

EGZAMINY POŻĄDANE I NIECHCIANE

Referat programowy VIII Ogólnopolskiej Konferencji z cyklu *Diagnostyka Edukacyjna*
DWA RODZAJE OCENIANIA SZKOLNEGO, Katowice, 21 - 23 października 2002 r.

Autor podnosi zagadnienie emocji towarzyszących egzaminom, w tym zwłaszcza u uczniów przyzwyczajonych do korzystania z ulgowego układu wymagań podczas oceniania wewnątrzszkolnego. Egzamin zewnątrzszkolny wykrywa „szarą strefę” osiągnięć uczniów, wobec których zastosowano w szkole obniżone wymagania. Bardzo ważna jest precyzyjna informacja nie tylko o sukcesach, ale i brakach w strukturze osiągnięć uczniów, bo tylko taka informacja jest użyteczna w planowaniu podnoszenia jakości kształcenia. Według autora pomyślnie wprowadzenie egzaminów zewnętrznych stwarza nowe perspektywy diagnostyce edukacyjnej na poziomie szkoły, województwa i kraju.

Jak tytuł sugeruje, mój referat będzie głównie dotyczył emocji związanych z egzaminami. Duża część tych emocji wynika z ujawniania nowego typu prawdy o edukacji. Ta prawda może być przeceniana lub niedoceniana, ale nigdy nie jest obojętna dla jej adresatów.

Nie wszystkie elementy *edukacji publicznej* są przyjmowane z jednakowym przekonaniem. Popierając ją jako całość, społeczeństwo może opierać się poszczególnym działaniom nie tylko ze względów poznawczych, ale i emocjonalnych. W tym drugim przypadku proces zmiany przedłuża się i wikła, a gdy nie jest właściwie kierowany, załamuje się.

To, co ma większe tradycje, budzi mniej oporu. „Lekcja” i „podręcznik” nie wywołują większych wątpliwości, choć zostały podważone przez radykalne szkoły alternatywne i coraz sprawniejsze multimedia. Uzasadnień natomiast potrzebują na przykład *gry dydaktyczne* jako metoda kształcenia, a z drugiej strony - *egzamin*, zwłaszcza w wersji masowej, wyżej zorganizowanej. Te pierwsze wydają się nazbyt rozluźniać, a te drugie - nazbyt usztywniać praktyki edukacyjne. Zderzają się tu stereotypy „pracowitego ucznia” (siedzącego nad książką) i „surowych wymagań” (bezwzględного egzaminatora), a przewyższanie stereotypów wymaga szczególnej cierpliwości (Wojciszke 2002, s. 116 - 119).

Nim zatem przejdę do kondycji egzaminu zewnętrznego w Polsce, zastanowię się nad znaczeniem wszelkiego egzaminu.

1. CZY POTRZEBUJEMY INFORMACJI O WYNIKU UCZENIA SIĘ?

Szeroko rozumiany *egzamin szkolny*, jako sprawdzanie osiągnięć wyodrębnione w procesie dydaktycznym, polega - w najprostszym ujęciu - na upewnianiu się o tych osiągnięciach.

Dlaczego sprawdzanie osiągnięć jest *wyodrębnione*, a nie wplecione w przebieg kształcenia, naturalne i mało widoczne? Dlatego, że jest ono koncepcyjnie i organizacyjnie trudne. Wymaga analizy treści kształcenia, diagnozy kontekstu uczenia się, doboru zadań, kierowania pracą ucznia nad zadaniami, wnioskowania statystycznego, odpowiedniego komunikowania wyniku sprawdzania. Żłudzenie, że to wszystko można uzyskać lekko

i mimochodem, starałem się rozwiać w referacie przedstawionym w Krakowie na VI Konferencji naszego cyklu (Niemierko 2001, s. 14). Tam też próbowałem ostrzec władze oświatowe przed wiarą, że wyprowadzenie egzaminu na zewnątrz szkoły i ujednoczenie zadań znakomicie i ostatecznie go zobiektywizuje (Niemierko 2001, s. 15 - 16).

Co wszakże nas skłania do podejmowania wysiłku *dokładnego sprawdzania* osiągnięć uczniów? Czy gra jest warta „świeczki”, a mianowicie - czasu i pieniędzy, a nade wszystko stresu, jakiego nie udaje się uniknąć w żadnej ze znanych form organizacyjnych wyodrębnionego sprawdzania osiągnięć? O pomoc w rozwiązaniu tak zasadniczej kwestii dydaktycznej zwracamy się, jak zwykle, do psychologów.

Nasz wybitny psycholog Wiesław Łukaszewski już trzydzieści lat temu zebrał dowody tezy, że *informacja o wyniku czynności*, niekoniecznie tylko pomyślna, podnosi jakość tej czynności w kolejnych powtórzeniach (Łukaszewski 1972), bo dążenie do znajomości własnych osiągnięć jest dla ucznia silniejszym motywem działania w sytuacji zadaniowej niż dążenie do poczucia powodzenia. Po trzydziestu latach - co zasługuje na uznanie - autor ten znów przestudiował nagromadzoną literaturę przedmiotu i przekonał się, że nadal nie ma jasności co do tego, w jakim stopniu sukcesy i porażki przyczyniają się do ulepszeń, ale *brak informacji zwrotnej* oraz jej niezróżnicowanie - w postaci wyłącznie wysokich albo wyłącznie niskich ocen - na dłuższą metę zawsze obniżają poziom wykonywania czynności (Łukaszewski 2002).

Przeskakując do przykładu „makro” można by zauważyć, że socjalizm, wyłączając obiektywny sprawdzian gry rynkowej w gospodarce i uczciwe badanie opinii publicznej w polityce, był skazany na zastój i pogarszanie się systemu. To były jednak zmiany dość powolne, nie dla wszystkich obywateli widoczne. Łatwo dziś znaleźć rodaków, którzy w tamtym systemie czuli się lepiej, bo nie byli zmuszeni do walki o jakość, do intensywnej pracy i, co bodaj dla nich jest najstraszniejsze, do podnoszenia kwalifikacji. Jako „masy pracujące” mogli wszak liczyć na komplementy władzy, niby dziecko bezwarunkowo kochane przez czułych rodziców.

Ta analogia makro nie jest jednak dostatecznie bliska szkole, bowiem dzieci nie wybierają sobie nauczycieli, nawet pseudodemokratycznie, i mają co najwyżej ledwie zauważalny wpływ na program kształcenia. Znaczna część uczniów może zatem traktować opanowywanie przewidzianych programem czynności jako *zadania obce*, narzucone im przez szkołę. Niektórzy uczniowie mogą rozwijać się w środowiskach nie doceniających szkoły, inni mogą czuć się w niej źle od początku z powodu braku sukcesów poznawczych i emocjonalnych, jeszcze inni, chyba najliczniejsi, zaczęli szkołę z wiarą w siebie, by stopniowo popaść w *bezradność intelektualną*, nabytą wskutek niemożności sprostania rosnącym wymaganiom programowym.

2. JAKOŚĆ UCZENIA SIĘ A INFORMACJA O WYNIKU

W cytowanym artykule Wiesław Łukaszewski dokonuje - dla pedagogów kardynalnego - podziału *zadań*, czyli wyobrażeń pożądanego wyniku, według dwu kryteriów:

- a) lokalizacji źródła zadania,
- b) precyzji wymagań („zawartości informacji dotyczących wyniku”).

To pierwsze kryterium prowadzi do odróżnienia wspomnianych już *zadań obcych*, z zewnętrzną i zwykle słabą motywacją do danego rodzaju działania, na przykład do uczenia się wybranego przedmiotu szkolnego, oraz *zadań własnych*, z wewnętrzną, zwykle znacznie silniejszą motywacją do działania. Analogią jest tu wiedza „zimna” i „gorąca” (Zajonc 1985). Rosnące liczby uczniów traktujących uczenie się w szkole jako zadanie obce niepokoją pedagogów wielu krajów. W podsumowaniu badań osiągnięć szkolnych *PISA 2000*, w których

Polska brała także udział, na pierwszym miejscu „głównych ustaleń” znalazło się następujące stwierdzenie (OECD 2001, s. 14):

„Biorąc pod uwagę poważne inwestycje w edukację i jej znaczenie dla przyszłego dobrobytu społeczeństw i jednostek rozczarowuje fakt, iż pokaźna część piętnastolatków wyraża negatywne postawy wobec uczenia się i wykazuje brak zaangażowania w szkole. W większości krajów ponad czwarta część ogółu uczniów oświadczyła, że szkoła jest miejscem, w którym nie chcą bywać. Dotyczy to od niespełna 20 procent uczniów w Danii, Meksyku, Portugalii i Szwecji - do ponad 35 procent w Belgii, Kanadzie, Francji, na Węgrzech, we Włoszech i w Stanach Zjednoczonych”.

Drugie kryterium dotyczy - w warunkach szkolnych - wymagań programowych. Od różnia (w mojej terminologii) *zadania wyraźne*, oparte na standardach wymagań programowych, sformułowanych ogólnie lub przez danego nauczyciela i możliwie konsekwentnie stosowanych, od *zadań niejasnych*, mglistych, w przewadze intuicyjnych, podatnych na zmienne interpretacje. Różnicę objaśnia *Mała encyklopedia logiki* (Marciszewski 1970):

„Wyrażenie, które jest zrozumiałe dla pewnej osoby, może mieć dla niej bądź znaczenie wyraźne, bądź znaczenie intuicyjne. Wyrażenie ma dla danej osoby w danym czasie znaczenie *wyraźne*, gdy - skrótkowo mówiąc - osoba ta potrafi w danej chwili wyjaśnić znaczenie tego wyrażenia. Jeśli natomiast jedna osoba w danym czasie potrafi tylko używać wyrażenia, nie umiejąc wyjaśnić jego znaczenia, to mówimy, że wyrażenie ma wówczas dla tej osoby znaczenie *intuicyjne*. (...) Istnieją więc wyrażenia, które się rozumie, ale nie w sposób wyraźny; są to właśnie wyrażenia, które rozumie się intuicyjnie.”

