Dobrze zdają sobie sprawę z ograniczeń obiektywizmu i trafności pomiaru dydaktycznego, ale chcieliby zwiększyć jego rolę w edukacji. Ci drudzy, reprezentujący podmioty edukacji, domagają się „prawdy” w imię prawa do własnej interpretacji faktów i regulowania swoich czynności. Skłonni są podejrzewać szkołę o koniunkturalizm, ukrywanie swoich braków i obciążanie uczniów odpowiedzialnością za brak zaangażowania i błędy metodyczne nauczycieli.
2. Strategię S, zorientowaną na mobilizację do uczenia się, wydają się stosować nauczyciele (Szyling, 2004), administracja, działacze i politycy. Jej zasadą jest optymizm pedagogiczny, wyrażający się przekonaniem o wartości wysiłku edukacyjnego w każdych warunkach i na każdym poziomie. Obiektywne dane mają dla nich wartość o tyle, o ile mogą przyczynić się do indywidualnego i zbiorowego postępu. Nie każda prawdziwa informacja jest pożyteczna w edukacji. Na przykład pomiar inteligencji ogólnej, stosunkowo łatwy do dokonania standardowymi technikami testowymi (Strelau, 1997, rozdz. 8) byłby niepotrzebny, a nawet szkodliwy ze względu na regulację uczenia się, bo uczeń, zarówno o wysokim, jak i o niskim wyniku pomiaru, poczułby się genetycznie zdeterminowany do pewnego poziomu osiągnięć i zwolniony z systematycznego wysiłku. Podobną rolę mogą odgrywać wyniki testów umiejętności, na przykład testów rozumienia czytanego testu, dość wysoko skorelowane z inteligencją werbalną ucznia (Nęcka, 2005, s. 200). Najbardziej użyteczna jest informacja prowadząca do zaplanowania czynności owocujących podniesieniem osiągnięć ucznia.
3. Doświadczenia szkolne, życiowe i zawodowe udziałowców kształcenia oraz wyniki badań naukowych jednoznacznie wykazują przewagę strategii S w praktyce. Jednolite standardy edukacyjne okazują się za wysokie dla jednych i za niskie dla innych uczniów; a zróżnicowane (wielostopniowe) standardy mimo woli utrwalają nierówności psychologiczne i społeczne. Nauczyciel podejmuje więc zadanie takiej interpretacji standardów, by ich sens („prawda”) został utrzymany, a zarazem by stawały

się atrakcyjnym i realnym („skuteczność”) celem pracy ucznia. W ten sposób, jak akcentuje Brookhart (2009b), ocenianie kształtuje uczenie się, o co w gruncie rzeczy chodzi wszystkim udziałowcom kształcenia szkolnego. „Pracujemy z takimi uczniami, jacy są, a nie z takimi, jacy powinni być” – stwierdzają doświadczeni pedagodzy.

5. Wartość dodana osiągnięć ucznia jako przedmiot oceniania

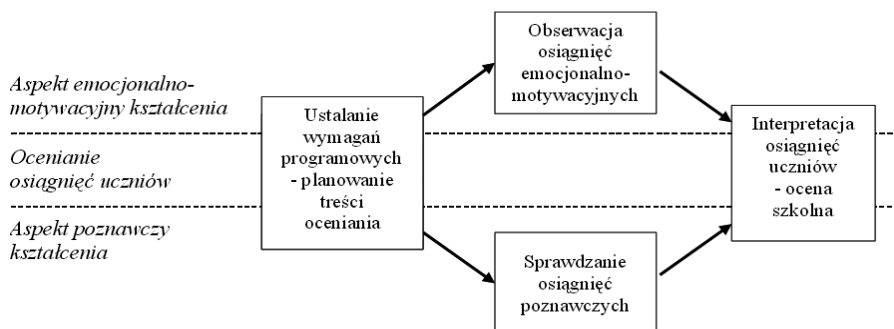
Dopiero dwadzieścia lat temu pedagodzy anglosascy zainteresowali się wartością dodaną osiągnięć uczniów (Gipps i Stobart, 1993; Saunders, 1999), a jej analizy w Polsce mają nie więcej niż pięć lat (Dolata, 2007; Niemierko, 2007, s. 362-365; Niemierko, 2009, s. 252-259). Edukacyjna wartość dodana, EWD (*educational value added*) jest przyrostem osiągnięć uczniów w wybranym zakresie programowym w określonym czasie. Do chwili obecnej zdecydowaną przewagę mają zastosowania tego pojęcia do oceny jakości pracy szkoły (Szmigel i Rappe, 2006; Szmigel, 2008; Pokropek, 2009) niż do oceniania osiągnięć pojedynczych uczniów, choć elementarnym procesem edukacyjnym jest uczenie się jednostek i ocena instytucji (szkoły) dokonywana na podstawie wyników jednostek, a nie odwrotnie. Przyczyny tej powściągliwości są zapewne trojaki:

