prof. dr hab. Grażyna Wieczorkowska
mgr Piotr Radkiewicz
Instytut Studiów Społecznych Uniwersytetu Warszawskiego

Magia liczb o ocenianiu w edukacji

Po okresie, kiedy w Polsce decydowało uznanie czynników politycznych a nie wartości merytoryczne, wahadło przemian przechyliło się w drugą stronę. Podejmowane są coraz częściej wysiłki, aby osiągnięcia edukacyjne zarówno osób, jak i instytucji przedstawiać w postaci liczbowej, nie pozostawiając marginesu swobody na interpretację. Artykuł ten jest efektem kilkuletnich doświadczeń Instytutu Studiów Społecznych związanych z pracami nad konstrukcją i diagnozą metodologiczną narzędzi pomiarowych w edukacji i analizą takich narzędzi wykorzystywanych w innych krajach. Traktując skrótowo opis studiów badawczych, chcemy przedstawić na ich tle podstawowe założenia i cele leżące u podstaw tworzenia testów osiągnięć oraz zwrócić uwagę na **zalety i wady tego typu narzędzi**. Jest to tekst o powodach, dla których warto zajmować się pomiarem edukacyjnym i stosować go w praktyce, ale także chcemy powiedzieć o ograniczeniach tego pomiaru i związanych z tym zagrożeniach.

Przedmiotem tego artykułu jest dająca się ostatnio zauważyć, coraz silniejsza tendencja do kwantyfikacji osiągnięć w edukacji — na różnych etapach procesu kształcenia. Zjawisko to znajduje swoje uzasadnienie, kiedy przyjrzymy się stosowanym w Polsce tradycyjnym sposobom oceniania. Nie ulega wątpliwości, że **informacja zwrotna o stopniu opanowania danej dziedziny wiedzy jest niezbywalnym elementem procesu nauczania**. Od zawsze jej źródłem był nauczyciel, który zadawał pytania i oceniał odpowiedzi. Ocena nawet najlepszego nauczyciela jest, niestety, w dużej mierze subiektywnym odbiciem jego osobistych preferencji, kontekstu sytuacyjnego a nawet samopoczucia. Pokazano, że to samo wypracowanie oceniane przez różnych nauczycieli może otrzymać oceny od 1 do 6. Więcej, to samo wypracowanie w zależności od kontekstu i samopoczucia może u tego samego nauczyciela uzyskać bardzo różne oceny. Ponadto, ta sama ocena u jednego nauczyciela często oznacza zupełnie co innego niż u innego nauczyciela. Ocena jest też często efektem zgeneralizowanej opinii o uczniu, ustalanej przez gremium nauczycielskie. Jak pokazują nasze badania, na podstawie jednej grupy przedmiotów (np. z języka polskiego lub z historii) można przewidywać stopnie uzyskane z pozostałych grup przedmiotów (np. matematyki, fizyki, chemii lub biologii). Nie ma żadnego powodu, aby zakładać, że zdolności i zainteresowania matematyczne i humanistyczne są ze sobą tak wysoko skorelowane. Wyniki w testach z eksperymentu wałbrzyskiego są dużo słabiej skorelowane niż stopnie z odpowiadających im przedmiotów. Inne badania (Cichomski, 1997) wskazują, że szkolne oceny semestralne silniej odzwierciedlają nierówności pochodzenia społecznego i różnice płci, niż różnice w poziomie zdolności intelektualnych. Tak więc ocena jest dodatkowo obciążona środowiskowym „dziedzictwem" ucznia.

Odwołując się do tych danych, bez wielkiej przesady można uznać, że zamiast wyrównywać szanse oceny szkolne konserwują, a nawet jeszcze bardziej pogłębiają nierówności społeczne. Co więcej, subiektywizm i arbitralność ocen w praktyce uniemożliwiają porównanie poziomu wiedzy uczniów oraz jakości kształcenia szkół. Staje się oczywiste, że **polska edukacja potrzebuje bardziej obiektywnych metod pomiaru wiedzy**. W tym celu musimy umieć konstruować **wystandaryzowane, obiektywne i rzetelne** narzędzia — testy osiągnięć.

W ilościowym i jakościowym rozwoju testów osiągnięć zdecydowanie przodują kraje anglosaskie, a w szczególności USA. W Stanach Zjednoczonych nikt już nie wyobraża sobie systemu edukacji bez takich narzędzi jak SAT, GRE czy cała gama testów przedmiotowych mierzących poziom osiągnięć z konkretnych dziedzin wiedzy (ang. Subjective Tests). Wiodącą instytucją, która zajmuje się tworzeniem testów, odpowiada za ich metodologiczną jakość oraz nadzoruje kolejne etapy i procedury ich obiegu (dystrybucja, egzamin,

sprawdzanie, przekazywanie i przechowywanie wyników) jest Educational Testing Service w Princeton. Trwałe miejsce testów osiągnięć w systemie edukacji USA opiera się na założeniu, że niezbędnym elementem procesu oceniania jest ujednolicony, rozciągnięty w czasie i oparty na **zewnętrznym kryterium** pomiar. Pomiar, który po przekształceniu w pewien obowiązujący system, zapewnia to, czego, ze względu na zróżnicowanie i arbitralność wskaźników lub subiektywizm ocen, nie zapewniają tradycyjne systemy oceniania — **porównywalność ocen uczniów i porównywalność jakości kształcenia w różnych szkołach**.