Traktując kryteria lokalizacji źródła zadania i precyzji wymagań jako niezależne, otrzymujemy tabelę, w której imiona uczniów są zapożyczone z książki *Ocenianie szkolne bez tajemnic*, tuż przed naszą konferencją opublikowaną przez Wydawnictwa Szkolne i Pedagogiczne (Niemierko 2002a).

Tabela 1. Cztery modele zadaniowe uczenia się

Zadania	obce	własne
niejasne	Andrzej	Dariusz
wyraźne	Celina	Beata

Posłużymy się teraz modelami Wiesława Łukaszewskiego i biografiami czterech uczniów, zawartymi w mojej książce, by scharakteryzować wybraną czwórkę i opisać jej style zachowań egzaminacyjnych.

Andrzej jest chłopcem silnym, odpornym i sprytnym, ale bez zainteresowań intelektualnych. Słabo czyta i pisze. Szkołę traktuje jako uciążliwy obowiązek, tym przykrzejszy, że z trudem przechodzi z klasy do klasy. Uważa, że to wina nauczycieli, którzy „się czepiają”. Nauczył się stosować strategie obronne w postaci zwlekania z odpowiedzią, przemyślnego ściągania klasówek, inspirowania zbiorowych aktów oporu, przeplatania postaw hardych i dobroduszych oraz wielu podobnych zachowań.

Andrzej nie wie, co będzie na egzaminach zewnętrznych, lecz przyjmuje je jako próbę umiejętności przetrwania. Zaopatruje się w „gotowce” i liczy na nauczycieli. „Niemożliwe, żeby nam nie pomogli. Przecież to oni mieli nas wszystkiego nauczyć!” - mówi. Rozważa też możliwość bojkotu egzaminów, ale nie śpieszy się z podjęciem ryzyka. Demonstruje postawę „spoko”, bo przecież w opalach jest w szkole na co dzień.

Przeciwny biegun motywacji reprezentuje **Beata**. Lubi zadania konkretne, nie wymagające wlotów, które zawsze wykonuje z wzorową starannością. Od swoich nauczycieli oczekuje objaśnienia, co trzeba „wykuć”, by otrzymać czwórkę. Dzięki swej solidności jest dobrą uczennicą. Nauczyciele ją cenią za stosunek do nauki, a uczniowie wybrali starszycią klasy.

Beata przykładą dużą wagę do egzaminów zewnętrznych, choć wypada na nich słabo. Żałuje, że nie potrafi przygotować się do nich dokładnie, bo niektóre zadania są „na inteligencję” i zawierają materiał spoza podręcznika. „Na co komu takie zadania?” - myśli Beata. Godzi się jednak z faktem, że są uczniowie zdolniejsi od niej i że nie wszystkie kierunki studiów będą dla niej dostępne.

Do najzdolniejszych w klasie należy **Celina**. Niestety, wykorzystuje swój talent raczej przeciw szkole. W złość wpada, gdy jest oceniana z lekcji na lekcję. Dobrze wie, czego się od niej wymaga, ale to wcale nie znaczy, że zamierza się tego uczyć. W szkole rozwija pożyteczną działalność ekologiczną i opozycyjną działal-

ność towarzyską. Materiał opanowuje „pięć przed dwunastą”, wprawiając w nieklamany podziw koleżanki. Dostaje trójki. Lubi szkołę za jej bezsilność wychowawczą.

Celina demonstracyjnie lekceważy egzaminy zewnętrzne, ale wie, że stanowią dla niej szansę. Wprawdzie niektóre zadania zamknięte ujawniają jej horrendalny brak wiedzy, lecz inne, rozszerzonej odpowiedzi, stanowią pole do popisu. Potrafi zabłysnąć stylem i oryginalnym pomysłem. „Na egzaminie pokażę figę swoim belfrom!” - chwali się koleżankom. Jest przekonana, że dostanie się na każdy kierunek studiów, na jaki się zdecydował, ale nauczyciele w to wątpią.

Zdolny i pracowity jest **Dariusz**, olimpijczyk w przedmiotach ścisłych. Uczy się nawet więcej niż Beata, ale brak mu jej równowagi psychicznej. Dariusz wciąż sięga do nowych dziedzin naukowych, co cieszy nauczycieli, lecz oddala go od klasy. Drobny, niezgrabny, w okularach, nie może się do niej dostosować. Wolalby nie chodzić do szkoły i uczyć się sam, przy komputerze.

Mimo piątek i szóstek w szkole Dariusz boi się testów, bo ma z nimi nienajlepsze doświadczenia. Często zdarza mu się niezrozumienie zadania, zwłaszcza wyboru wielokrotnego, które, jego zdaniem, „nie może być tak proste”. Bywa, że w schematach punktowania zadań otwartych nie przewidziano teorii, którą właśnie opanował. To wszystko denerwuje Dariusza i prowadzi do huśtawki nastrojów. Myśli o sobie krytycznie i sukces go dziwi. „Egzaminy niszczą wiedzę” - zauważa. - „Zmuszają do uczenia się tego, co nie jest ani ciekawe, ani uczniom potrzebne.” Bardzo dokładnie jednak studiuje sylabusy, narzekając na wieloznaczność standardów.

Można spodziewać się, że Andrzej uzyska na egzaminie zewnętrznym wynik najniższy, Beata niski, Celina - gdy szczęście jej dopisze - średni, a Dariusz wysoki. Jak to odbiorą uczniowie o tak różnych charakterystykach zadaniowych?

Andrzej zgłosi pretensję do egzaminatorów. Raczej nie otwarcie, na forum szkoły, ale w mniejszym gronie, wobec rodziców, kolegów i podobnych rozmówców. Jego zdaniem, organizacja sprawdzania jest zła, zadania dziwaczne, atmosfera policyjna, wynik niesprawiedliwy. Nie widać szans na to, by Andrzej zaaprobował egzamin szkolny w jakiegokolwiek formie.

Beata powstrzyma się raczej od krytyki egzaminu, ale popatrzy i westchnie w sposób, który powie wszystko. Jej pilność i dojrzałość, które zapewniają jej wysoką pozycję w klasie szkolnej, nie przydają się w testowaniu. Im wyższa jest jakość zadań, tym głębiej sięgają one do uzdolnień intelektualnych ucznia, a tym oszczędniej do jego pamięci. Beata stara się jednak wyciągnąć wnioski z egzaminu i postanawia dokładniej nauczyć się tego, co na nim było.

Celinę egzamin zewnętrzny upewnia, że to, czego wymagają nauczyciele, jest mało ważne. Szkołę można przejść indywidualnym szlakiem, orientując się na to, co będzie potrzebne egzaminie końcowym. Większość materiału nie przyda się, więc po co się nim przejmować? Egzamin daje Celinie podstawę do uwolnienia się od podległości nauczycielom i rodzicom.

Egzaminy zewnętrzne zbyt wiele kosztują *Dariusza*. Mimo serii sukcesów nie wierzy w siebie. Dobrze wie, że wszystkiego nie można umieć, a nadto ma spore kłopoty z testową formą egzaminu, w tym zwłaszcza z zadaniami wyboru wielokrotnego. Nie może sobie dawać żadnego straconego punktu, obwinia przy tym zawsze tylko siebie. Słyszając to koledzy pukają się w czoło, co dodatkowo go przygnębia. Rodzice są dumni z *Dariusza*, ale czy na pewno go dobrze rozumieją?

Podsumowując te charakterystyki, możemy stwierdzić, że egzamin zewnętrzny, potraktowany jako *egzamin doniosły*, to jest taki, w którym wynik (promocja, dyplom, przyjęcie do szkoły) bardziej liczy się niż komentarz dydaktyczny, wykazał następujące właściwości:

1. *Zwężenie treści kształcenia*. Nie sposób reprezentować zadaniami wszystkiego, co mieści się w programie kilku lat kształcenia, tym bardziej zaś tego, co wprowadził nauczyciel. Dokonuje się wybór, kierowany interpretacją standardów edukacyjnych oraz względami administracyjnymi, w tym przede wszystkim czasem pracy ucznia i egzaminatora. Celina i Beata, dla których te standardy są w miarę wyraźne, zaczynają więc uczyć się „pod egzamin”.

2. *Neutralność wychowawcza*. Na co dzień nauczyciele wspierają przykładną pilność Beaty i zainteresowania naukowe *Dariusza*, a starają się zwalczyć opór pozostałej dwójki wo-

bec obowiązków szkolnych. Tu żadne wcześniejsze zasługi się nie liczą i do głosu dochodzą uzdolnienia, a także, pośrednio, czynniki społeczne, takie jak kultura językowa środowiska rodzinnego (Konarzewski 1999). Dlatego leniwa i krnąbrna Celina może wyprzedzić pracowitą i układną Beatę, wbrew oczekiwaniom wychowawców tych dziewcząt.