1. Zjawiskiem społecznym, które spowodowało poszukiwanie nowego wskaźnika osiągnięć uczniów, było różnicowanie się poziomu kształcenia między szkołami na skutek konkurencji na rynku edukacyjnym podsycanej rankingami egzaminacyjnymi (Gipps i Stobart, 1993; Dolata, 2009). Szkoły plasujące się wyżej w „tabelach ligowych” pozyskiwały coraz lepszych kandydatów, a szkoły plasujące się niżej – coraz słabszych. Ta tendencja jest sprzeczna z zasadami demokracji i nieefektywna ekonomicznie, bo rozwarstwia społeczeństwo i ogranicza zasób dobrze wykształconej kadry. Należało zatem rozliczać szkoły według postępów, jakie zrobili ich uczniowie o nierównym przygotowaniu „na wejściu”, co potraktowano priorytetowo.
2. Pomiar zmiany, czyli różnicy między dwoma kolejnymi stanami układu, zakłada możliwość skalowania pionowego (*vertical scaling*) danej cechy, czyli rejestrowania różnic nie między jednostkami, lecz między poziomami właściwości jednostki uzyskanymi w procesie wzrostu. Możliwość takiego skalowania jest w edukacji ograniczona do podstawowych umiejętności uczniów, rozwijanych w sposób ciągły przez wiele lat, zwykle „od przedszkola do matury” ($K - I2$), jak w przypadku czytania, pisania, rozumowania matematycznego oraz interpretacji przyrody i społeczeństwa. To właśnie, z grubsza biorąc, mierzą nasze egzaminu zewnętrzne. Poza możliwością skalowania pionowego pozostają wiadomości, pamiętane w okresie opanowywania materiału i ewentualnych obowiązkowych powtórek, ale zwykle zapominane niedługo potem.
3. Trzecim z powodów nieufności pedagogów wobec EWD jako miary osiągnięć ucznia jest jej niepewność – duży błąd pomiaru różnicy między dwoma stanami układu. Gdy przy skalowaniu pionowym jako miarę osiągnięć ucznia zastosujemy równoważnik klasy, czyli

informację o klasie i miesiącu nauki, w którym średni wynik testowania ma daną wartość surową (punktową), to przy rzetelności pojedynczego wyniku testowania nie przewyższającej 0,8, standardowy błąd pomiaru wynosi w typowych warunkach około 6 miesięcy, a przy rzetelności wynoszącej 0,9 – około 4,5 miesiąca (Niemierko, 2009, s. 132). Wskutek tego w diagnozie edukacyjnej tylko różnice równoważnika w latach, a nie w miesiącach czy kwartałach, są istotne statystycznie. Stosujący ten wskaźnik częściej niż co pół roku napotka zjawisko nieregularnych, losowych wahań oceny umiejętności.

Mimo tylu niesprzyjających okoliczności edukacyjna wartość dodana osiągnięć ucznia jest informacją, która może wzmocnić strategię skuteczności kształcenia (S) w procedurze oceniania orientującego. Ocena postępu (*gain score*) ma cenne właściwości, jakich brak zarówno typowym metodom oceniania kształtującego, jak i procedurom oceniania sumującego. Są one następujące:

1. EWD jest nie tylko miarą osiągnięć ucznia w dziedzinie poznawczej, lecz także, pośrednio, wskaźnikiem jego osiągnięć emocjonalno-motywacyjnych, obejmujących uczucia, motyw, postawy i wartości. Te osiągnięcia, stanowiące równorzędny z poznawczymi wynik kształcenia (Niemierko, 2007, s. 34-36), nie podlegają zwykłym procedurom sprawdzania wyników uczenia się (odpytywaniu, klasówkom, testom, egzaminom), lecz wpływają na ocenianie społeczno-wychowawcze, dominujące w praktyce szkolnej (Brookhart, 1993). Rys. 1. przedstawia ten układ zależności (Niemierko, 2007, s. 270):



Rys. 1. Złożoność kształcenia i oceniania szkolnego w praktyce szkolnej

2. Inaczej niż w przypadku oceniania kształtującego, nastawionego na szybką korektę wykonywanych czynności, ocenianie orientujące, jak to akcentuje jego polska nazwa, działa w pewnym dystansie od celu kształcenia. Jego zadaniem jest dostarczać informacji o przebytej drodze, o położeniu celu i tempie zbliżania się do niego. Cel jest odległy nie tylko w wymiarze poznawczym, czyli pod względem umiejętności, lecz także w wymiarze emocjonalnym, od motywacji zewnętrznej („uczę się tego, bo muszę”) do wewnętrznej („chcę się tego uczyć”), od biernego uczestnictwa w lekcjach do systemu samokontroli postępu. Orientacja we własnych brakach jest warunkiem i dowodem samodzielności uczenia się. Oto przykład:

W początkowym okresie prac nad nową maturą znakomita polonistka (E. Kuczyńska), pracująca w znakomitym liceum, zaproponowała oryginalny schemat dziewięciokryterialnego holistycznego punktowania wypracowania szkolnego (Niemierko, 1999, s. 99-103), który kilkakrotnie zastosowała w nauczanych przez siebie oddziałach. Po gorących dyskusjach przyjęto jednak dla kraju model analityczny punktowania wypracowań, polegający na sporządzaniu szczegółowych wzorów wypracowań egzaminacyjnych, co do dziś budzi kontrowersje, bo twórcze prace uczniów, a tym bardziej – ekspertów, nie poddają się takiemu modelowaniu. Na pytanie, czy nie jest rozczarowana karierą swojego schematu, jego autorka odpowiedziała, że jest zadowolona, bo jej uczniowie samorzutnie przyjęli go jako metodę samokontroli jakości pisanych wypracowań, oceniając według niego swoje braki i możliwości ich przewyżczenia (zdobycia dodatkowych punktów i wyższych stopni). Schemat działa doskonale jako środek dydaktyczny organizujący rozwój umiejętności uczniów. Mniej zaawansowani i mniej ambitni uczniowie nie wykazaliby raczej tyle inicjatywy w zakresie samooceny, ale przykład wskazuje drogę, którą będzie zapewne podążać ocenianie szkolne w przyszłych dziesięcioleciach: rosnącą rolę ucznia, jego pracę nad błędami i potrzebę przygotowania nauczyciela do kierowania uczeniem się samokontroli (Niemierko, 2007, rozdz. 11).

3. Technologia informacyjno-komunikacyjna, jako zespół środków i metod przetwarzania, przechowywania i wykorzystywania informacji, ułatwia analizy edukacyjnej wartości dodanej, zwiększa ich znaczenie orientujące i umożliwia oddanie wielu procedur edukacyjnych w ręce uczniów. Dotyczy to w szczególności (Sysło, 2005):
 - a. pożądaných modyfikacji i „negocjowania” treści kształcenia według potrzeb, zainteresowań i osiągnięć ucznia,
 - b. samodzielnego poszukiwania potrzebnej informacji i przyspieszenia dostępu do niej,
 - c. doboru ćwiczeń polegających na rozwiązywaniu zadań i programowaniu zastosowań informacji w sytuacjach typowych i problemowych,
 - d. organizacji uczenia się, gospodarki czasem i wysiłkiem w pracy indywidualnej i zbiorowej,
 - e. wyboru czasu i miejsca uczenia się, sprawdzania i oceniania,
 - f. zastosowania technik gromadzenia i modeli interpretacji danych o wynikach uczenia się.

Wiele wskazuje na to, że technologia informacyjno-komunikacyjna będzie rozwijać się i upowszechniać systematycznie nadal, co doprowadzi po latach do utworzenia „społeczeństwa sieciowego”, w którym zmienna organizacja zadaniowa weźmie górę nad tradycyjnymi więzami instytucjonalno-lokalnymi (tamże, s. 45). W takim społeczeństwie uczenie się i samokontrola jednostek będzie odgrywać zasadniczą rolę.

Można spodziewać się, że ocenianie informujące, w tym analiza indywidualnej wartości dodanej podstawowych umiejętności, będzie „karmić się” wynikami standaryzowanych serii testów w miarę wzrostu jakości i dostępności takich serii. Pierwsze polskie próby zastosowań podobnego podejścia pojawiły się już w badaniach Okręgowej Komisji Egzaminacyjnej w Gdańsku (Kutajczyk i Przychodzeń, 2008) oraz w koncepcji „eduBarometru” Ośrodka Diagnozy Edukacyjnej w Krakowie.

6. Prawidłowości rozwoju osiągnięć ucznia

Wspieranie rozwoju ucznia wchodzi w konflikt z modelem szkoły jednolitej, opartej na założeniu, że wszyscy uczniowie w normie psychologicznej uzyskują podobne, dokładnie przewidziane programem kolejnych klas i szczebli systemu edukacji, osiągnięcia. Gdy rozwój ucznia staje się głównym celem jego edukacji (Kohlberg i Mayer, 1993), sztywna struktura dydaktyczna szkoły jednolitej okazuje się utrudnieniem kształcenia i źródłem dysfunkcji oceniania.