W tym miejscu chcemy przejść do bliższego opisu dwóch prób konstrukcji i zastosowania pomiarów testowych na różnych poziomach kształcenia, w które to przedsięwzięcia, za każdym razem na nieco innym etapie, zaangażował się Instytut Studiów Społecznych UW. Pierwsza z nich wiąże się z problemem **oceny jakości kształcenia na studiach wyższych**, a wynika głównie z uzasadnionych obaw — potwierdzanych codzienną obserwacją i raportami kompetentnych instytucji — dotyczących jakości kształcenia i wymogów, które powinny spełniać uczelnie. Ma to przede wszystkim związek z żywiołowym rozwojem płatnych studiów oferowanych zarówno przez nowo powstające szkoły prywatne, jak i uczelnie państwowe

Pojęcie *jakości kształcenia* w szkolnictwie wyższym nie jest jednoznaczne (por. przegląd E. Wnuk-Lipińskiej i M. Wójcickiej (1997)). Najczęściej znajdują tu zastosowanie dwa podejścia: wynikowe, zorientowane na rezultaty (Cross-Durrant 1992) oraz podejście nawiązujące do standardów jakości, oparte na analizie warunków uznanych za sprzyjające osiąganiu oczekiwanego rezultatu (ISO 1987).

Większość prób określania jakości kształcenia **na podstawie wyniku czy rezultatu kształcenia** prowadzi do szerokiego wykorzystywania różnego rodzaju wskaźników osiągnięć. Używa się ponad 50 różnego rodzaju wskaźników (Scott 1994). Są one odnoszone zarówno do rzeczywistości uczelnianej (w relacji do studenta), jak i poza nią wykraczającej (w relacji do absolwenta). Dla przykładu, można tu wymienić wskaźniki dotyczące zatrudnienia absolwentów, sukcesu w zawodzie, przydatności w pracy wiedzy i umiejętności zdobytych na studiach, czy takie jak: terminowość kończenia studiów i poszczególnych etapów kształcenia (semestr, rok), koszty kształcenia. Możliwość odwołania się do tak szerokiego zakresu wskaźników wydaje się mało realna i nieprzydatna w sensie poznawczym.

Drugi sposób podejścia — **jakość w relacji do szeroko rozumianych warunków kształcenia** — oznacza koncentrację na podejmowanych w uczelniach działaniach na rzecz zapewnienia i utrzymania wysokiego poziomu kształcenia. W tym przypadku ocenie podlega nie sam **proces czy wynik kształcenia**; na pierwszy plan wysuwają się strategie instytucjonalne (uczelnia, wydział/kierunek) i w relacji do nich rozpatrywane są szeroko rozumiane warunki realizacji przyjętych założeń edukacyjnych. Inaczej mówiąc, nie tyle „produkt"' czy wynik kształcenia, ile raczej próba odpowiedzi na pytanie, czy mechanizmy i procedury stosowane w danej jednostce organizacyjnej mogą dawać wystarczające gwarancje jakości. Takie podejście można określić jako *wewnętrzny system zapewniania jakości*.

Instytut Studiów Społecznych, przy wsparciu Fundacji im. S. Batorego, podjął próbę budowy narzędzia oceny jakości kształcenia, którego cele i założenia metodologiczne idą w nieco innym kierunku niż opisane powyżej metody. Doceniając zalety zarówno podejścia zorientowanego na rezultaty, jak i drugiego, opartego na standardach jakości, stwierdziliśmy, że niezbędnym elementem procesu oceniania jest wystandaryzowany i opierający się na **zewnętrznym kryterium** pomiar podstawowych umiejętności akademickich. Pomiar eliminujący zróżnicowanie i arbitralność wskaźników oraz subiektywizm ocen, a w konsekwencji dający szansę na to, czego nie zapewniają wymienione wyżej podejścia — **porównywalność jakości kształcenia na różnych uczelniach**.

Uznaliśmy, że planując ogólną strategię monitorowania jakości kształcenia, od której kształtu zależałby wybór konkretnej metody pomiaru, można zdefiniować przedmiot pomiaru jako **przyrost określonych umiejętności w czasie**. Obiektywny, ilościowy pomiar dostarczałby dodatkowych informacji, których nie można uzyskać stosując tradycyjne w szkolnictwie wyższym metody oceniania. Ponadto, przyjmując, że możliwe jest zdefiniowanie pojęcia podstawowych umiejętności akademickich, jest on dobrą podstawą porównywania poziomu kształcenia pomiędzy różnymi uczelniami. Założenie to oparte jest na przekonaniu, że niezależnie od wiedzy, którą nabywają studenci na różnych uczelniach i kierunkach studiów, istnieją pewne ogólne umiejętności, których kształtowanie i rozwój powinny być podstawowym zadaniem każdej uczelni. Ich wysoki

poziom zdecydowanie zwiększa szansę na to, że uczelnie będą opuszczały osoby kreatywne, łatwo adaptujące się do zróżnicowanych i zmieniających się warunków na rynku pracy.