3. *Stresujący przebieg*. Psychologowie dowodzą, że najsilniejszym pozafizjologicznym motywem naszego działania jest *autowaloryzacja*, rozumiana jako „dążenie do obrony, podtrzymania lub nasilenia dobrego mniemania o sobie” (Wojciszke 2002, s. 22 i 149 - 151). Autowaloryzacja prowadzi do przeceniania własnej wartości. Zarówno uczniowie, jak i nauczyciele systematycznie przeceniają osiągnięcia szkolne (Niemierko 1990, s. 290 - 291). Egzaminy ściągają obie strony na ziemię, a stres, jaki przeżywa w tym przypadku uczeń, może być porównany tylko z sytuacją, gdy matka małemu dziecku mówi, że „już nie kocha” lub „sobie pójdzie”. Najbardziej denerwuje się Dariusz, który ma jednak tak niską samoocenę, że wynik egzaminu może mu przynieść tylko ulgę i polepszenie samopoczucia.

4. *Wielopostaciowość odbioru*. Wydawałoby się, że egzamin zewnętrzny, uwolniony od lokalnego kolorytu i personalnych zaszłości, standaryzuje doświadczenia edukacyjne uczniów. Nic podobnego! Każdy z uczniów odbiera go po swojemu: Andrzej jako wolną amerykankę, Beata jako krzyż pański, Celina jako partię szachów, a Dariusz jako ciężką chorobę. Egzamin zewnętrzny nie niweluje psychologicznych różnic indywidualnych, lecz je pogłębia.

3. EMOCJONALNA WARTOŚĆ EGZAMINU

W swobodnych wypowiedziach uczniów na temat egzaminów doniosłych przeważają oceny ujemne. „Matura kojarzy mi się wyłącznie ze stresem, wrzodami na żołądku, nerwową atmosferą w rodzinie. *Ucz się, bo nie zdasz matury!!!* Wartość - żadna” - napisał jeden z licealistów (Fronckiewicz 1998). Egzaminy to loteria i sądy kapturowe, kara boska za popełnione grzechy, a w najlepszym przypadku - zmowa ludzi dorosłych przeciwko młodzieży.

Potrzebę wytworzenia cieplejszych skojarzeń dostrzegł Marek Legutko, dyrektor Okręgowej Komisji Egzaminacyjnej w Krakowie. W artykule *Oswajanie z nowością* (2000) zaproponował aż cztery porównania, które (w nieco zmienionej kolejności) poddam tu systematycznej analizie.

Tabela 2. Wzmacnianie wartości emocjonalnej egzaminu szkolnego przez porównania (według pomysłu M. Legutki)

Objekt odniesienia	Podstawa porównania	Spodziewane emocje ucznia	Akcentowana właściwość pomiaru	Ciężkość egzaminu
<i>Masowe biegi (przełajowe)</i>	Pokonywanie słabości, wartość udziału	chęć udziału, motywacja do wysiłku	niezależność sytuacji egzaminacyjnej (uczniwa próba sił)	Obowiązkowy udział
<i>Obowiązkowe badania lekarskie</i>	uciążliwa procedura, porada specjalistyczna	cierpliwość i zaufanie do prowadzących	rzetelność pomiaru (prawdziwości dokładność informacji)	Brak zapewnienia terapii
<i>„Śpiewać każdy może”</i>	okazja do popisu, wielopostaciowość sukcesu	redukcja lęku, pewność sukcesu	trafność pomiaru (zakres sprawdzanych kompetencji)	Świadomość normy jakości
<i>Jazda figurowa na lodzie</i>	jasność wymagań, profesjonalni sędziowie	przeżycie prawdy, przeżycie piękna	obiektywizm pomiaru (stosowanie standardów)	Ograniczenie wyboru dyscypliny

Analiza dowodzi, że cztery umowne obrazy egzaminu dadzą się zinterpretować w kategoriach niezbędnych właściwości *miaru sprawdzającego*. To jednak wcale nie znaczy, że łatwo je wytworzyć w umysłach uczniów i rodziców. Wręcz przeciwnie, takie rozważania zobowiązują nas do wielu ulepszeń egzaminu, by doświadczenia pokoleń składały się na podobny obraz. Realia i marzenia nie powinny być przeciwstawne.

Nim zajmujemy się realiami, warto przypomnieć, że „każde porównanie kuleje”. Prawa skrajna kolumna tabeli pokazuje, na czym polega ta niedoskonałość. Na egzamin uczeń zwykle nie zgłasza się dobrowolnie, na pomoc bywa za późno, nie każdy „śpiew” jest sukcesem i nie zawsze można wybrać ulubiony przedmiot. Z ułomności porównań ich autor zdawał sobie sprawę i dlatego zaproponował ich od razu kilka. Podobnie postępuje każdy doświadczony wychowawca, gdy pragnie ukształtować korzystne emocje w sercach swych podopiecznych.

4. SZARA STREFA OSIĄGNIĘĆ SZKOLNYCH

Badania osiągnięć szkolnych pokazują, że szkoły nie stosują oficjalnych, deklarowanych wymagań programowych wobec wszystkich uczniów. Przykładem będą wyniki Ogólnopolskich Badań Osiągnięć Uczniów, Nauczycieli i Szkół, dokonanych w 1984 roku (Patrzalek 1988, Nowik 1988).

Tabela 3. Wyniki testów sprawdzających z języka polskiego i matematyki w ogólnopolskich badaniach osiągnięć uczniów szkół podstawowych i średnich w 1984 roku

Przedmiot	Klasa	Zakres programowy	Rodzaj zadań*)	Liczba badanych	Norma ilościowa	Stopnie ndst
Język polski	IV SP	<i>Osiągnięcia konieczne</i>	Z + O	4 460	60 %	25 %
	IV SP	<i>Czytanie ze zrozumieniem</i>	Z	4 463	65 %	30 %
	IV SP	<i>Gramatyka i ortografia</i>	Z	4 450	60 %	48 %
	VIII SP	<i>Wiedza o języku</i>	Z + O	4 400	45 %	49 %
	VIII SP	<i>Czytanie utworu epickiego</i>	Z + O	4 496	35 %	51 %
	VIII SP	<i>Pisanie sprawozdania</i>	O	4 448	70 %	53 %
	IV LO	<i>Wiedza o języku</i>	Z + O	3 289	35 %	37 %
	IV LO	<i>Odbiór poezji</i>	O	3 287	(brak)	59 %
	IV LO	<i>Czytanie poezji współczesnej</i>	Z	3 269	35 %	31 %
	IV LO	<i>snej</i>	Z	3 298	60 %	62 %
	Razem	<i>Środki samokształcenia</i>	Z + O	39 860	52 %	44 %
		Język polski				
Matematyka	IV SP	<i>Osiągnięcia konieczne</i>	O	4 437	75 %	52 %
	IV SP	<i>Test wielostopniowy</i>	O	4 444	75 %	69 %
	VIII SP	<i>Osiągnięcia konieczne</i>	O	2 153	75 %	73 %
	VIII SP	<i>Test wielostopniowy</i>	Z	2 243	75 %	74 %
	IV LO	<i>Test wielostopniowy</i>	Z	3 910	75 %	62 %
	Razem	Matematyka	O + Z	17 187	75 %	66 %

*) Z - zadania zamknięte (wyboru wielokrotnego), O - zadania otwarte (krótkiej i rozszerzonej odpowiedzi)

Próby uczniów wykorzystane w badaniach, których dotyczy tabela 2, były nie tylko duże, lecz także wysoce reprezentatywne, to jest poprawnie wylosowane i uprawniające do wnioskowania o odpowiednich populacjach. Grupy uczniów nie spełniających normy oceny „dostatecznej” są w nich przerażająco wielkie, gdyż sięgają blisko połowy badanych w języku polskim, a dwu trzecich w matematyce. Zważywszy jednak na to, że w pierwszym z tych przedmiotów zastosowano niskie normy ilościowe wymagań (niekiedy nawet 35%), to trzeba uznać, że *większość* uczniów nie spełniła wymagań programowych - z wyjątkiem może języka polskiego w IV klasie szkoły podstawowej - i to niezależnie od szczebla szkoły i formy zadań.

Dlaczego autorzy testów, doświadczeni nauczyciele i wytrawni dydaktycy poszczególnych przedmiotów, zbudowali tak trudne zadania? Nie mogli zbudować łatwiejszych zadań, jeśli mieli reprezentować obowiązujące programy kształcenia, wyrażające w większym stopniu ambicje naukowe, narodowe i polityczne działaczy niż osobiste doświadczenia peda-

gogiczne nauczycieli! Po badaniach pilotażowych (w 1983 roku) przeprowadzono dyskusję i obniżono niektóre normy (poloniści), ale treść zadań została tylko nieznacznie skorygowana.

Podobnie szokujące konfrontacje zdarzyły się także w innych krajach. Na przełomie lat siedemdziesiątych i osiemdziesiątych ubiegłego stulecia Stany Zjednoczone podjęły wysiłek eliminacji *analfabetyzmu funkcjonalnego* u absolwentów szkół średnich, przejawiającego się kłopotami z czytaniem i pisanem, błędami w pomiarze czasu, nieumiejętnością posługiwania się mapą i książeczką czekową, niezrozumieniem tabeli i wykresu oraz podobnymi brakami ogólnocywilizacyjnymi. Ufundowano *programy minimum kompetencji*, oparte na pomiarze *osiągnięć koniecznych*, czyli umiejętności pozwalających uczniom korzystać z różnych źródeł informacji i wykonywać proste zadania życia codziennego. Przepaścią nie do pokonania okazały się różnice między społecznie akceptowalnym minimum osiągnięć maturalnych a wynikami testowania najsłabszych uczniów.