Tradycyjna dydaktyka była mocno związana z systemem klasowo-lekcyjnym, „polegającym na podziale uczniów na klasy, które realizują ustalone centralnie programy nauczania” (Okoń, 1998). W tym systemie narodził się także pomiar sprawdzający, w którym układem odniesienia dla każdego wyniku ucznia są wymagania programowe, a w spopularyzowanej w Polsce wersji wielostopniowej (Niemierko, 1990) wymagania są uporządkowane według skali stopni szkolnych. Taki rodzaj pomiaru silnie przemawiał do nauczycieli i administracji oświatowej, ale główna słabość systemu klasowo-lekcyjnego, jakim jest założenie planowego, wyrównanego rozwoju osiągnięć całej populacji uczniów, dawała znać o sobie coraz mocniej. Po trzydziestu latach od „wynalezienia” pomiaru sprawdzającego (Glaser, 1963) uznano, że miał on wprawdzie dobroczynne skutki dla świadomości dydaktycznej nauczycieli (Popham, 1994), ale obiecywał za wiele (Linn, 1994) i nie wszystkich obietnic udało mu się dotrzymać (Millman, 1994; Hambleton, 1994). W szczególności dotyczy to wymagań koniecznych, formułowanych jako minima programowe. Już znacznie wcześniej wybitny dydaktyk amerykański Henry Brickell (Airasian i in., 1979, s. 99) ostrzegął:

„W tradycji edukacyjnej minima są czymś, co szkoły starają się osiągnąć, a nie, co gwarantują. Są celami, a nie standardami. *Praca bez braków* nie jest dewizą szkoły. Jest nią *praca według możliwości ucznia*. A możliwości, jak nie mają górnej granicy, tak nie mają i dolnej”.

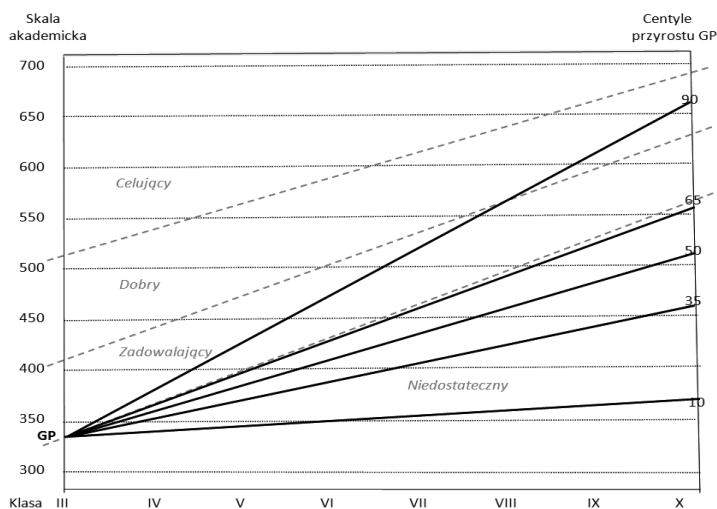
Mimo niepowodzeń programu *Minimum kompetencji* (Pipho, 1978; Niemierko, 2007, s. 160-161), w którym dla wielu szkół „minimum stało się maksimum”, oraz *Aktu edukacyjnego* prezydenta George’a Busha (seniora), przewidującego na rok 2000 światowe przodownictwo szkół amerykańskich w zakresie przedmiotów ścisłych, Stany Zjednoczone zdecydowały się na program „Żadne dziecko nie zostaje w tyle” (*No Child...*, 2002). Ten program jest nasycony pomiarem sprawdzającym (ustawowo co roku) w klasach III – VIII szkoły ogólnokształcącej. W miarę jak zbliża się jego końcowe rozliczenie, rok 2014, narasta niepokój o realność sytuacji, w której wszyscy uczniowie w normie psychologicznej osiągną, według standardów poszczególnych stanów, poziom biegłości (*proficiency*) w zakresie głównych umiejętności szkolnych. Z tego powodu federalny Departament Edukacji dopuścił zastosowanie modeli wzrostu (*growth models*), powiększające grono osiagających standardy o uczniów będących w trakcie (*on track*) osiagania standardów, z dostatecznie wysokim prawdopodobieństwem ich osiagnięcia przed ukończeniem ósmego roku nauki (*U.S. Department...*, 2005; *U.S. Department...*, 2009). To dało specjalistom pomiaru okazję pokazania, jak ograniczone są szanse na przyspieszenie rozwoju umiejętności słabszych

uczniów (*Educational Measurement...*, 2009b). Szczególnie wiele do myślenia daje wykres pokazujący sprzeczność sprawdzającej i różnicującej interpretacji wartości dodanej osiągnięć uczniów (Betebenner, 2009, s. 49; wykres jest umieszczony także na okładce monograficznego numeru czasopisma).

Betebenner przyznaje, że wzorem dla niego były table wzrostu i wagi dziecka stosowane w pediatrii, służące jako normy empiryczne rozwoju fizycznego. Takie table pozwalają ocenić nie tylko stan organizmu ze względu na te wskaźniki, lecz także jego rozwój (zmianę) w stosunku do innych dzieci o tych samych wartościach wskaźnika w poprzednim pomiarze.

