

Znaczenie edukacyjne egzaminu doniosłego

Prof. dr hab. Bolesław Niemierko

W artykule przedstawię teoretyczną koncepcję egzaminu doniosłego, zagraniczne i krajowe doświadczenia w jej wdrażaniu oraz – ze szczególną starannością – ostrzeżenia przed nadinterpretacją wyników egzaminu.

Nadinterpretacja, dokonywana na zasadzie *pars pro toto* (część za całość), pojawiająca się uporczywie we wszystkich krajach wprowadzających systemy egzaminów zewnętrznych, jest sprowadzaniem oceny osiągnięć ucznia, nauczyciela i szkoły do wyniku jednego egzaminu, którego trafność jest zawsze ograniczona *kontekstem* (okolicznościami) i *procedurą* (jakością pomiaru). Przypomina wiarę laików w cudownie działające lekarstwo, wynikającą z braku gruntownej wiedzy medycznej.

Artykuł ma poszerzyć wiedzę czytelników o egzaminach doniosłych, a tym samym zapobiec zarówno ich przecenianiu, jak i zbyt ostrej krytyce, wynikającej z rozczarowań.

Dwa rodzaje egzaminów szkolnych

Egzaminem szkolnym nazywamy sprawdzanie i ocenianie osiągnięć wyodrębnione w procesie kształcenia. Egzaminy mogą otwierać, regulować lub zamykać kształcenie, a więc spełniać funkcję *diagnostyczną, kształtującą* lub *sumującą*. Obejmują określone osiągnięcia *poznawcze* (dydaktyczne) uczniów, nigdy zaś wprost nie obejmują ich osiągnięć *emocjonalno-motywacyjnych* (wychowawczych, charakterologicznych),

będących równie ważnym wynikiem kształcenia, możliwych do oszacowania przez obserwację procesu uczenia się.

Egzaminy mają różną wagę i różny rozkład akcentów między *informacją* o wyniku uczenia się i *komentarzem*:

1. *Egzamin powszedni* („niskich stawek”) to egzamin, w którym znaczenie komentarza dydaktycznego jest większe niż znaczenie informacji o wyniku uczenia się, a więc kształtowanie osiągnięć przeważa nad ich sumowaniem. Każda sytuacja dydaktyczna zapowiedziana, a nawet tylko traktowana przez nauczyciela jako sprawdzanie i ocenianie osiągnięć uczniów, jest tego rodzaju egzaminem.

2. *Egzamin doniosły* („wysokich stawek”) to egzamin, w którym znaczenie informacji o wyniku uczenia się jest większe niż znaczenie komentarza dydaktycznego, a więc sumowanie osiągnięć przeważa nad ich kształtowaniem. Ze względu na konsekwencje („stawkę”) uzyskanego wyniku (promocja, wstęp do szkoły wyższego szczebla) ten wynik jest ważniejszy dla ucznia niż wszelkie, spóźnione już, zachęty i dobre rady.

Nie wystarczą egzaminy wewnętrzne!

Jeżeli pominąć sporadyczne akcje szerszego badania osiągnięć uczniów, egzamin powszedni jest w Polsce *wewnętrzny*, organizowany przez nauczyciela. Sprawdzanie prac domowych, odpytywanie,

kartkówki i klasówki powszedniej uczniom. Na takim egzaminie nie można wiele stracić, a w każdym razie można dość łatwo stratę wyrównać. Można za to na nim sporo zyskać: zdobyć cenne doświadczenie i, gdy komentarz jest dydaktycznie dojrzały, wskazówki dotyczące skutecznego uczenia się.

Zwolennicy ograniczenia ewaluacji osiągnięć uczniów do egzaminów wewnętrznych padają ofiarą złudzeń co do warunków, w jakich one przebiegają. Te złudzenia są następujące (Niemierko, 2002, s. 254):

1. Egzamin powszedni przebiega *w dogodnych warunkach* dla ucznia i nauczyciela. Te warunki to spokój, swobodność, życzliwość i wzajemne zaufanie jako przeciwstawienie surowej obcości egzaminu doniosłego.

Doświadczeni pedagodzy nie wątpią w możliwość wytworzenia takich warunków, wiedzą jednak, że często zaufanie jest ograniczone („Nauczyciel uwziął się na mnie!”), atmosfera jest napięta, krzewi się oszustwo (podpowiadanie, ściąganie).

2. Dzięki dobrej znajomości uczniów egzamin powszedni jest *zindywidualizowany*. Nauczyciel może dobierać treść i formę egzaminu do osobowości ucznia, czego nie dopuszcza zobiektywizowana procedura egzaminu doniosłego.

Indywidualizacja egzaminu jest cenna, ale ryzykowna. Łatwo o wypaczenia, postrzegane przez uczniów jako rażąca niesprawiedliwość: faworyzowanie

wybrańców, nagradzanie za uległość i współpracę.

3. Nauczyciel *jasno mówi* swoim uczniom, czego będzie od nich wymagał. Treść i forma sprawdzania osiągnięć w ramach egzaminu powszedniego nie budzą wątpliwości i obaw, jakie wiążą się z udziałem komisji zewnętrznych.

Ten warunek najsilniej bodaj różnicuje nauczycieli. Obok takich, których stopnie, jakie postawią za określoną wypowiedź, uczniowie bezbłędnie przewidują, są też i tacy, których pytania i oceny są dla uczniów zaskoczeniem, źródłem stresu. Planowanie wyników wydaje się polepszać sprawę, ale do uczniów przenika ono raczej powoli.

4. Już przed egzaminem nauczyciel *zna prawdę* o osiągnięciach uczniów, a egzamin tę prawdę jedynie potwierdza. Holistyczne widzenie ucznia, oparte na intuicyjnym obrazie jego pracy, jest bardziej trafne niż statystyka wyników zadań egzaminacyjnych.