Henry Brickell, dyrektor jednego z amerykańskich instytutów ewaluacji dydaktycznej, opatrzył doświadczenia programów minimum kompetencji następującą refleksją: „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” (Airasian, Madaus i Pedulla 1979, s. 99). Nie dziwny się więc, że po kilku latach ciężkich zmagani pedagogicznych, politycznych i sądowych, programy minimum kompetencji w USA upadły.

Drugim układem wymagań egzaminacyjnych nazwiemy wymagania programowe ograniczone ze względu na specjalne potrzeby edukacyjne uczniów. *Specjalne potrzeby edukacyjne* wykazują uczniowie, którzy „nie potrafią podolać wymaganiom programu nauczania”. Obok dzieci i młodzieży z zaburzeniami rozwoju umysłowego, dysleksją i nadpobudliwych lub zahamowanych ruchowo, których łączna liczba sięga 20 procent populacji, można tu zaliczyć uczniów, których materialne lub moralne warunki uczenia się pozbawiły wstępnego przygotowania poznawczego i charakterystyki emocjonalnej, niezbędnych do opanowanie przewidzianej treści kształcenia. Spośród zwórki uczniów, jakich poznaliśmy w tym referacie, należą do nich Andrzej oraz, w pewnym stopniu, Beata i Celina.

Pionierskie badania nad tym układem przeprowadziła w Stanach Zjednoczonych Susan Brookhart (1993). Nauczyciele, zdaniem Brookhart, starają się wychować uczniów do wytrwałej pracy. „Sens pojęcia *stopień szkolny*” - napisała - „jest dla nauczycieli ściśle związany z zagadnieniem nakładu pracy ucznia. Nauczyciele używają zwrotów w rodzaju *praca, jaką wykonał(a)*, by opisać stopnie i ich stawianie. (...) Stopnie są dla nich czymś, co uczniowie zarabiają - są wynagrodzeniem za wkład pracy na pewnym poziomie. Osiągnięcia mieszczą się w tym pojęciu, ale go nie wyczerpują.”

Jakie, wobec tego, znaczenie dla oceniania szkolnego ma poziom osiągnięć ucznia? „Istnieją dwa standardy sprawiedliwości oceniania: uczeń przeciętny lub lepszy otrzymuje *na co zasłużył*, podczas gdy uczeń słabszy od przeciętnego otrzymuje *szansę*, jeżeli tylko da się to jakoś uzasadnić. (...) *Nie mogłabym źle ocenić ucznia, który się stara*, ponieważ uczeń, który *pilnie pracuje* nie *zasługuje* na złą ocenę. (...) Zalecane procedury oceniania, sugerujące bezkompromisowość, mają ograniczoną wartość dla nauczycieli. Wyniki badań sugerują, że nauczyciele łączą role sędziego i adwokata w różny sposób wobec uczniów o różnych zdolnościach i że to jest właśnie sedno aktu wartościowania” - konkluduje Brookhart.

O drugim układzie wymagań nauczycieli, tworzącym „szarą strefę” osiągnięć uczniów, wiemy na razie niewiele. Jego wstępna charakterystyka obejmuje następujące właściwości:

1. Jest zjawiskiem *międzynarodowym* i *międzysystemowym*. O przykłady nietrudno ani w Polsce, ani w innych krajach. Dotyczy każdej - w zasadzie - nieselekcyjnej szkoły obowiązkowej i wielu szkół wyższego szczebla.

2. Wynika z *ingerencji kontekstu uczenia się* w proces edukacyjny. Nie wszyscy uczniowie mają sprzyjające warunki rozwoju - wewnętrzne i zewnętrzne. W wielu przypadkach te warunki są bardzo dalekie od założeń autora programu kształcenia.

3. Jest *nie w pełni legalny*, zwłaszcza gdy obejmuje uczniów bez formalnych orzeczeń medycznych i psychologicznych. Szkoły legitymują się wymaganiami reprezentującymi pełne programy kształcenia, nie zaś wymaganiami z konieczności obniżonymi.

4. Obejmuje jedynie *dolną część* skali stopni szkolnych. Nauczyciele bronią pilnie, choć nie zawsze skutecznie, wysokich stopni przed inflacją. Wobec dolnej części skali - od stopnia „dopuszczającego” do „dostatecznego” lub „dobrego” - bywają bezradni, gdyż przykre fakty są silniejsze od szlachetnych zamierzeń.

5. Przynosi wyraźny *pożytek indywidualny i społeczny*. Porażki szkolne nigdy nie podnoszą aspiracji edukacyjnych, a strach pomyśleć o społeczeństwie przepelnionym osobnikami, którzy poodpadali ze szkół, do jakich uczęszczali.

6. Może wykazywać *skuteczność dyscyplinarną*, gdyż wprowadza zagrożenie wycofaniem ulgi, ale jest *niewychowawcza*, zwłaszcza wtedy, gdy „zasady gry” są niejasne i zmienne. Uczeń „z laski” promowany traci wiarę w siebie i motywację wewnętrzną.

Zakres zastosowań drugiego układu wymagań jest zwykle wymuszony warunkami pracy szkoły, ale stosunek nauczycieli do tych działań i, co ważniejsze, do uczniów, których działania dotyczą, jest bardzo zróżnicowany. Można tu wyróżnić cztery *strategie dydaktyczne*:

1. Strategia *otwarta - przyjazna*: przyzwolenie.
2. Strategia *otwarta - nieprzyjazna*: zwalczanie.
3. Strategia *ukryta - przyjazna*: zrozumienie.
4. Strategia *ukryta - nieprzyjazna*: potępienie.

Strategie otwarte są lepsze niż ukryte, a przyjazne są lepsze niż nieprzyjazne. Rozwiązanie problemu układów wymagań egzaminacyjnych przyniesie jednak dopiero ewolucja standardów edukacyjnych, programów i metod kształcenia w kierunku *wielopoziomowości kształcenia*, czyli dużego zróżnicowania poziomów uczenia się w jednej, wspólnej klasie. Wymagania najniższe muszą być obniżone, a wymagania najwyższe mogą być podwyższone. Chodzi o to, by - w miarę podnoszenia jakości kształcenia i pomnażania wyposażenia szkoły - zarówno uczeń wybitny, jak i uczeń czyniący niewielkie postępy, mieli warunki spokojnego, skutecznego i ucziwego rozwoju osiągnąć.

5. POMIAROWE ZALETY I WADY DWU RODZAJÓW EGZAMINU

Rozróżnienie *egzaminu wewnętrznego*, regulowanego przez własnego nauczyciela danych uczniów, i *egzaminu zewnętrznego*, regulowanego spoza nauczycielskiego systemu kształcenia, nie jest ostre. Istnieje wiele form pośrednich, „zewnętrzno-wewnętrznych”, jak „stara” matura, podatna na efekt „szarej strefy” wymagań programowych, oraz „wewnętrzno-zewnętrznych”, jak stosowanie baterii testów standaryzowanych, na przykład takich, jak konstruowane w Instytucie Badań Kompetencji w Wałbrzychu. O typie egzaminu decyduje nie oficjalny organizator, lecz alokacja normowanie wyników. Tak więc o stopniu ucznia na „starej” maturze decyduje wyłącznie jego szkoła, natomiast układem odniesienia wyniku badania kompetencji są normy staninowe populacji ogólnopolskiej (Mulała i Sroka 2001).

Gdy porównamy egzaminy wewnętrzne (wewnątrzszkolne) i egzaminy zewnętrzne ze względu na główne postulaty pomiaru sprawdzającego (Niemierko 1990, s. 112 - 121), zarysują się między nimi istotne różnice:

1. Egzaminy zewnętrzne stoją znacznie wyżej od wewnętrznych pod względem *bezsbronności*, rozumianej jako zapewnianie wszystkim uczniom jednakowo sprzyjających wa-

runków do wykazania się opanowaniem przewidzianej treści kształcenia. Nauczyciel nie jest zasadniczo w stanie uwolnić się od swoich wyobrażeń o uzdolnieniach i nakładzie pracy ucznia, toteż wielkie wołanie o „sprawiedliwość” rozlega się we wszystkich ankietach na temat osobowości nauczyciela. Gdy uczeń występuje jako numer kodowy, pokusy nagradzania ulubieńców i zwalczania oponentów nie mają do egzaminatora dostępu.

Dlaczego jednak do ideału jest wciąż jeszcze daleko? Głębiej rozumiana bezstronność obejmuje nie tylko ujednoczenie (lub wyrównanie) zadań i organizacji sprawdzianu, lecz także podobne przygotowanie uczniów, nie wyłączając kulturowych wpływów środowiskowych (Hornowska 1999), oraz pewnego rodzaju *indywidualizację*, niezbędną we wszelkich przypadkach specjalnych potrzeb edukacyjnych. Wyważenie zakresu i głębi tej indywidualizacji jest skomplikowane, co potwierdziły tegoroczne egzaminy w szkołach podstawowych i gimnazjach. Wahania tak rozumianej stronniczości egzaminu zewnętrznego są jednak zaledwie ułamkiem amplitudy niekontrolowanych postaw i nastrojów typowego nauczyciela.

2. *Dokładność punktowania*, rozumiana jako jego zgodność z właściwym kluczem lub schematem punktowania zadań, jest także wyższa w egzaminie zewnętrznym. Decyduje o tym wkład pracy w przygotowanie egzaminu.