Dane, jakimi posługuje się autor analiz, dotyczą matematyki. Rys. 2. zbudowano następująco:

1. Na lewym marginesie zaznaczono skalę akademicką (o średniej 500 i odchyleniu standardowym 100) osiągnięć matematycznych ogółu uczniów klas III – VIII.
2. Na prawym marginesie zaznaczono centyle, czyli punkty na skali odpowiadające określonemu procentowi badanych, rozkładowi wyników w kolejnych klasach uczniów o osiągnięciach „ledwie dopuszczających” (*barely proficient*) z matematyki w klasie III szkoły elementarnej, którą będziemy nazywali dalej grupą progową, GP.
3. W ramce wykresu skalę akademicką reprezentują linie kropkowane, normy osiągnięć „zadowalających” (*partly proficient*), „dobrych” (*proficient*) i „celujących” (*advanced*) wykresiono linią przerywaną, a „trajektorie” rozwoju uczniów z grupy GP w III klasie – linią ciągłą. Wszystkie linie, w oryginale lekko wypukłe ze względu na efekt pułapu, przybliżono tu jako proste.



Rys. 2. Prognoza rozwoju osiągnięć matematycznych uczniów klasy III o wyniku „ledwie dopuszczającym” w klasie III (GP) na tle norm wymagań w szkole amerykańskiej (według D. Betebennera)

Wykres rozwoju osiągnięć matematycznych uczniów z grupy GP, lokujących się na granicy oceny pozytywnej w klasie III szkoły podstawowej, ilustruje następujące prawidłowości:

1. Wiązka trajektorii rozwoju tej grupy wykazuje silny efekt wachlarzowy (*fan effect*) w postaci systematycznego wzrostu rozrzutu osiągnięć uczniów. Początkowo jednorodna, po siedmiu latach kształcenia GP staje się bardzo różnorodna, a jej środkowe 80% członków rozprasza się na niemal trzy odchylenia standardowe skali akademickiej, od 370 do 660 punktów.
2. Tylko około 10% członków grupy progowej spełnia wymagania biegłości (oceny „dobrej”) w klasie VIII, zakładanej dla wszystkich w programie „Żadne dziecko nie zostaje w tyle”. Nie wiemy, jacy to są uczniowie. Możemy sądzić, że pokonali trudności komunikacyjne, typowe dla wieloetnicznego społeczeństwa amerykańskiego, w którym język angielski jest drugim językiem w wielu rodzinach. Być może korzystali z korepetycji lub innych form zorganizowanej pomocy w nauce. Około 1/3 członków GP utrzymuje się na linii wymagań minimalnych (ocena „dopuszczająca”, *barely proficient*) lub wspina się powyżej tej linii. Można sądzić, że warunki rozwoju tych uczniów są korzystniejsze od reszty i ich ambicja podołania wymaganiom jest większa.
3. Typowy uczeń (mediana) z grupy progowej traci stopniowo pozycję wobec normy oceny „zadowolającej” (*partly proficient*) z roku na rok, powiększając populację niespełniających nawet częściowych wymagań programowych z matematyki.
4. Najwolniej rozwijający się uczniowie, stanowiący najśłabsze 10% GP, wykazują niewielki postęp, nie przekraczający 1/3 odchylenia standardowego, od 340 do 370 punktów skali akademickiej. Ich szanse na spełnienie najniższych wymagań programowych maleją z roku na rok, gdyż te wymagania rosną znacznie szybciej.

Wykres Betebennera łączy zastosowania pomiaru sprawdzającego (wielostopniowe normy wymagań programowych w układzie klas szkolnych) i pomiaru różnicującego (skala akademicka, skalowanie pionowe, centyle). Wyznacza nam pole funkcji oceny szkolnej wobec uczniów o różnych osiągnięciach, w tym zwłaszcza z grupy progowej, a tym bardziej „podprogowej”, rozwijających się wolniej od oczekiwań, o niskim prawdopodobieństwie spełnienia wymagań programowych. *Dzisiaj forsowane są nakazy osiągnięć jako moralny imperatyw wysokich standardów dla wszystkich dzieci, z niewielką troską o możliwość ich spełnienia przez uczniów – skarży się autor wykresu (2009, s. 50).*

7. Prorozwojowe funkcje oceny szkolnej

Dawid Betebrenner kończy swoje analizy konkluzją, iż *tylko przez równoczesne rozważanie tego, co jest, co powinno być i co jest możliwe, jest w stanie sprawić, że systemy odpowiedzialności edukacyjnej będą sprawiedliwe, uczciwe i prawdziwe (s. 50).*