Tak bywa, ale nieczęsto. Nadzieja na to, że niezależnie pracujący nauczyciele będą jednym głosem mówić o wynikach swoich uczniów, nie spełnia się nigdzie na świecie.

Ograniczenia egzaminu zewnętrznego

W ramach reformy edukacji już w 1999 roku pojawiły się w Polsce pierwsze egzaminy *zewnętrzne*, organizowane przez wyspecjalizowane komisje egzaminacyjne, centralną i okręgowe. Wraz z nimi pojawiły się nadzieje na ostateczne rozwiązanie kwestii obiektywizmu oceniania osiągnięć uczniów, nauczycieli oraz szkół ogólnokształcących i zawodowych. Wytworzyły się nowe złudzenia, które można streścić następująco (Niemierko, 2002, s. 267-269):

1. Standardy edukacyjne *dokładnie wyznaczają* poziom osiągnięć uczniów.

Uważna lektura informatorów egzaminacyjnych, a tym bardziej – każda próba ich zastosowania do oceniania osiągnięć uczniów przekonuje, że standardy są chwiejne. Nawet najściślejsze opisy czynności uczniów nie są w stanie ustanowić normy pozwalającej na bezbłędne rozstrzygnięcie, czy uczeń opanował te czynności. Trudność testów waha się w kolejnych latach (przykład biologii na maturze) i między przedmiotami (historia a geografia), zaś punktowanie zadań znacząco waha się między egzaminatorami, wytwarzając niepożądany *efekt egzaminatora* (Dolata i in., 2004; Dubiecka i in., 2006).

2. Krajowi specjaliści wytwarzają zadania *najlepiej mierzące* osiągnięcia uczniów.

Taka opinia podnosi autorytet centralnych władz oświatowych i zatrudnionych przez nie ekspertów, ale nie da się utrzymać w świetle teorii pomiaru dydaktycznego. Nie potrafimy zbudować w pełni trafnych zadań, nie znając autorskich planów wynikowych i, co nie mniej ważne, sposobu ich realizacji w danej szkole (Niemierko, 2002, s. 177-178).

3. Ujednolicenie narzędzi zapewnia nam *pełną porównywalność* wyników egzaminu.

Niespecjaliści pomiaru (urzędnicy, dziennikarze, rodzice) łatwo zapominają, że egzamin jest tylko niewielką próbą czynności, reprezentującą wielki obszar opanowywanej treści kształcenia. Każde wnioskowanie z próby o populacji jest obciążone określonym błędem, a ujednolicenie zadań nie zwiększa precyzji wnioskowania statystycznego.

4. Egzaminy doniosłe pozwalają na obiektywną ocenę *pracy nauczycieli i szkół*.

To złudzenie jest najtrudniejsze do wykorzenia ze świadomości administracji szkolnej, rodziców i opinii publicznej. Od lat wiadomo, że rankingi i „tabele ligowe” wyrządzają wiele zł szkółom pracującym w trudniejszych warunkach (Porworow-

ski, 2000). W oczach laików każdy egzamin zewnętrzny, na wzór parlamentarnej komisji śledczej, odkrywa „całą prawdę” o szkole i demaskuje jej braki, bez względu na różnice wstępnego przygotowania uczniów i kulturę ich środowiska.

Sprawdzanie i różnicowanie

Sprawdzanie osiągnięć uczniów – zarówno wewnątrz szkoły, jak i w szerszym układzie – pozwala jedynie na *ocenianie wąskodydaktyczne* osiągnięć ucznia, ograniczone do wiadomości i umiejętności, a sfera społeczno-wychowawcza pozostaje na uboczu. Z tego względu wyniki egzaminu zewnętrznego lub badania osiągnięć szkolnych będą na ogół niższe niż oceny uzyskiwane przez uczniów wewnątrz szkoły (Niemierko, 1990, s. 45-47). Kto tego nie wie, doznaje przykrego rozczarowania.

Głębszą przyczyną nieporozumień jest *znaczenie treściowe* wyników egzaminu zewnętrznego, nieuchronnie ograniczone do obowiązkowej części wspólnej programów kształcenia, a więc do *podstawy programowej*, z pominięciem szczegółowych tematów i lektur. Z tego względu w testach (sprawdzianach) szerokiego użytku akcentowane są podstawowe umiejętności (kompetencje), a ich wyniki są wysoko skorelowane z wynikami testów uzdolnień (Anastasi, 1988, s. 411-415). Części autorskie programów kształcenia, które są zwykle najcenniejsze dla szkoły, nie mogą stać się treścią egzaminu zewnętrznego.

Przykładu doboru treści egzaminu zewnętrznego dostarczy nam sprawdzian osiągnięć uczniów szkoły podstawowej zastosowany w całym kraju w 2004 roku. Na tym przykładzie prześledzimy też wpływ, jaki na znaczenie treściowe egzaminu wywiera przedmiot i forma zadań mierzących poszczególne umiejętności. Pokazuje to tab. 1. (Niemierko, 2004):

Tabela 1. Znaczenie treściowe sprawdzianu osiągnięć uczniów po szkole podstawowej w 2004 roku

Umiejętność	Plan			Sprawdzanie			Różnicowanie		
	Rodzaj zadań	Liczba punktów	Udział	Łatwość zadania*)	Wynik punktowy	Udział	Moc różnicująca zadania*)	Wskaźnik rzetelności testki**)	Udział
Czytanie	zamknięte	10	25%	,82	8,19	32%	,38	1,39	18%
Pisanie	otwarte	12	30%	,69	8,25	32%	,45	2,40	31%
Rozumowanie	zamknięte otwarte	8	20%	,48 ,44	3,68	14%	,44 ,54	1,81	23%
Korzystanie z informacji	zamknięte otwarte	2	5%	,74 ,75	1,49	6%	,48 ,48	0,42	5%
Wykorzystywanie wiedzy w praktyce	zamknięte otwarte	8	20%	,62 ,29	3,95	16%	,40 ,61	1,78	23%
R a z e m	zamknięte otwarte łącznie	20 20 40	50% 50% 100%	,70 ,58 ,64	14,0 11,6 25,6	55% 45% 100%	,40 ,49 ,45	3,41 4,42 7,83	44% 56% 100%

*) Wartości średnie. W zadaniach otwartych każdą wyodrębnioną, osobno punktowaną „czynność” (aspekt rozwiązania) potraktowano jako osobne zadanie.