Dokładność punktowania to jednak nie tylko statystyczna zgodność między egzaminatorami, lecz także poprawność odczytania odpowiedzi uczniów. Nieporadności w konstrukcji i pomyłki w kluczu punktowania *zadań zamkniętych* można ograniczyć w toku standaryzacji testów, ale nigdy nie da się ich całkowicie wyeliminować i historia pomiaru dydaktycznego zna przypadki wielkich katastrof w tym zakresie (Niemierko 1990, s. 69 - 70). Profesjonalnych egzaminatorów można wyćwiczyć w działaniach zapewniających zgodność wyników punktowania *zadań otwartych*, ale nie wykluczy to, nawet w matematyce (Chodnicki 2002), „gubienia diamentów” w postaci twórczych interpretacji zadań i oryginalnego zapisu działań.

3. Trzecim zakresem widocznej przewagi egzaminu zewnętrznego nad egzaminem wewnętrznym jest *rzetelność*, czyli powtarzalność wyniku egzaminu w podobnych warunkach. Tę przewagę uzyskuje on dzięki czterem okolicznościom: 1) staranniejszej (standaryzowanej) konstrukcji narzędzia pomiaru, 2) wydłużeniu czasu egzaminowania pojedynczego ucznia (w stosunku do jednorazowego odpytywania, ale nie w stosunku do ogółu prac domowych i klasowych), 3) bardziej analitycznemu (więcej zadań, więcej kryteriów) podejściu i 4) zwiększeniu rozrzutu osiągnięć uczniów w próbie przez systematyczne różnice poziomu pracy szkół.

Losowe błędy pomiaru są nie do wyeliminowania ze względu na ograniczoną liczebność każdej realnej próby zadań w stosunku do *uniwersum* zadań, niepełną dokładność punktowania zadań otwartych, zgadywanie odpowiedzi na zadania zamknięte i kilka pomniejszych czynników statystycznych, ale udział głównych źródeł błędu jest lepiej kontrolowany w testach *standaryzowanych*. Zastosowanie takich testów na masową skalę w egzaminach ogólnopolskich oraz upowszechnienie komputerów stwarza w naszym kraju perspektywę stosowania *probabilistycznych modeli pomiaru dydaktycznego*, znacznie bardziej precyzyjnych i elastycznych niż oszacowania klasycznego błędu pomiaru (Szaleniec 2002, Kryniowski 2002).

4. *Trafność pomiaru* powinna być rozumiana szeroko, jako użyteczność wniosków opartych na jego wynikach do określonego celu. Gdy tym celem uczynimy przewidywanie wyników dalszego uczenia się w szkole wyższego szczebla, egzaminy wewnętrzne wykażą pewną przewagę nad egzaminami zewnętrznymi.

Jak to się dzieje, że na przykład wyniki znakomitego testu *Scholastic Aptitude Test* niżej korelują z osiągnięciami studentów pierwszego roku amerykańskiego kolegium niż mniej rzetelne i jawnie zależne od poziomu pracy szkoły oceny dyplomowe (Nitko 1983, s. 550 i n.)? Oddziałują tu silnie dwie spośród przedstawionych już właściwości egzaminu doniosłego: zwięźlenie treści kształcenia i neutralność wychowawcza. To, czego uczeń rzeczywiście uczy się w swojej szkole, i

to, jak się tego uczy, ma większe znaczenie dla jego przyszłych postępów niż próba wybranych umiejętności. Ponieważ jednak różnice wymagań programowych są w naszym szkolnictwie duże, a w szkolnictwie amerykańskim - ze względu na różnorodność kulturową - jeszcze większe, połączenie wyników dwu rodzajów egzaminowania polepsza przewidywanie skuteczności dalszego uczenia się (Niemierko 2001, s. 21).

Trafność jest wyższą właściwością pomiaru niż rzetelność (a także niż bezstronność i dokładność punktowania wyników), toteż jej niedostatek w egzaminach zewnętrznych budzi uzasadnioną troskę specjalistów pomiaru dydaktycznego. Dla zilustrowania problemu przytoczę raz jeszcze model Petera Knighta, rozwinięty przez Arletę Porębę-Konopczyńską (2001):

Tabela 4. Rzetelność i trafność egzaminów szkolnych

Właściwości egzaminu	Wysoka rzetelność	Umiarkowana rzetelność	Niska rzetelność
Wysoka trafność	Perły profesjonalizmu	Nauczyciel - światły profesjonalista	Nauczyciel obdarzony intuicją pedagogiczną
Umiarkowana trafność	<i>Egzaminy zewnętrzne</i>	Nauczyciel - dobry rzemieślnik	Codzienna praktyka szkolna
Niska trafność	P o l a m i n o w e e g z a m i n o w a n i a !		

Tabela ostrzega przed niską trafnością sprawdzania osiągnięć uczniów, gdzie „pola minowe” zagrażają procesowi kształcenia szkolnego. Wysoka rzetelność, nie do uzyskania poza analityczną strategią pomiaru dydaktycznego i bez zastosowania testów standaryzowanych, pojawia się tylko na egzaminach zewnętrznych i u najwyżej zaawansowanych „pomiarowców”. Trafność egzaminu zewnętrznego jest ograniczona jego wyobcowaniem z nauczycielskiego systemu kształcenia. Typowe egzaminy wewnętrzne odbywają się przy niskiej lub umiarkowanej rzetelności, ale przy umiarkowanej lub wysokiej trafności sprawdzania osiągnięć uczniów. Nauczyciel sprawdza - jak umie - to, czego naucza.

5. Najwięcej problemów egzaminacyjnych wiąże się z *obiektywizmem pomiaru*, rozumianym jako reprezentowanie ustalonych wymagań programowych. Ta koronna właściwość pomiaru sprawdzającego jest trudna do analizy i jeszcze trudniejsza do zapewnienia egzaminom. Główną trudność stanowi wskaźnikowanie jakości osiągnięć, opisanej standardami edukacyjnymi, przez liczbę punktów, jakie uczeń musi uzyskać w wyniku egzaminu. Istnieje wiele wyrafinowanych statystycznie technik budowania norm wymagań, bądź na podstawie analizy zadań, bądź na podstawie wyników testowania wybranych grup uczniów (Niemierko 1990, s. 338 - 351), ale żadna z nich nie rozwiązuje problemu społecznego uczniów „tuż pod kreską”.

W egzaminach wewnętrznych ta trudność jest pokonywana przez incydentalne, w miarę potrzeb i - w założeniu - niejawne, *obniżanie wymagań* w postaci:

1. drugiego układu wymagań programowych (redukcji zakresu wymagań),
2. minimalnej normy ilościowej („jeśli odpowie na jakiegokolwiek pytanie”).

W egzaminach zewnętrznych „szara strefa” osiągnięć uczniów ulega prześwietleniu. Ponieważ wszelkie istotne procedury są jawne, drugi układ wymagań mógłby istnieć tylko w postaci wielopoziomowej struktury egzaminu, napotykejacej ostrej sprzeciw polityków oświatowych, protestujących przeciw dyplomom „lepszym” i „gorszym”. Nim społeczeństwo oswoi się z upadkiem mitu „szkoły jednolitej”, pozostaje nam tylko operowanie normą ilościową.

Zastosowanie pomiaru sprawdzającego w egzaminach obejmujących całą populację wiekową przy ustalonej normie ilościowej jest obecnie w Polsce praktycznie niemożliwe! Nawet bardzo niskie normy, konsekwentnie stosowane, doprowadziłyby do katastrofy edukacyjnej wielu uczniów i wiele szkół. Oto jakie byłyby skutki zastosowania wybranych poziomów norm w tegorocznych egzaminach zewnętrznych:

Tabela 5. Wybrane ilościowe normy wymagań a procent uczniów, którzy nie zaliczyliby egzaminu zewnętrznego w 2002 roku^{*)}

Egzamin	Liczba uczniów	Norma ilościowa				
		20%	30%	50%	70%	80%
„Sprawdzian” w kl. VI szkoły podstawowej	537 614	0,7	1,3	9,3	34,4	72,4
Egzamin w kl. III gimnazjum - język polski	562 542	1,0	5,0	26,9	64,1	85,4
Egzamin w kl. III gimnazjum - matematyka	558 257	0,9	5,9	36,6	73,8	87,9

^{*)} Wykorzystano tu raporty Centralnej Komisji Egzaminacyjnej, napisane przez A. Brożka i T. Chrostowską.

Tylko w szkole podstawowej dałoby się utrzymać normę 50 procent, a i to zapewne pod warunkiem zorganizowania kursów wyrównawczych i stworzenia uczniom możliwości powtarzania egzaminu. W gimnazjum - racjonalna norma nie mogłaby przekraczać 30 procent, zaś w przypadku próbnego egzaminu maturalnego z matematyki nawet tak niska norma zbulwersowała społeczeństwo i wpłynęła, jak wiadomo, na odroczenie *Nowej Matury*.

Niskie normy osłabiają *znaczenie treściowe* wyników egzaminu, to jest możliwość ustalenia, co umie uczeń uzyskujący te wyniki. Normy 20 - 30% nazbyt zbliżają się do wartości oczekiwanej (10% i 13%) wyniku „czystego” zgadywania odpowiedzi na zadania wyboru wielokrotnego w obu testach, dopuszczają fragmentaryczną i mało funkcjonalną wiedzę ucznia. Także emocjonalny efekt egzaminu, w którym uczeń nie rozwiązał większości zadań, jest ujemny. W pomiarze sprawdzającym najczęściej stosuje się normy 60 - 80% (Niemierko 1990, s. 332 - 338). „Lepiej mniej a pewniej” jest generalną zasadą budowania standardów edukacyjnych, organizacji kształcenia i sprawdzania osiągnięć szkolnych.