Przenieśmy tę syntezę ze szczebla systemowego na szczebel oceniania osiągnięć uczniów w pojedynczym oddziale i przedmiocie. Prorozwojowe funkcje oceny szkolnej we w pełni upowszechnionej edukacji, przy dużym zróżnicowaniu stanu i przyrostu osiągnięć uczniów, są zasadnicze. Już dawno zauważono, że drugoroczność nie jest skuteczna wobec wolno rozwijających się dzieci, bo niewielkie (patrz dolna część rys. 2.) wyrównanie podstawowych umiejętności jest okupione dotkliwą degradacją społeczną dziecka. Mnożenie liczby uczniów odsianych, zniechęconych lub w inny sposób „wykluczonych” z dobrodziejstw wykształcenia (Kwieciński, 2002) jest niewątpliwą porażką systemu edukacji. Wydaje się zatem celowe:

1. Traktowanie rozwoju, a nie stanu osiągnięć uczniów jako pierwszoplanowego przedmiotu oceniania, ze świadomością, że o rozwój jest trudniej na niższych poziomach osiągnięć.
2. Przedstawianie uczniom prawdy o ich osiągnięciach w celu zwiększenia samodzielności w budowaniu strategii osiągania standardów edukacyjnych.
3. Wdrażanie uczniów do samokontroli skuteczności uczenia się przez regulowanie aspektów emocjonalnych i poznawczych tego procesu.
4. Akcentowanie pracy ucznia w procesie uczenia się, a nie jego uzdolnień ogólnych ani przedmiotowych.

Ostatni z tych postulatów eliminuje ocenę niedostateczną ze skali stopni szkolnych wobec wszystkich uczniów wykazujących motywację do uczenia się przedmiotu. *Nie mogłabym źle ocenić ucznia, który się stara, ponieważ uczeń, który pilnie pracuje, nie zasługuje na złą ocenę* – stwierdza typowa nauczycielka amerykańska (Brookhart, 1993, s. 140) i to samo zapewne powiedziała by nauczycielka polska. Wartość kształtująca negatywnej oceny osiągnięć ucznia jest w istocie znikoma, orientująca – wątpliwa, a sumująca – żadna. Kształtowanie wymaga instrukcji co do lepszego wykonywania czynności, orientowanie – pokazania, na czym można się oprzeć, a sumowanie – przedstawienia pozytywnego dorobku. Teorię dydaktyczną opartą na tych zasadach można by nazwać prorozwojową teorią oceniania szkolnego.

Wróćmy teraz do anegdoty kolonijnej przytoczonej na początku referatu. W świetle prorozwojowej teorii oceniania szkolnego zinterpretujemy ją następująco:

1. Dzieci mają prawo do prawdy o swoim rozwoju fizycznym. Nie wolno jednak przeceniać różnic wzrostu i wagi, a tym bardziej podnosić ich zmiany do rangi moralnego obowiązku. Są cenniejsze właściwości, o które trzeba się starać, także na koloniach.
2. Wdrożenie do samokontroli fizycznego stanu organizmu jest pożyteczne, zwłaszcza w obecnej epoce chipsów i batonów. Warto by udostępnić dzieciom wzrostomierz i wagę osobową.
3. Bezpośrednie wiązanie zmian wzrostu i wagi z obyczajami żywieniowymi jest nadinterpretacją prawidłowości biologicznej. Intencja zachęcenia dzieci do jedzenia nie usprawiedliwia podawania im fałszywej informacji.

4. Należy postarać się o skuteczne metody wdrażania dzieci do spokojnego spożywania posiłków. Codzienne ocenianie ich zachowań przy stole będzie kształtujące o tyle, o ile rzeczywiście oddziała na te zachowania.
5. Ocenianie orientujące kultury jedzenia, na przykład cotygodniowe i chyba lepiej grupowe, powinno opierać się na swobodnej obserwacji dokonywanej przez wychowawców. Wskaźniki metryczne rozwoju fizycznego i stanu zdrowia uzyskiwane w gabinecie lekarskim nadają się raczej do zastosowania w dłuższym okresie.
6. Sumujące ocenianie korzyści z kolonii musiałyby mieć zakres daleko szerszy niż pomiar wzrostu i ciężaru ciała. Przy tym, podobnie jak w ocenianiu osiągnięć szkolnych, negatywna ocena końcowa udziału dziecka w kolonii byłaby pozbawiona skuteczności wychowawczej.

8. Podsumowanie

Przegląd stanowisk prezentowanych w literaturze zagranicznej i analiza krajowych doświadczeń w zakresie oceniania szkolnego prowadzą do następujących uogólnionych wniosków:

- I. Mit *szkoły jednolitej* ciąży nad teorią i praktyką oceniania osiągnięć uczniów. By ten mit utrzymać, nauczyciele są zmuszani do niekonsekwencji w ocenianiu.
- II. *Podejście prorozwojowe* ogranicza rolę standardów edukacyjnych, które pozostają celami, ale przestają być normami uczenia się. Brak normy eliminuje negatywną ocenę szkolną.
- III. Ocenianie jest *kształtujące* wtedy, gdy oddziałuje na ucznia i podnosi wyniki jego uczenia się. Polega na znajdowaniu drogi do poprawnego wykonywania opanowywanych czynności.
- IV. Ocenianie jest *orientujące* wtedy, gdy uczeń i nauczyciel interpretują osiągnięcia ze względu na cele uczenia się. Pomiar dydaktyczny może dostarczać danych do takiego oceniania.
- V. Ocenianie *sumujące* dostarcza informacji o wynikach kształcenia wszystkim udziałowcom edukacji, od ucznia do polityka. Jako zamykające proces, ma najmniejszą wartość dydaktyczną.
- VI. Ponieważ trzy rodzaje oceniania szkolnego różnią się co do treści i funkcji w procesie kształcenia nie można oczekiwać od nich zgodności wyników. Taka zgodność byłaby jałowa.
- VII. Szkoła zaspokaja potrzeby edukacyjne uczniów o szybkim, przeciętnym i powolnym tempie rozwoju. W każdym przypadku *wartość dodana* liczy się nie mniej niż *stan* osiągnięć ucznia.
- VIII. Funkcja diagnostyczna oceniania jest służebna wobec *skuteczności* kształcenia. Z tego względu *ocenianie kształtujące* odgrywa rolę podstawową w prorozwojowym systemie oceniania.

Bibliografia:

1. Airasian P. W., G. F. Madaus, J. J. Pedulla (red.) (1979). *Minimal Competency Testing*. Englewood Cliffs 1979, Educational Technology Publications.
2. Betebenner, D. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice* (2009b) *Special Issue: Growth Model Pilot Program: Consequences and Alternatives*, nr 4.
3. Black, P., Wiliiam, D. (1998) Assessment and classroom learning. *Educational Assessment: Principles, Policy, and Practice*, 5.
4. Bloom, B.S., Hastings, J.T., Madaus, G.F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.
5. Brookhart, S.M. (1993). Teacher's grading practices: Meaning and values. *Journal of Educational Measurement*, 2.
6. Brookhart, S.M. (2004). *Grading*. Upper Saddle River: Pearson Education.
7. Brookhart, S.M. (2009a) Editorial. *Educational Measurement: Issues and Practice*, nr 1.
8. Brookhart, S.M. (2009b) Editorial. *Educational Measurement: Issues and Practice. Special Issue: The Validity of Formative and Interim Assessment*, nr 3.
9. Brophy, J. (2002). *Motywowanie uczniów do nauki*. Warszawa: WN PWN.
10. Brzezińska, A. (2000). *Społeczna psychologia rozwoju*. Warszawa: Scholar.
11. Dolata R. (2007). Krytyczna analiza metody edukacyjnej wartości dodanej. *Egzamin, Biuletyn badawczy* Nr 14, Centralna Komisja Egzaminacyjna w Warszawie.
12. Dolata, R. (2009). Cicha rewolucja w polskiej oświacie – proces różnicowania się gimnazjów w dużych miastach. [w:] B. Niemierko, M.K. Szmigel (red.) *Badania międzynarodowe i wzory zagraniczne w diagnostyce edukacyjnej*. Kraków: PTDE.
13. *Educational Measurement: Issues and Practice* (2009a). *Special Issue: The Validity of Formative and Interim Assessment*, nr 3.
14. *Educational Measurement: Issues and Practice* (2009b). *Special Issue: Growth Model Pilot Program: Consequences and Alternatives*, nr 4.
15. Gipps, C., Stobart, G. (1993). *Assessment. A teachers' guide to the issues*. London: Hodder & Stoughton.
16. Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 18.
17. Hambleton, R.K. (1994). The rise and fall of criterion-referenced measurement?
18. *Educational Measurement: Issues and Practice*, 4.
19. Heritage, M., Kim, J., Vendlinski, T., Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice. Special Issue: The Validity of Formative and Interim Assessment*, nr 3.
20. Kohlberg L., Mayer R. (1993). Rozwój jako cel wychowania. W: Z. Kwieciński, L. Witkowski (red.) *Spory o edukację. Dylematy i kontrowersje we współczesnych pedagogiach*. Toruń: Edytor.
21. Koretz, D.M., Hamilton, L.S. (2006). Testing for accountability in K – 12. W: R.L. Brennan (red.) *Educational measurement. Fourth edition*. Westport: American Council on Education – Praeger.
22. Krause, A. (2009) Iluzja dydaktycznej sprawczości – profesjonalizacja niekompetencji. W: L. Hurlo,
23. D. Klus-Stanska i M. Łojko *Paradygmaty współczesnej dydaktyki*. Kraków: Impuls.
24. Kutajczyk, T., Przychodzeń, B. (2008). *Sposoby szacowania przyrostu osiągnięć uczniów w gimnazjum*. Gdańsk: OKE.
25. Kwieciński, Z. (2002). *Wykluczanie*. Toruń: Wydawnictwo UMK.