**) Wskaźnik rzetelności testki jest iloczynem jej odchylenia standardowego i mocy różnicującej (Niemierko, 1975, s. 261-262).

Tab. 1 przedstawia trzy rodzaje znaczenia treściowego egzaminu:

1. *Założone znaczenie treściowe*, oparte na planowaniu treści egzaminu i na konstrukcji narzędzia pomiaru. W przypadku analizowanego sprawdzianu założono, podobnie jak w kilku poprzednich latach, że pięć podstawowych umiejętności, mierzonych przez pięć wyodrębnionych *testek*, będzie reprezentowanych w proporcjach podanych w kolumnie „udział”, a zadania zamknięte i zadania otwarte złożą się na wynik egzaminu w równych proporcjach, dając maksymalnie po 20 punktów, a więc łącznie 40 punktów.

2. *Uzyskane znaczenie sprawdzające*, oparte na analizie wskaźników łatwości zadań rozwiązywanych przez uczniów. Porównując kolumnę „sprawdzanie – udział” z kolumną „plan – udział”, zauważamy zmianę proporcji znaczenia na korzyść „czytania” i na niekorzyść elementów przedmiotów ścisłych: „rozumowania” i „wykorzystania wiedzy w praktyce”. Istotne znaczenie ma tu także rodzaj zadań: zadania wyboru wielokrot-

nego, wprowadzone jako wyłączne w „czytaniu”, dostarczyły ogółem o 10 punktów procentowych większą część wyniku ogólnego niż – przy tej samej maksymalnej liczbie punktów – zadania otwarte. Silną pozycję sprawdzającą ma jednak także „pisanie”, którego jakość jest szacowana wyłącznie na podstawie zadań otwartych.

Taki wynik może cieszyć polonistów, których przedmiot jest w sprawdzaniu silniejszy niż w założeniach. Zadania matematyczne, jako na ogół trudniejsze, dostarczyły uczniom mniej punktów i mniej wpłynęły na ich dorobek punktowy niż zadania językowe. Ta prawidłowość jest tym silniejsza, im bardziej oddalamy się od wyniku maksymalnego ku wynikom niskim. To naturalne, że dla słabych uczniów każdy test pisemny jest przede wszystkim testem czytania, ale w grę może tu wchodzić także podatność zadań wyboru wielokrotnego na oszukiwanie (ściąganie).

3. *Uzyskane znaczenie różnicujące*, określone wkładem zadań do rozrzutu wyników pomiaru. Zadania otwarte lepiej różnicują egzaminowanych uczniów niż zadania zamknięte, co

nie jest oczywiście niespodzianką, bo samodzielność i czas pracy ucznia są na ogół większe w zadaniach otwartych, pozbawionych gotowych odpowiedzi. Procentowy udział testek mierzących pięć umiejętności objętych egzaminem pokazuje wzrost znaczenia „rozumowania” i „wykorzystywania wiedzy w praktyce” w stosunku do ich udziału w sprawdzaniu, a więc elementów przedmiotów ścisłych, średnio o ponad połowę, oraz dramatyczny spadek, niemal do połowy, znaczenia „czytania”.

Teraz górą wydają się być matematycy i przyrodnicy, których zadania lepiej mierzą umiejętności uczniów w sensie różnicowania osiągnięć (Stożek, 2006). Tylko „pisanie” utrzymuje wysoką pozycję przy obydwu podejściach, sprawdzającym i różnicującym. Ogółem, zadania otwarte wniosły wyraźnie więcej do różnicowania niż zadania zamknięte. Zwolennicy zadań otwartych zyskują więc potwierdzenie swoich, dobrze skądinąd uzasadnionych teoretycznie, przekonań.

Okrojenie treściowe oraz silna zależność wyniku od formy zadań powodują, że *trafność wewnętrzna* egzaminu ze-

wewnętrznego i badań osiągnięć szkolnych jest niższa niż trafność wewnątrzszkolnego sprawdzania osiągnięć (Knight, 2000; Poręba-Konopczyńska, 2001; Niemierko, 2002, s. 177-178). Zyskują one na niższych właściwościach sprawdzania – bezstronności, dokładnym punktowaniu i rzetelności – tracą natomiast na trafności i *obiektywizmie sprawdzania*, rozumianym jako reprezentowanie tych wymagań programowych, przed jakimi uczeń był postawiony w swojej szkole.

Stosunek nauczycieli do egzaminów zewnętrznych

To zagadnienie prześledzimy na przykładzie krajów anglosaskich. Niezadowalająca trafność programowa testów szerokiego użytku jest powodem powściągliwości nauczycieli w korzystaniu z ich wyników. Jeszcze w latach sześćdziesiątych ubiegłego stulecia w Stanach Zjednoczonych tylko co czwarty nauczyciel uważał, że przygotowywanie uczniów do takich testów należy do jego obowiązków i że wyniki tych testów świadczą o jego pracy (Goslin, 1967, s. 71 i 110), a trzydzieści lat później tylko co trzeci nauczyciel był zdania, że „informacja uzyskana z testów standaryzowanych powinna być wykorzystana do podniesienia jakości nauczania” (Plake i in., 1993, s. 12).

Badania przeprowadzone w Irlandii, gdzie w początku lat osiemdziesiątych eksperymentalnie wprowadzono testy standaryzowane, wykazały, że informacja o wynikach pomiaru jest wykorzystywana przez nauczycieli głównie wtedy, gdy zgadza się z ich własną opinią o uczniach, a zaufanie do testów rośnie w miarę zdobywania pewności, że wyniki testowania nie zburzą tej opinii (Kelleghan i in., 1982; Niemierko, 1990, s. 387). Tak więc w przekonaniu nauczycieli ocenianie wewnątrzszkolne ma zdecydowane pierwszeństwo.