Proces urealniania, a dokładniej - *poziomowania* (stratyfikacji) standardów wymagań egzaminacyjnych, musi w Polsce potrwać co najmniej kilka lat. Jest nieunikniony. Rezygnacja z zewnętrznego sprawdzania osiągnięć uczniów z powodu „złej nowiny”, jakie przynosi, byłaby bowiem przysłowiowym „wylaniem dziecka z kąpielą”, odsunięciem zadania zrationalizowanego kierowania edukacją narodową na dalsze lata.

Teoria pomiaru psychologicznego i dydaktycznego oferuje zaawansowane modele analizy zadań ze względu na prawdopodobieństwo ich rozwiązania przez uczniów o różnym poziomie przygotowania (uzdolnienia). Są to, wspomniane już, *probabilistyczne modele wyniku zadania testowego* (Hambleton i Swaminathan 1985). Podjęte w ostatnich latach i miesiącach, w kilku ośrodkach w Polsce, prace z tego zakresu mogą korzystnie wpłynąć na jakość projektowania narzędzi egzaminu zewnętrznego.

6. ADMINISTRACYJNE ZALETY I WADY DWU RODZAJÓW EGZAMINÓW

Jakie walory mają egzaminy w oczach uczniów, rodziców, nauczycieli, dyrekcji szkół, administracji oświatowej i społeczeństwa? Rozważmy ich „opcje” po kolei.

1. Z pozycji *ucznia* - od Andrzeja do Dariusza - egzamin zewnętrzny ma niewiele zalet, a wiele wad zasadniczych. Teoria czynności sugeruje, że lepiej powinna go znosić młodzież mająca potrzebę uczenia się (Beata i Dariusz) i zorientowana w wymaganiach programowych (Beata i Celina), ale rzeczywistość jest znacznie bardziej skomplikowana. Uczeń najlepiej dostosowany (Beata) najwięcej traci na wycofaniu kryteriów społeczno-moralnych, pozapoznawczych, uczeń w konflikcie wychowawczym z nauczycielami (Celina) może na tym raczej zyskać, uczeń wyobcowany (Andrzej) zademonstruje spryt i „twardość”, a uczeń zdolny i wrażliwy (Dariusz) - lęk i poczucie winy.

Na razie nie widać uczniów, którzy cieszyliby się na egzamin zewnętrzny jak na szlachetne wyzwanie, jak na fascynujący mecz sportowy, indywidualny lub w barwach szkoły. Jeśli egzaminy powszednie są dla nich małym złem koniecznym, to egzaminy doniosłe - dużym. Każdy postęp uczyniony w tym zakresie należałoby przywitać z entuzjazmem.

2. *Rodzice*, jak można sądzić z reportaży, są głównie zainteresowani dwoma aspektami egzaminów zewnętrznych:

- a. żeby ich dziecko mogło dostać się do jak najlepszej szkoły,
- b. żeby w toku egzaminu go nie skrzywdzono.

Przy takich oczekiwaniach łatwo o żal do nauczycieli (o słabe przygotowanie uczniów) i do egzaminatorów (o niewłaściwe ich traktowanie). Gdyby egzamin zewnętrzny, konsekwentnie sprawdzający osiągnięcia, zdyskwalifikował „szarą strefę” oceniania szkolnego, żal byłby o wiele większy. „Dlaczego szkoła udaje, że uczeń umie, co powinien, a egzaminy pokazują, że tego nie umie?” - pytano by. Na takie właśnie pytanie nie potrafili odpowiedzieć organizatorzy amerykańskich egzaminów „minimum kompetencji” (Pipho 1978). Po dwudziestu latach coraz bardziej skłonni jesteśmy sądzić, że żadna - w zasadzie - szkoła nie jest w stanie utrzymać wymagań programowych, jakie deklaruje i zobowiązuje się stosować.

3. Badania pedagogiczne pokazują, że *nauczyciele* zachowują spory dystans wobec egzaminów zewnętrznych, a szerzej - wobec wszelkich wyników testów standaryzowanych (Kelleghan i in. 1982). Przybiera to postać następującą:

- a. „informacja o wynikach pomiaru jest wykorzystywana przez nauczycieli głównie wtedy, gdy zgadza się z ich własną opinią o uczniach”,
- b. „zaufanie nauczycieli do testów rośnie w miarę zdobywania pewności, że wyniki testowania nie zburzą ich opinii o uczniach” (Niemierko 1990, s. 387).

Dlaczego osławianie się nauczycieli z pomiarem trwa długo? By to wyjaśnić, musimy wrócić do cytowanych już poglądów Susan Brookhart i Henry’ego Brickella na temat rozbieżności między postulowanym i realnym poziomem kształcenia. Stała konieczność prowadzenia pracy dydaktycznej z uczniami, *jacy są*, przy obowiązujących programach, podstawach programowych, planach pracy szkoły i innych dokumentach objaśniających, *jacy powinni być*, powoduje, że klasa szkolna tworzy rodzaj „układu odosobnionego”, rodzinnego, nieprzenikliwego, chronionego przed zbyt silnym wpływem otoczenia.

Ten układ można rozbić, uzależniając karierę ucznia wyłącznie od wyników egzaminów zewnętrznych, jednak zmniejszenie roli nauczycielskiego systemu kształcenia niesie ryzyko radykalnego obniżenia jakości uczenia się w nader licznych przypadkach wyłącznie zewnętrznej motywacji i jawnej niechęci do szkoły (Ornstein i Hunknins 1998, s. 364 - 366).

4. Znacznie bardziej otwartą postawę wobec egzaminów zewnętrznych wykazują *dyrektorzy szkół* i przedstawiciele niższych szczebli *administracji oświatowej*. Wydają się je aprobować pod niemal każdym względem (Niemierko 1998; Niemierko 1999, s. 190 - 192), co stwarza zagrożenie zbyt pośpiesznym wdrażaniem tej innowacji.

Skłonność administracji oświatowej do bezkrytycznego ufania wynikom pomiaru dydaktycznego wymaga pogłębionych badań. Gdy słyszy się, że „trzeba nauczycielom pokazać, jakie mają braki”, powstaje obawa, że egzaminowanie uczniów mogłoby posłużyć nie tyle do wspierania procesu kształcenia, ile do rozgrywania batalii z niepokornym personelem szkoły.

Przed takim niebezpieczeństwem przestrzega nas Jan Potworowski, dobrze znający przykre doświadczenia angielskie (Potworowski 2000). W tamtym kraju władze dość szybko przeszły od „wytykania i zawstydzania” szkół o niskich wynikach egzaminów zewnętrznych do „drakońskich środków” w postaci nieustannych wizytacji takich szkół, ich zamykania lub „rekonstytucji”, to jest do karnego zwalniania nauczycieli (Wiliam 2000). Skutki tego rodzaju działań były oplakane. Teza, iż „polityka oświatowa ma chronić ucznia przed złym nauczycielem i złą szkołą, a dobrego nauczyciela i dobrą szkołę przed administracją oświatową” (Kruszewski 2002) okazuje się bardzo trudna do zastosowania.

5. Wszelkie błędy popełnione na szczeblu *Ministerstwa Edukacji* mają rangę polityczną, co znaczy, że wchodzi tu w grę stosunki władzy. Egzaminy jako filar jednej z czterech reform, które miały ostatecznie zerwać z niesłuszną przeszłością, zostały zaangażowane w walkę z nauczycielskim konserwatyzmem. W tym duchu zredagowano wstępny manifest systemu egzaminacyjnego, dobrano czołową kadrę i powołano dekoracyjną Radę Programową.

Inwazja postępowości musiała przynieść pewne skutki ujemne. Tak pisze Maria Groenwald (2002): „Zabrakło wyrozumiałości wobec krytycznych opinii, jak i szacunku dla ich autorów. Zamiast pożądanej tolerancji pojawiło się zjawisko nietolerancji (...) W duchu nietolerancji, osobom krytycznie wypowiadającym się o zewnętrznym egzaminowaniu oraz o jego nie zawsze pozytywnych skutkach przypisywano etykietkę inności i dystansowano się od nich. Z czasem przybierano wobec nich postawę niechęci czy nawet wrogości”. Taką politykę obciąża autorka odpowiedzialnością za przypadki „osobliwej tolerancji” wielu nauczycieli wobec jawnego odpisywania rozwiązań wybranych zadań przez niektórych uczniów w toku egzaminu gimnazjalnego. Nietolerancja zaowocowała złą, fałszywą tolerancją nieuczciwości - „tolerancja została postawiona przed prawdą i odpowiedzialnością”.

Pozytywne i negatywne doświadczenie wielu krajów uczy, że o sukcesie reformy oświatowej decyduje pozyskanie zaufania i współpracy nauczycieli, a nie głoszona ideologia (Kontinen 1993). To jeden z powodów, dla których zmiany nie mogą być nagłe i z reguły nie mieszczą się w jednej kadencji parlamentarnej. Nad przebudową nauczycielskich systemów kształcenia trzeba pracować długo, wytrwale i z niemalym nakładem środków. Mądra kontynuacja zmian zapoczątkowanych przez poprzedników bywa miarą dojrzałości nowej władzy.