26. Linn, R.L. (1994). Criterion-referenced measurement: A valuable perspective clouded by surplus meaning. *Educational Measurement: Issues and Practice*, 4.
27. Millman, J. (1994). Criterion-referenced testing 30 years later: promise broken, promise kept. *Educational Measurement: Issues and Practice*, 4.
28. Nęcka, E. (2005). *Inteligencja. Geneza – Struktura – Funkcje*. Gdańsk: GWP.
29. Nichols, P.D., Meyers, J.L., Burling, K.S. (2009). A response to Shepard. *Educational Measurement: Issues and Practice. Special Issue: The Validity of Formative and Interim Assessment*, nr 3.
30. Niemierko B. (1990). *Pomiar sprawdzający w dydaktyce. Teoria i zastosowania*. Warszawa: PWN.
31. Niemierko, B. (1999). *Pomiar wyników kształcenia*. Warszawa: WSiP.
32. Niemierko, B. (1997). *Między oceną szkolną a dydaktyką. Bliżej dydaktyki*. Warszawa: WSiP.
33. Niemierko, B. (2002). *Ocenianie szkolne bez tajemnic*. Warszawa: WSiP.
34. Niemierko, B. (2007). *Kształcenie szkolne. Podręcznik skutecznej dydaktyki*. Warszawa: WAIp.
35. No Child Left Behind Act of 2001 (2002). *Public Law No. 107 – 110, 115 Stat. 125*.
36. Okoń W. (1998) *Nowy słownik pedagogiczny*. Warszawa: Żak.
37. Perie, M., Marion, S. i Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice. Special Issue: The Validity of Formative and Interim Assessment*, nr 3.
38. Phipo C. (red.) (1978). *Minimum Competency Testing*, „Phi Delta Kappan. Special Issue”, s. 585-629.
39. Pokropek, A. (2009). Metody statystyczne wykorzystywane w szacowaniu trzyletnich wskaźników egzaminacyjnych. W: B. Niemierko, M.K. Szmigel (red.) *Badania międzynarodowe i wzory zagraniczne w diagnostyce edukacyjnej*. Kraków: PTDE.
40. Sadler, D.R. (1989) Formative evaluation and the design of instructional systems. *Instructional Science*, s. 119-144.
41. Saunders L. (1999). A brief history of educational „value added”: How did we get to where we are? *School Effectiveness and School Improvement*, 2.
42. 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.
43. Shepard, L.A. (2009). Commentary: Evaluating the validity of formative and interim assessment. *Educational Measurement: Issues and Practice. Special Issue: The Validity of Formative and Interim Assessment*, nr 3.
44. Strelau, J. (1998). *Inteligencja człowieka*. Warszawa: Żak.
45. Stróżyński, K. (red.) (w druku). *Ocenianie kształtujące po polsku. Kurs dla doradców metodycznych. Scenariusze zajęć*. Warszawa: CODN.
46. Sysło, M. (2005). Rozwój technologii informacyjnej a edukacja – stan, kierunki, wyzwania. W: B. Niemierko,
47. G. Szyling (red.) *Holistyczne i analityczne metody diagnostyki edukacyjnej. Perspektywy informatyczne egzaminów szkolnych*. Gdańsk: Fundacja Rozwoju Uniwersytetu Gdańskiego.
48. Szmigel, M.K. (2008). Wykorzystanie edukacyjnej wartości dodanej w lokalnej polityce oświatowej. W: B. Niemierko i M.K. Szmigel (red.) *Uczenie się i egzamin w oczach nauczycieli*. Kraków: PTDE.
49. Szmigel, M.K., Rappe, A. (2006). Stałość wskaźnika edukacyjnej wartości dodanej i opinie dyrektorów o jego znaczeniu. W: B. Niemierko, M.K. Szmigel (red.) *O wyższą jakość egzaminów szkolnych*. Lublin: PTDE.

50. Szyling, G. (2004) Problemy nauczycielskich strategii. „Szara strefa” oceniania w oczach uczniów. „Nowa Szkoła”, nr 9.
51. Tripp, D. (1996). *Zdarzenia krytyczne w nauczaniu. Kształtowanie profesjonalnego osądu*. Warszawa: WSiP.
52. U.S. Department of Education (2005). Secretary Spelling announces growth model pilot, addresses chief state school officers’ annual policy forum in Richmond. *U.S. Department of Education press release*. 2005/11/11182005.html.
53. U.S. Department of Education (2005). Secretary Spelling approves additional Growth Model Pilots For 2008-2009 school year. *U.S. Department of Education press release*. 2009/01/01082009a. html.
54. Walukiewicz S. (2009). *Kapitał ludzki i społeczny jako przedmiot badań pedagogicznych*. W: B. Niemierko i M.K. Szmigel (red.) *Badania międzynarodowe i wzory zagraniczne w diagnostyce edukacyjnej*. Kraków: PTDE.