Specjaliści pomiaru dydaktycznego dostrzegają i dokumentują wiele zagrożeń i równie wiele korzyści z testowania i egzaminowania zewnętrznego. Najważniejsze

zagrożenia są następujące (Smith i Rottenberg, 1991; wykaz uzupełniony):

1. redukcja czasu przydzielonego na „zwykłe” nauczanie,
2. zaniedbywanie materiału pozaegzaminacyjnego przez nauczycieli,
3. dostosowywanie metod kształcenia do zadań egzaminacyjnych,
4. ograniczanie swobody wyboru przedmiotów i uczenia się,
5. ujemny wpływ etyczny na nauczycieli,
6. postrzeganie egzaminu przez uczniów, zwłaszcza młodszych, jako surowej kary,
7. budzenie w uczniach niechęci do przedmiotów szkolnych,
8. obniżanie samooceny uczniów,
9. możliwa dyskryminacja środowiskowa i etniczna uczniów,
10. przecenianie wartości zadań wyboru wielokrotnego.

Równie długa jest lista *korzyści*, jakie przynoszą testowanie i egzaminy zewnętrzne (Cizek, 2001):

1. więcej podstaw do istotnych decyzji pedagogicznych,
2. pobudzanie rozwoju zawodowego nauczycieli,
3. wzrost zainteresowania uczniami o specjalnych potrzebach edukacyjnych,
4. wzrost wiedzy nauczycieli o pomiarze dydaktycznym,
5. nastawienie na zbieranie i wykorzystywanie informacji z różnych źródeł,
6. poszukiwanie alternatywnych sposobów kształcenia,
7. wspomaganie planowania finansowego edukacji,
8. pogłębianie znajomości macierzystych dyscyplin naukowych przez nauczycieli,
9. wzrost jakości testów egzaminacyjnych,
10. niewielki, ale znaczący wzrost uczenia się.

Ta ostatnia, najważniejsza z korzyści nie jest zapewniona przez samo testowanie (Niemierko, 1990, s. 385 i 387), bo „czę-

ste mierzenie temperatury nie wpływa na obniżenie gorączki” (Cizek, 2001). Uzyskiwana jest wysiłkiem nauczycieli. Obie listy dotyczą *efektu zwrotnego* egzaminów, czyli ich konsekwencji jako zmiany systemowej. Która lista w danym przypadku przeważa, to kwestia strategii i realizacji zmiany.

Negatywne argumenty, dowody i przykłady egzaminacyjne łatwiej przemawiają do wyobraźni niektórych pedagogów niż pozytywne. „Jest mnóstwo opisów rozpaczy dobrych uczniów, którym odmówiono dyplomów w wyniku egzaminu doniesłego – skarży się Gregory Cizek (2001) – oraz opowieści o tym, jak to testy zawężają programy kształcenia, frustrują najlepszych nauczycieli, wywołują dławiący niepokój nawet u najzdolniejszych uczniów i powodują u małych dzieci wymioty albo płacz, albo jedno i drugie”. Wciąż za mało jest natomiast chłodnych analiz i solidnej wiedzy o egzaminach.

Podstawowy błąd atrybucji osiągnięć uczniów

W reprezentatywnych dla Stanów Zjednoczonych badaniach opinii nauczycieli o testach standaryzowanych (Herman i Golan, 1993) znalazły się pytania o *presję* wywieraną na nauczycieli, by podnieść wyniki uczniów w tych testach. Jako główne źródło presji badani wskazali źródki masowego przekazu (wskaźnik 3,6 w skali 1–5) – znacznie silniejsze w swym oddziaływaniu niż dyrekcja szkoły oraz inspektorat oświaty (3,2) – i rodziców oraz społeczność lokalna (2,8). Nacisk mediów na szkoły w związku z egzaminami zewnętrznymi jest potężny także w Wielkiej Brytanii i w Polsce.

Co powoduje, że media i opinia publiczna wykazują więcej poparcia dla egzaminów zewnętrznych niż nauczyciele? „[Gdy] społeczeństwo popiera testy standaryzowane – pisze William Rich (2003) – można podejrzewać że lubi ono nie testy, lecz prostotę informacji. Chyba naprawdę odpowiada mu taka łatwa droga do zrozumienia pracy szkoły: za-

stosuj jakiś test i poranguj szkoły od 1 do 10". Następnie Rich przypomina politykę Wielkiego Skoku w Chinach: gdy wytępiono ptaki, które wyjadały zboże, owady uwolnione od swego największego wroga wyjadły zboże doszczętnie. W szkołach także nie wolno naruszać równowagi ekologicznej. „Lekcja, jakiej udzielił nam Mao – podsumowuje autor artykułu – dowodzi, że wielka polityka zesłizguje się z celu i naraża na ryzyko. (...) Gdy chodzi o edukację, to nauczyciele mogą zdecydować, by nie kształcić uczniów w pewnych ważnych dziedzinach”.

Jakie skutki przynosi publikowanie rankingów szkół, celnie nazwanych w Anglii „tabelami ligowymi” (Gipps i Stobart, 1993, s. 44)? „Złej sławy tabele ligowe – napisał Jan Potworowski (2000, s. 56) – podając do wiadomości publicznej rezultaty każdej szkoły, przynoszą tym z górnej partii tabeli nagrodę w postaci zwiększonego naboru [i wzrostu państwowej dotacji], mniej zaś efektywne szkoły karane są utratą uczniów”. To wywołało protesty i interpelacje parlamentarne, na ogół mało skuteczne (William, 2001). Także w Polsce zaobserwowano wzrost różnic w naborze uczniów do poszczególnych gimnazjów i liceów, wynikający z kilkuletniej działalności systemu egzaminów zewnętrznych i publikowania rankingów szkół.