6. Najwyższą instancją obywatelską w systemach demokratycznych jest *społeczeństwo*, którego nastroje wyrażają - lepiej lub gorzej - środki masowego przekazu, zaś sumieniem bywają zaangażowani uczeni, specjaliści nauk społecznych.

Do takich uczonych należy Zbigniew Kwieciński, który w swym dziele pod alarmującym tytułem *Wykluczanie* zauważa, że - w związku z upowszechnieniem kształcenia - „pozytywna selekcja do szkół średnich w roku 1972 została zastąpiona w roku 1998 negatywną selekcją do szkół zasadniczych. Wspinanie się do elit poprzez szkołę średnią zastąpione zostało spychaniem najslabszych na margines” (Kwieciński 2002, s. 32).

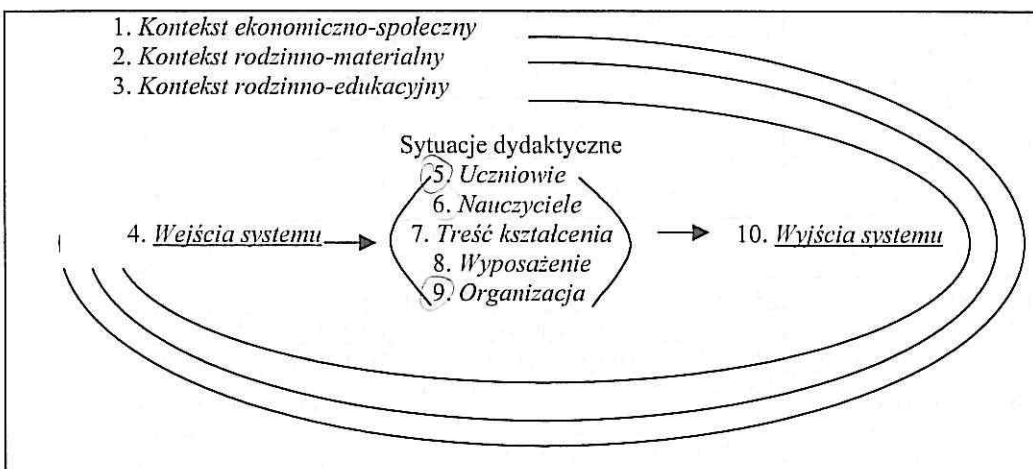
Autor tych spostrzeżeń spodziewa się, że likwidacja szkół zasadniczych zmniejszyłaby „margines” analfabetyzmu funkcjonalnego, choć zapewne - wobec złożoności wszelkich zjawisk z zakresu zdolności i motywacji - nie byłaby w stanie go wyeliminować, skoro nie udaje się to nawet w USA, gdzie szkoła średnia jest od dawna ujednolicona. Pewien optymizm wnosi raport Instytutu Studiów Społecznych, pokazujący wzrost pokoleniowy sztuki czytania, jednak - podobnie jak w badaniach Kwiecińskiego - słabnący w latach dziewięćdziesiątych (Cichomski 2001). Metodologiczną trudnością obydwu badań pozostaje konstrukcja ilościowej normy wymagań, rozdzielającej „analfabetów funkcjonalnych” od „alfabetów funkcjonalnych”.

Spoleczeństwo jako całość nie dysponuje tak szczegółową informacją, jak uczeni i wielkie urzędy, tym bardziej więc jest skłonne do emocjonalnego reagowania na zaobserwowane zjawiska. Jeżeli egzaminy nie mówią całej prawdy o uczniu, to może są niepotrzebne? Jeżeli wyniki wielu uczniów są słabe, to kto jest winien i jak będzie ukarany? Jeżeli niektóre szkoły są słabsze niż reszta, to dlaczego je utrzymujemy? Jeżeli nie ma całkowitej jasności co do wartości egzaminów zewnętrznych, to dlaczego je wprowadzamy? Na takie pytania trzeba odpowiadać spokojnie, rozważnie, z poszanowaniem przekonania i uczuć pytającego. Konferencja, w której uczestniczymy, może zbliżyć nas do takiego wzoru.

7. JAK EGZAMIN ZEWNĘTRZNY WZBOGACA DIAGNOSTYKĘ EDUKACYJNĄ?

Selekcja szkolna jest wybieraniem jednych, a odrzucaniem innych uczniów i z tego względu zawsze będzie podlegać uzasadnionej krytyce społecznej. Inaczej jest z *diagnostyką edukacyjną*, polegającą na rozpoznawaniu i wyjaśnianiu procesów edukacyjnych. Zdobyta w ten sposób wiedza może być wykorzystana do *podniesienia wyników kształcenia*, przeciw czemu nikt rozsądny nie protestuje. Do tego celu musimy znać obecny stan tych wyników, możliwy do rzetelnego zarejestrowania albo w odpowiednio szerokich badaniach osiągnięć szkolnych, albo w toku obowiązkowego egzaminu.

By naszkicować metodologię takiej diagnozy, posłużę się modelem zbudowanym na użytek Instytutu Badań Kompetencji w Wałbrzychu (Niemierko 2002b):



Rysunek 1. Model systemu kształcenia

Model pokazuje, że wyniki kształcenia są uzyskiwane w wielowarstwowym *kontekście uczenia się* (1, 2, 3) w ten sposób, że wejściowy stan osiągnięć uczniów (4) jest przekształcany poprzez sytuacje dydaktyczne (5 - 9) na stan wyjściowy (10). *Stan wejściowy* i *stan wyjściowy* mogą być mierzone wynikami egzaminów zewnętrznych lub wewnętrznych, co kierunkuje diagnozę wężej, *poznawczo*, lub szerzej, ale mniej dokładnie, *wychowawczo*.

Pojęcie *kontekstu*, to jest okoliczności nie poddających się działaniom podmiotu, nie ogranicza się do zmiennych o numerach 1 - 3. W przypadku działalności szkoły niezależna jest także zmienna 5, w przypadku ucznia - także zmienne 6 - 9. Kontekst uczenia się jest zatem niezwykle bogaty, a model przedstawiony na rysunku 1 jest tylko „konstruktem”, umownym zarysem struktury najważniejszych uwarunkowań wyników kształcenia.

Tegoroczne analizy kontekstu wyników krajowego sprawdzianu w VI klasie szkoły podstawowej i egzaminu w III klasie gimnazjum nie wykroczyły poza ekonomiczno-społeczne charakterystyki powiatów, określenie typu gminy oraz porównanie wyników dziewcząt i chłopców. Aby poszerzyć zakres analiz potrzebne są dwa rodzaje wskaźników:

1. *Wskaźniki rzeczowe*, obejmujące dane „twarde”, mierzalne (policzalne) bezpośrednio. Ich dobór może być wzorowany na koncepcji *OECD*, od około 20 lat systematycznie stosowanej i ulepszanej do celu monitorowania rozwoju systemów edukacyjnych w skali międzynarodowej, „makro” (Bogaj i in. 2001).

2. *Wskaźniki szacunkowe*, obejmujące dane „miękkie”, mierzalne pośrednio, poprzez odczucia i poglądy uczniów, nauczycieli i dyrektorów szkół. Każdy z tych wskaźników powinien być liczbowym wynikiem skali postaw (opisowej, szacunkowej), możliwie starannie wystandaryzowanej do swej funkcji. Jego trafność jest zwykle wyższa w skali „mikro”, bo poszerzanie kręgu respondentów rozmywa znaczenie wskaźnika. Pewne postępy w zakresie konstrukcji takich wskaźników uczyniono pod (buńczucznie brzmiącym) hasłem „mierzenia jakości pracy szkoły” (Wlazło 1999).

Plan i przykłady operacjonalizacji modelu systemu kształcenia przedstawia tabela 6:

Tabela 6. Wskaźniki rzeczowe i szacunkowe dla dziesięciu zmiennych w modelu systemu kształcenia przedstawionym na rysunku 1

Zmienna	Wskaźnik rzeczowy		Wskaźnik szacunkowy - respondent
	treść (przykład)	źródło	
1. Kontekst ekonomiczno-społeczny	Stopa bezrobocia w powiecie	Dane GUS	Dyrektor szkoły
2. Kontekst rodzinno-materialny	Liczba osób na pokój w mieszkaniu	Uczeń	Uczeń
3. Kontekst rodzinno-edukacyjny	Wykształcenie rodziców	Uczeń	Nauczyciel
4. Wejścia systemu	Osiągnięcia poznawcze uczniów w szkole niższego szczebla	OKE	Dyrektor szkoły
5. Uczniowie	Przewidywane wykształcenie ucznia	Uczeń	Nauczyciel
6. Nauczyciele	Stopnie zawodowe nauczycieli	Dyrektor szkoły	Uczeń
7. Treść kształcenia	Średnia ocen łatwości zadań testu	Nauczyciel	Uczeń
8. Wyposażenie	Liczba uczniów na jeden komputer w szkole	Dyrektor szkoły	Uczeń
9. Organizacja	Liczba uczniów na jednego nauczyciela w szkole	Dyrektor szkoły	Nauczyciel
10. Wyjścia systemu	Osiągnięcia poznawcze uczniów w danej szkole	OKE	Dyrektor szkoły

Przydatność wyników egzaminów wewnętrznych i zewnętrznych oraz innych wskaźników zestawionych w tabeli 6 do interpretacji pracy szkoły podlega weryfikacji empirycznej. Wskaźniki szacunkowe wydają się niezbędne, gdyż mimo - a może raczej dzięki - swej „miękkości” mogą dobrze rejestrować emocjonalny aspekt kształcenia i stan emocji związanych z procesem uczenia się i egzaminowania w wybranym systemie kształcenia.