Gdy szkoły są traktowane jako „dobre” i „słabe” na podstawie jednej tylko miary, wyniku egzaminu zewnętrznego, popełniany jest *podstawowy błąd atrybucji* osiągnięć uczniów, polegający na przecenianiu roli szkoły w uzyskiwaniu tych osiągnięć przez uczniów, a niedocenianiu znaczenia sytuacji na wejściu oraz kontekstu w systemie kształcenia. Podobnie jak w przypadku atrybucji dokonywanych przez ucznia, przypisujących sukcesy raczej sobie, a porażki raczej czynnikom zewnętrznym, społeczeństwo przypisuje nauczycielom sukcesy i porażki uczniów w egzaminach zewnętrznych, bez względu na selekcję kandydatów do tych szkół i warunki, w jakich pracują. Ludzie „upatrują przyczynowej roli w tym, na czym skupia się ich uwaga”

(Wojciszke, 2002, s. 105): czytając listę szkół ułożoną według średnich wyników egzaminu, oceniają pracę szkoły, a rozmawiając ze swoim dzieckiem o szkole, oceniają pracę dziecka.

Już Carl Rogers, twórca pedagogiki humanistycznej, stwierdził, że „nauczanie jest działaniem mocno przecenianym” (1969, s. 103). Gdy zaś, zgodnie z tezą Rogersa, akcentujemy *uczenie się*, musimy je rozważać w pełnym cyklu, rozpoczynającym się w domu rodzinnym, a przebiegającym w środowisku i w szkołach niższego szczebla, nie tylko w szkole przygotowującej do danego egzaminu (Fenstermacher i Soltis, 1999, s. 26n). „Na wyniki kształcenia wpływa blisko tysiąc czynników: część z nich wiąże się z pracą szkoły, inne są od niej niezależne” – napisał Krzysztof Kruszewski (1987, s. 9). Trudno wymagać od dziennikarzy, by zajmowali się badaniem tych czynników, ale gdy – na niemal całym świecie – śpieszą oni do czytelników z informacją, które szkoły są „dobre” (a więc zasługują na posyłanie tam dzieci), a które są „słabe” (a więc powinno się ich unikać), potęgują procesy selekcyjne i wywołują rozgorzczenie wśród nauczycieli szkół „słabych”.

Wartość dodana osiągnięć uczniów

Dobrze uzasadnionym teoretycznie sposobem wykorzystania informacji o osiągnięciach uczniów uzyskanej z egzaminu zewnętrznego jest przedstawienie jej w formie *wartości dodanej osiągnięć uczniów*, czyli jako przyrostu osiągnięć uczniów w wybranym zakresie programowym w określonym czasie. Wartość dodana ma pokazać nie stan osiągnięć uczniów, lecz ich *rozwój*, co odpowiada postulatowi psychologii pedagogicznej (Brzezińska, 2000). Średnia *wartości dodanej osiągnięć uczniów* w pewnej szkole może być wskaźnikiem wkładu szkoły w ten rozwój.

O „postępach uczniów” mówi się w pojęciu nauczycielskim i na spotkaniach z rodzicami uczniów od dawna, ale miary tych postępów zaczęto wprowadzać (w Anglii i USA) dopiero w latach

dziewięćdziesiątych ubiegłego stulecia (Saunders, 1999; Doran, Izumi, 2004). Przedtem uważano, że *pomiar zmiany* w osiągnięciach uczniów jako przyrostu ich osiągnięć jest rażąco nierzetelny ze względu na kumulowanie się błędów dwu pojedynczych pomiarów, wstępnego i końcowego, co powodowało, że – z psychometrycznego punktu widzenia – „historia mierzenia przyrostów osiągnięć była długa i smutna” (Hambleton, 1978). Wskutek rozwoju systemów egzaminów zewnętrznych i masowego publikowania „tabel ligowych” stało się konieczne sięgnięcie, w obronie szkół pracujących ze słabiej przygotowaną młodzieżą, do problematyki pomiaru zmiany (Millman, 1997), ale *komplikacja procedur* bynajmniej nie ustąpiła. Jej charakter jest czworaki:

1. *Ograniczenia treściowe*. Podwojenie liczby danych nie poszerza znaczenia treściowego wyników, lecz przeciwnie – wnosi do niego dodatkowe ograniczenia. Zasadnie porównywać możemy tylko poziomy *umiejętności* rozwijanych w sposób ciągły w danym okresie, takich jak czytanie, pisanie, rozumowanie, rozwiązywanie problemów, nie zaś poziomy *wiadomości*, które są pamiętane tylko przez pewien czas. Określając różnice, a nie stany osiągnięć, wpadamy w „psychometryczne sidła”, gdyż mierzymy przede wszystkim wrodzone uzdolnienia umysłowe, a więc – upraszczając – „to, czego szkoła nie nauczyła uczniów” (Popham, 1978, s. 78-85). Wartość dodana jest dla ucznia rozwojowo cenna, ale nigdy nie reprezentuje całej treści kształcenia szkolnego, nawet tylko dziedziny poznawczej.

2. *Ograniczenia kontekstowe*. Możemy wprawdzie założyć, że wstępne osiągnięcia ucznia zawierają w sobie pewną informację o warunkach, w jakich się rozwija, ale urządzenie i organizacja szkoły pozostają poza zasięgiem wskaźnika wartości dodanej. Z tego powodu działanie edukacyjne szkół uboższych materialnie i kadrowo może być niedocenione. Pełniejsza *kontekstualizacja warto-*

ści *dodanej* musiałyby opierać się na oszacowaniach wpływu co najmniej kilku najważniejszych zmiennych kontekstowych osiągnięć szkolnych (Goldstein, 2001), co niewątpliwie gmatwałoby obliczenia, ale zbliżało proces pomiarowy do oceniania społeczno-wychowawczego.

3. *Ograniczenia dydaktyczne.* W niewielu dziedzinach ma sens stosowanie tych samych, a dokładnej – równoległych – wersji narzędzi pomiaru po kilku latach, na przykład tego samego testu matematycznego na początek i na koniec gimnazjum. Wyższy poziom jakości tej samej kategorii czynności w drugim pomiarze powoduje, że *wartość dodana bezwzględna*, wyrażona przyrostem osiągnięć w skali wyników pierwszego pomiaru, nie może być ustalona. Zawsze możliwe jest natomiast oszacowanie *wartości dodanej względnej*, wyrażonej jako zmiana pozycji wyników w populacji badanych uczniów albo szkół. Znaczna większość zastosowań wskaźnika wartości dodanej ogranicza się do interpretacji jego wartości względnej, a więc jest *różnicowaniem* osiągnięć szkolnych, a nie ich sprawdzaniem.

4. *Błędy statystyczne.* Ograniczona rzetelność pierwszego pomiaru i niepełna zgodność treści dwu pomiarów powodują, że *wartość oczekiwana* wyniku drugiego pomiaru, odejmowana, dla uzyskania *wartości dodanej względnej*, od *wartości otrzymanej*, wypada zawsze bliżej średniej wyników niż wynik pierwszego pomiaru. Tę dość złożoną prawidłowość statystyczną objaśniają podręczniki statystyki jako „regresję ku średniej” (Ferguson i Takane, 1997, rozdz. 8). By ją wykorzystać, trzeba albo zastosować „równanie regresji” wobec wyników punktowych dwu pomiarów (Dolata, 2005), albo, co jest prostsze, zmniejszyć odchylenie odpowiedniego stanina od średniego stanina (wynoszącego 5) w proporcji współczynnika korelacji dwu pomiarów, wynoszącego zwykle około 0,7 (Niemierko, 2005). Na przykład, gdy uczeń uzyskał 8 staninów,

to wartość oczekiwana wyników drugiego pomiaru wyniesie około 7 staninów i ona będzie podstawą oceny postępu osiągnięć. Gdyby zaś uczeń uzyskał w pierwszym pomiarze tylko 2 staniny, to wartość oczekiwana wyników drugiego pomiaru wyniosłaby około 3 staninów. Losowe błędy pomiaru, zwiększające rozrzut wyników, najsilniej obciążają wyniki mocno odległe od średniej i gdyby nie brać ich pod uwagę, uczniowie i szkoły o wysokich wynikach mieliby w znacznej większości ujemną wartość dodaną względną, a uczniowie i szkoły o niskich wynikach – dodatnią wartość dodaną względną.

Zawiłości interpretacyjne *wskaźnika wartości dodanej* są trudniejsze do przezwyciężenia niż problemy obliczeniowe, gdzie

w sukurs przychodzi nam technika komputerowa. Powodem, dla którego warto się nim zajmować, jest wyższa trafność oszacowania jakości pracy szkoły za pomocą *wartości dodanej osiągnięć* uczniów niż za pomocą pojedynczego wyniku końcowego w egzaminach zewnętrznych (Bartmańska, 2004). Pomyślnie wypadły próby oszacowania wartości dodanej gmin, szkół i uczniów przez trzyletnią naukę w gimnazjum (Szmigel i Rappe, 2005a), a dyrektorzy gimnazjów stwierdzili, że „chcieliby otrzymywać taką informację, [gdyż ona] może pomóc w planowaniu pracy i motywowaniu zarówno uczniów, jak i nauczycieli” (Szmigel i Rappe, 2005b).

Maria Krystyna Szmigel i Anna Rappe podały przykłady *komentarzy do wartości dodanej* osiągnięć ucznia w gimnazjum, a wśród nich następujące (2005b, s. 152):

KOMENTARZ DO WARTOŚCI DODANEJ OSIĄGNIĘĆ UCZNIĄ

Postęp mały (wynik końcowy niższy o co najmniej 1,5 stanina): Adamie, masz powody do zadowolenia, bo stwierdzono u ciebie postęp w zakresie wiedzy i umiejętności w ciągu trzech lat gimnazjum. Jednak w stosunku do rówieśników twój postęp oceniono jako mały. Sam odpowiedz sobie na pytanie, czy nie poświęć za mało czasu na naukę w gimnazjum? Sądzę, że stać cię na korzystniejsze wyniki kolejnych egzaminów. Zachęcam cię, byś zaczął poważniej traktować swoje obowiązki szkolne! Na dalszy etap kształcenia zapalamy, ku przestrodze, czerwone światło.

Postęp średni (wynik końcowy wyższy lub niższy w granicach 0,5 stanina): Gratuluję ci, Zosiu, że w licznej gromadzie uczniów gimnazjalnych uzyskałaś wynik równy przewidywanemu, a więc w stosunku do swoich rówieśników zachowałaś taką samą pozycję. Czy ten wynik ci wystarcza? Czy jest zgodny z twoimi dalszymi planami edukacyjnymi? Warto, żebyś wiedziała, że są w twojej szkole i klasie uczniowie, którzy w tym samym czasie uczynili większy postęp. Jeśli masz ambicję dorównania czołowce, to nie trać czasu! Czeka cię jeszcze intensywniejsza praca. Powodzenia! Zapalamy żółte światło na początek drogi.

Postęp wysoki (wynik końcowy wyższy o co najmniej 1,5 stanina): Heniu, jestem ciekawa, jak bardzo wynik egzaminu gimnazjalnego cię satysfakcjonuje? Chyba bardzo. Podnieś głowę wyżej i bądź dumny z siebie! Mam dla ciebie dobrą wiadomość: jesteś w grupie uczniów, którzy w ciągu trzech lat nauki w gimnazjum uczynili największy postęp w stosunku do wyniku sprawdzianu po szkole podstawowej. Brawo! Twoje zdolności oraz chęć uczenia się mogą być źródłem wielu sukcesów w przyszłości. Życzę ci wytrwałości! Stawiaj przed sobą coraz trudniejsze zadania. Nie zwalniam, masz zielone światło!