PODSUMOWANIE

Główne tezy mojego referatu można streścić następująco:

1. Egzaminy doniosłe obfitują w *emocje*, zwłaszcza w początkowej fazie wdrożeń. Nie wszystkie emocje podpowiadają racjonalne rozwiązania, ale żadnych nie wolno lekceważyć.
2. *Informacja* o wyniku kształcenia jest niezbędna do podnoszenia jego jakości. Dotyczy to nie tylko informacji o sukcesach, lecz także informacji o brakach.
3. Uczniowie reagują na egzamin doniosły w zgodzie ze swym *stylem uczenia się*. Wysoka motywacja i jasność wymagań mogą, ale nie muszą sprzyjać wysokim wynikom.
4. Większości uczniom egzamin szkolny, w tym zwłaszcza egzamin zewnętrzny, kojarzy się ujemnie. Dominuje poczucie *zagrożenia i bezradności*.
5. Wytworzenie *dodatnich skojarzeń* egzaminu jest trudne. By je uzyskać, trzeba nie tylko oddziaływać na emocje uczniów, lecz także dostarczyć im właściwych doświadczeń.
6. W większości krajów około 1/5 - 1/3 uczniów w wieku gimnazjalnym oświadcza, że wolaliby nie chodzić do szkoły. Ta grupa określa zadania szkolne jako *obce*.
W każdym kraju egzaminy zewnętrzne wykrywają rozległą „szarą strefę” osiągnięć uczniów, wobec których zastosowano obniżone wymagania. Winą obarcza się nauczycieli.
8. Egzaminy zewnętrzne górują nad wewnętrznymi właściwościami *organizacyjnymi i technicznymi*, ale egzaminy wewnętrzne górują *trafnością* programową i rolą *wychowawczą*.
9. *Obniżanie* normy wymagań do poziomu poniżej 50 procent powoduje utratę znaczenia treściowego wyników. Dotyczy to zwłaszcza testów złożonych z zadań zamkniętych.
10. Ze względu na „szarą strefę” osiągnięć uczniów jednopoziomowe zewnętrzne egzaminy *sprawdzające* są praktycznie niemożliwe. Pozostaje *różnicowanie* osiągnięć uczniów.
11. Egzaminy zewnętrzne mają wiele zalet i zastosowań z pozycji *administracji* szkolnej. W oczach uczniów, nauczycieli i opinii publicznej ich zalety wyłaniają się wolniej.
12. Pomyślnie przeprowadzenie egzaminów zewnętrznych na dwu szczeblach systemu szkolnego stwarza nowe perspektywy *diagnostyce edukacyjnej* w skali szkoły i w skali kraju.

Bolesław Niemierko

LITERATURA

- Airasian P. W., Madaus G. F., Pedulla J. J., 1979, *Minimal competency testing*, Englewood Cliffs, Educational Technology Publications.
- Bogaj A., Kwiatkowski S.M., Piwowarski R., 2001, *Wskaźniki edukacyjne: Polska 2000*, Warszawa, IBE.
- Brookhart S., *Teachers' grading practices: Meaning and values*, „Journal of Educational Measurement” 1993 nr 2.
- Chodnicki J., 2002, *Porównanie efektywności różnych schematów oceniania zadań otwartych z matematyki na egzaminach powszechnych*. Maszynopis.
- Cichomski B. (red.), 2001, *Determinanty i konsekwencje niskiego stopnia alfabetyzmu funkcjonalnego Polaków. Rekomendacja dla reformy systemu edukacyjnego*. Warszawa, Instytut Studiów Społecznych UW.
- Fronckiewicz E., 1998, *Matura jako źródło stresu*, (w:) B. Niemierko i E. Kowalik, *Perspektywy diagnostyki edukacyjnej*, Gdańsk, Wyd. UG.
- Groenwald M., 2002, *Egzamin z (nie)tolerancji*, Referat przygotowany na VIII Konferencję z cyklu „Diagnostyka Edukacyjna” w Katowicach.
- Hambleton R.K., Swaminathan H., 1985, *Item response theory. Principles and applications*, Boston, Kluwer-Nijhoff.
- Hornowska E., 1999, *Stromniczość testów psychologicznych. Problemy - kierunki - kontrowersje*, Poznań, Humaniora.
- Kellegan T., Madaus G. F., Airasian P.W., 1982, *The effects of standardized testing*, Boston, Kluwer, Nijhoff.

- Konarzewski K., 1999, *Dylematy oceniania osiągnięć szkolnych*, „Kwartalnik Pedagogiczny” nr 2/1999.
- Konttinen R., 1993, *Teachers' perception of the effects of a high stakes examination. The case of the Finnish matriculation examination*, Tekst zaprezentowany na konferencji Council of Europe Educational Research Workshop. Jyväskylä, Finlandia.
- Kruszewski K., 2002, *Od tłumacza*. Wstęp do książki *Klasa szkoły*. Maszynopis.
- Kryniewski M., 2002, *Zastosowanie probabilistycznej teorii wyniku zadania testowego do analizy testów*, (w:) B. Niemierko (red.), *Ocenianie szkolne, ekonomika i polityka oświatowa, probabilistyczne teorie pomiaru*. Skrypt nr 3 dla uczestników III Podyplomowego Studium Ewaluacji Dydaktycznej na Uniwersytecie Gdańskim.
- Kwieciński Z., 2002, *Wykluczanie*, wydanie drugie poszerzone, Toruń, Wyd. UMK.
- Legutko M., 2000, *Oswajanie z nowością*, „Gazeta Wyborcza. Gazeta w Krakowie” 27 listopada.
- Łukaszewski W., 1972, *Informacje dotyczące osiągniętych wyników a poziom wykonania zadań umysłowych*, Wrocław, Ossolineum.
- Łukaszewski W., 2002, *Zwrotne informacje o wyniku czynności*, (w:) I. Kurcz i D. Kądzielawa *Psychologia czynności. Nowe perspektywy*, Warszawa, Scholar.
- Mulawa J. i Sroka D., 2001, *Od standardów wymagań do standardów osiągnięć uczniów klas trzecich szkół podstawowych*, (w:) B. Niemierko i W. Malecki (red.), *Dawne i nowe formy egzaminowania*, Wrocław, Wyd. DSWE.
- Niemierko B., 1998, *Ewaluacja sprawdzianu kompetencji przeprowadzonego w szkołach podstawowych województw śląskiego i piłskiego*. Maszynopis.
- Niemierko B., 1999, *Pomiar wyników kształcenia*, Warszawa, WSiP.
- Niemierko B., 2001, *Chłodne oblicze egzaminu zewnętrznego*, (w:) B. Niemierko i M. K. Szmigel (red.), *Teoria i praktyka oceniania zewnętrznego*, Kraków, PANDIT.
- Niemierko B., 2002a, *Ocenianie szkolne bez tajemnic*, Warszawa, WSiP.
- Niemierko B., 2002b, *Projekt instrumentarium pomiaru kontekstu uczenia się w szkole podstawowej i gimnazjum na użytek Instytutu Badań Kompetencji w Wałbrzychu*. Maszynopis.
- Nowik J. (red.), 1988, *Matematyka. Osiągnięcia uczniów z języka polskiego. Materiały ogólnopolskich badań osiągnięć uczniów, nauczycieli i szkół. Tom V*, IKN.
- OECD, 2001, *Knowledge and skills for life. First results from PISA 2000. Executive summary*. www.pisa.oecd.org.
- Ornstein A.C., Hunkins F.P., 1998, *Program szkolny. Założenia, zasady, problematyka*, Warszawa, WSiP.
- Patrzalek T. (red.), 1988, *Osiągnięcia uczniów z języka polskiego. Materiały ogólnopolskich badań osiągnięć uczniów, nauczycieli i szkół. Tom V*, IKN.
- Pipho C. (red.), 1978, *Minimum competency testing*, „Phi Delta Kappan”, Special Issue.
- Poręba-Konopczyńska A., 2001, *Moje refleksje nad ocenianiem na podstawie układu rzetelności i trafności wyników sprawdzania zaproponowanego przez Petera T. Knight*, (w:) B. Niemierko i M. K. Szmigel (red.), *Teoria i praktyka oceniania zewnętrznego*. Kraków, PANDIT.
- Potworowski J., 2000, *Pierwiastek angielski w transformacji polskiej oświaty*, (w:) K. Kruszewski (red.), *Pedagogika w pokoju nauczycielskim*, Warszawa, WSiP.
- Sędek G., 1995, *Bezradność intelektualna w szkole*, Warszawa, PAN.
- Szalencic H., *Probabilistyczne modele wyniku zadania testowego (Item Response Theory)*, (w:) B. Niemierko (red.), *Ocenianie szkolne, ekonomika i polityka oświatowa, probabilistyczne teorie pomiaru*. Skrypt nr 3 dla uczestników III Podyplomowego Studium Ewaluacji Dydaktycznej na Uniwersytecie Gdańskim.
- William D., 2000, *Integrating summative and formative functions of assessment*. Maszynopis.
- Wlazło S., 1999, *Mierzenie jakości pracy szkoły. Części: pierwsza, druga i trzecia*, Wrocław, MarMar.
- Wojciszke B., 2002, *Człowiek wśród ludzi. Zarys psychologii społecznej*, Warszawa, Scholar.
- Zajonc R. B., 1985, *Uczucia a myślenie: nie trzeba się domyślać, by wiedzieć, co się woli*, „Przegląd Psychologiczny” r. 28, s. 27 - 72.