Ten ogólny komentarz, utrzymany w przyjaznym i krzepiącym tonie, powinien być koniecznie uzupełniony informacją o szczegółowych zakresach większego

i mniejszego (niezadowolającego) postępu, tak by jego sprzężenie zwrotne było wzmocnione i usprawnione.

Dydaktyka postępu ucznia

Upowszechnienie szacowania *wartości dodanej osiągnięć* uczniów dałoby szansę na uprawianie *dydaktyki postępu ucznia*, opartej na planowaniu, kontraktowaniu i sprawdzaniu zmiany korzystnej (przyrostu) tych osiągnięć. Ze względu jednak na obciążenia nauczyciela, który byłby tu zobowiązany nie tylko do wielopoziomowego kierowania uczeniem, lecz także do pełnej indywidualizacji treści i metod kształcenia, taka dydaktyka będzie możliwa tylko w przyszłym „społeczeństwie sieciowym”, w którym uczeń uzyska wysoko wydajne środki komunikacji i wielu konsultantów.

Sieci uczenia się, jako związki między uczącymi się, uzyskiwane dzięki coraz tańszemu i coraz bardziej wydajnym środkom komunikacji, będą zapewne w przyszłości stopniowo wyręczać szkoły. O scenariuszu „sieci kształcenia w łonie społeczeństwa sieciowego” – jednym z sześciu scenariuszy przyszłości szkoły do 2020 roku, sporządzonych przez międzynarodowe organizacje ekonomiczno-edukacyjne – tak pisze Maciej Sysło (2005, s. 45):

„Większa waga jest przykładana do kształcenia w różnych kulturach, wartościach, za pomocą sieci pozostających w dyspozycji różnych grup społecznych (...). Powszechnie staje się kształcenie zindywidualizowane, w małych grupach, w domu. (...) Zanika nauczyciel w tradycyjnym znaczeniu, zacierają się lub zanika różnice między nauczycielem a uczniem, rodzicem a nauczycielem, edukacją a społecznością. Pojawia się nowa profesja konsultanta, wykorzystywana w nauczaniu sieciowym, zdalnym, w doradztwie”.

To brzmi dziś jak fantazja. Perspektywy dydaktyki postępu ucznia są jednak rozległe i kuszące. Możemy przewidywać:

1. dalszy wzrost znaczenia *podstawowych umiejętności* ucznia, kosztem encyklopedycznych wiadomości,
2. lepsze dostosowanie *treści i metod kształcenia* do poziomu przygotowania poszczególnych grup uczniów,

w tym zwłaszcza uczniów wolniej uczących się,

3. wyprowadzenie z cienia *szkół i regionów*, które słabo wypadają w dotychczasowych w rankingach, a przyniosą uczniom duży postęp,
4. uprawnienie *komputeryzacji* planowania i oceniania osiągnięć uczniów.

Trzeba będzie uchronić się przed *liberalizmem*, jaki może być uprawiany pod hasłem: „gdy nie ma stałych wymagań, to każdy przejaw postępu jest dobry”, prowadzącym do obniżenia poziomu osiągnięć uczniów. Możliwe jest także ukrywanie umiejętności na wejściu, by wykazać się postępowaniem i zbyt pochopne wnioskowanie o postępie lub regresie osiągnięć ucznia, szkoły lub regionu. Zmiana systemowa spowodowana dydaktyką postępu ucznia byłaby ogromna, co nakazuje wielką ostrożność w próbach jej wprowadzania.

Cytowana literatura

Anastasi A. *Psychological testing. Sixth edition*, New York 1988, Macmillan.

Bartmańska A. *Wartość dodana w komunikowaniu wyników egzaminu a przeciwdziałanie niewłaściwej interpretacji wyników egzaminów zewnętrznych* [w:] B. Niemierko i H. Szaleniec (red.) *Diagnostyka edukacyjna. Standardy wymagań i normy testowe w diagnostyce edukacyjnej*, Kraków 2004, PTDE.

Brzezińska A. *Społeczna psychologia rozwoju*, Warszawa, 2000, Scholar.

Cizek G. *More unintended consequences of high-stakes testing*, „Educational Measurement: Issues and Practice”, 2001, nr 4.

Dolata R., Putkiewicz E., Wilkomirska A. *Reforma egzaminu maturalnego – oceny i rekomendacje*, Warszawa 2004, Instytut Spraw Publicznych.

Dolata R. *Wykorzystanie wyników egzaminów zewnętrznych do oceny jakości kształcenia*, „Biuletyn Badawczy CKE Nr 3”, 2005.

Doran H.C., Izumi L.T. *Putting Education to the Test: A Value-Added Model for California*, San Francisco, 2004, Pacific Research Institute.

Dubiecka A., Szaleniec H., Węziak D. *Efekt egzaminatora w egzaminach zewnętrznych* [w:] B. Niemierko i K. Szmigiel (red.) *O wyższą jakość egzaminów szkolnych. Część I. Etyka egzaminacyjna i zagadnienia ogólne*, Kraków 2006, PTDE.

Fenstermacher G.D., Soltis J.F. *Style nauczania*, Warszawa 2000, WSiP.

Ferguson G.A., Takane Y. *Analiza statystyczna w psychologii i pedagogice*, Warszawa 1997, PWN.

Gipps C. i Stobart G. *Assessment. A teachers' guide to the issues*, London 1993, Hodder.

Goldstein H. *The difficulty of ranking schools. The limits to value added*, „New Economy” 2001, s. 197-198.

Goslin D. *Teachers and testing*, New York 1967, Sage.

Hambleton R.H. *On the use of cut-off scores in educational settings*, „Journal of Educational Measurement” 1978, s. 277-290.

Herman J.L., Golan S. *The effects of standardized testing on teaching and schools*, „Educational Measurement: Issues and Practice” 1993, nr 4.

Kelleghan T., Madaus G.F., Airasian P.W. *The effects of standardized testing*, Boston 1982, Kluwer-Nijhoff.

Knight P.T. *Summative assessment in higher education: An appraisal in disarray*, maszynopis, 2000.

Kruszewski K. *Zrozumieć szkołę*, Warszawa 1987, MAW.

Millman J. *Grading teachers, grading schools. Is student achievement a valid educational measure?* Thousand Oaks, CA, 1997, Corvin Press.

Niemierko B. *Testy osiągnięć szkolnych. Podstawowe pojęcia i techniki obliczeniowe*, Warszawa 1975, WSiP.

Niemierko B. *Pomiar sprawdzający w dydaktyce. Teoria i zastosowania*, Warszawa 1990, PWN.

Niemierko B. *Ocenianie szkolne bez tajemnic*, Warszawa 2002, WSiP.

Niemierko B. *Założone i uzyskane, sprawdzające i różnicujące znaczenie treściowe egzaminacyjnej skali pomiarowej* [w:] B. Niemierko i H. Szaleniec (red.) *Diagnostyka edukacyjna. Standardy wymagań i normy testowe w diagnostyce edukacyjnej*, Kraków 2004, PTDE.

Niemierko B. *Wartość dodana osiągnięć uczniów, szkół i regionów. Studium wstępne do wykorzystania przez CKE*, „Biuletyn Badawczy CKE Nr 3”, 2005.

Plake B.S., Impara J.C., Fager J.J. *Assessment competencies of teachers: A national survey*, „Educational Measurement: Issues and Practice” 1993, nr 4.

Poręba-Konopczyńska A. *Moje refleksje nad ocenianiem na podstawie układu rzetelności i trafności wyników sprawdzania zaproponowanego przez Petera T. Knighta* [w:] B. Niemierko i M.K. Szmigel (red.) *Teoria i praktyka oceniania zewnętrznego*, Kraków 2001, PANDIT.

Potworowski J. *Pierwiastek angielski w transformacji polskiej oświaty* [w:] K. Kruszewski (red.) *Pedagogika w pokoju nauczycielskim*, Warszawa 2000, WSiP.

Rich W. *Historical high-stakes policies relating to unintended consequences of high-stakes testing*, „Educational Measurement: Issues and Practice” 2003, nr 1.

Rogers C. *Freedom to learn*, Columbus 1969, Merrill.

Saunders L. *A brief history of educational „value added”: How did we get to where we are?*, „School Effectiveness and School Improvement” 1999, nr 2.

Smith M.L., Rottenberg C. *Unintended consequences of external testing in elementary schools*, „Educational Measurement: Issues and Practices” 1991, nr 4.

Stożek E. *Zadania „rozrywające” w testach* [w:] B. Niemierko i K. Szmigel (red.) *O wyższą jakość egzaminów szkolnych. Część I: Etyka egzaminacyjna i zagrożenia ogólne*, Kraków 2006, PTDE.

Szmigel M.K., Rappe A. *Przydatność uproszczonej (staninowej) metody szacowania wartości dodanej osiągnięć uczniów, szkół i jednostek administracyjnych oświaty* [w:] B. Niemierko i G. Szyling (red.) *Holistyczne i analityczne metody diagnostyki edukacyjnej. Perspektywy informatyczne egzaminów szkolnych*, Gdańsk 2005, Fundacja Rozwoju Uniwersytetu Gdańskiego.

Szmigel M.K., Rappe A. *Komunikowanie wartości dodanej osiągnięć szkol-*

nych uczniom, nauczycielom i dyrektorom szkół [w:] B. Niemierko i G. Szyling (red.) *Holistyczne i analityczne metody diagnostyki edukacyjnej. Perspektywy informatyczne egzaminów szkolnych*, Gdańsk 2005, Fundacja Rozwoju Uniwersytetu Gdańskiego.

Wiliam D. *Integrating summative and formative functions of assessment*, maszynopis, 2001.

Wojciszke B. *Człowiek wśród ludzi. Zarys psychologii społecznej*, Warszawa 2002, Scholar.

Autor jest naukowcem i pedagogiem, specjalizuje się w dydaktyce ogólnej. Autorytet w dziedzinie pomiaru dydaktycznego

Czy sprawiedliwie oceniał postępowanie uczniów badanych?

Czy zatem sumiennie ważył stwierdzone rzeczywiste zasoby ich wiedzy i na tychże zasobach opierał swój sąd o postępie uczniów, czy też może uwzględniał tylko wykryte w ich wiedzy braki, częstokroć nieuniknione, przypadkowe, pozorne lub dla dalszej nauki i ogólnego wykształcenia zgoła obojętne?

Czy żądając zawsze wiedzy dokładnej i pewnej, większą przywiązywał wagę do rzeczy ważnych, zasadniczych, niż do rzeczy drobnych, podrzędnych?

Czy w żądaniach swych nie zapomniał o tym, że sam powoli w ciągu wieloletnich studiów nie bez trudu zdobył sobie to, czego wymaga od uczniów z bezwzględną niemal dokładnością?

Czy oceniając postępowanie uczniów, zwracał uwagę nie tylko na rzeczywistą ich wiedzę, lecz także na samodzielność ich sądu i rozwój zdolności myślenia?

Czy równą miarą mierzył postępowanie wszystkich uczniów?

Czy bezmyślniej, mechanicznej pracy pamięci nie przedkładał nad myślącą pracę rozumu?

Czy ogólny sąd o postępie uczniów opierał na dostatecznej ilości sądów szczegółowych, zebranych w przeciągu dłuższych okresów czasu?

T. Sołtysik: *Wskazówki dla praktykantów zawodu nauczycielskiego. Zbiór pytań do oceny lekcji próbnych i popisowych* – Kraków 21 grudnia 1914.