Definiowanie jakości kształcenia w kategoriach przyrostu określonych umiejętności narzuca dwa rodzaje podejść badawczych. Z jednej strony przyrost umiejętności można śledzić powtarzając pomiar tych samych studentów w czasie. Niestety, ze względu na ograniczenia czasowe programu, metoda ta została wykluczona. Alternatywą jest porównywanie wyników studentów poszczególnych lat studiów. Takie rozwiązanie zastosowano w Kanadzie, gdzie tzw. wartość dodaną mierzy się poprzez porównywanie stopni z egzaminów, jakie studenci uzyskują na 1 i 4 roku studiów. Niestety, w warunkach polskich przeszkodą jest brak obiektywnych i sprawdzonych zasad oceniania w zakresie konkretnych przedmiotów akademickich. Oceny wykładowców nie są porównywalne. Najlepszym rozwiązaniem w tej sytuacji jest próba skonstruowania wystandaryzowanych, zewnętrznych narzędzi oceny dających podstawę do dokonywania porównań pomiędzy uczelniami. Powinny one mieć formę testu zamkniętego, o wysokiej rzetelności i zobiektywizowanych procedurach przeprowadzania i sprawdzania. Treść pytań testowych powinna odnosić się do takich umiejętności, których rozwój jest szczególnie ważny w procesie kształcenia akademickiego i względnie niezależny od profilu uczelni.

Założyliśmy, że przedmiotem porównań będą wyniki w przygotowanych przez nas testach. Wyniki te pozwoliłyby odpowiedzieć na pytanie, **czy przyrosty kompetencji akademickich w uczelni A są wyższe od przyrostów w uczelni B**. W tym celu wybraliśmy dwie szkoły wyższe różniące się poziomem kształcenia. Zaplanowaliśmy, że w obydwu szkołach, w ciągu roku akademickiego, zbadany zostanie poziom umiejętności dwóch roczników — studentów I roku i studentów III roku. Różnice w poziomie osiągniętych wyników pomiędzy oboma rocznikami traktowane są jako wskaźnik przyrostu umiejętności studentów, a analiza różnic pomiędzy rocznikami może służyć do porównania poziomu kształcenia pomiędzy szkołami. Zdolność narzędzi testowych do trafnego wykrywania takich różnic jest miarą ich przydatności w procesie monitorowania jakości kształcenia w szkołach wyższych. Można zakładać, że studenci renomowanych uczelni otrzymają zawsze lepsze wyniki niż studenci „gorszych" uczelni, ponieważ mają lepszych kandydatów i prowadzą selekcję „na wejściu". Przyjęta metoda powinna być niewrażliwa na wyjściowy poziom umiejętności poznawczych studentów. Szkoła Wyższa B może uzyskać w tym porównaniu lepsze wyniki, nawet jeżeli jej studenci na pierwszym roku mieli dużo gorsze wyniki niż studenci pierwszego roku Uczelni A. Narzędzia skonstruowane w taki sposób umożliwiają określanie wpływu dydaktyki prowadzonej w szkołach wyższych na rozwój (bądź zahamowanie) kompetencji akademickich studentów.

Uznaliśmy, że aby narzędzia pomiarowe dobrze spełniały swoje funkcje powinny mieć **charakter ogólny, tzn. niepowiązany z treściami nauczania**. W oparciu o doświadczenia amerykańskiego systemu edukacji (GRE) i badania polskich naukowców udało się zdefiniować grupę umiejętności, które wydają się być szczególnie istotne dla oceny akademickich dokonań studenta, a jednocześnie ich rozwój można traktować jako względnie niezależny od profilu uczelni. Dobrze skonstruowany test powinien odzwierciedlać poziom tych zdolności i wielkość wysiłku włożonego w ich rozwój. Projekt ten znajduje się nadal w fazie eksperymentalnych. Tym niemniej warto zauważyć, że część hipotez udało się już dziś potwierdzić. I tak, na przykład, studenci renomowanej uczelni warszawskiej uzyskali lepsze wyniki niż studenci prywatnej uczelni o tym samym profilu (mieszczącej się również w Warszawie); pojawiły się także istotne, choć w przypadku jednej uczelni odwrotne od oczekiwanych, różnice pomiędzy wynikami studentów 1 i 3 roku. Podstawowym problemem w interpretacji otrzymanych wyników jest **trudność określenia wpływu poziomu motywacji**. Wyniki uzyskane przez polskich studentów (w porównaniu ze studentami amerykańskimi) są bardzo niskie. Można przypuszczać, że byłyby dużo lepsze, gdyby studenci byli przekonani, że jest to egzamin bezpośrednio wpływający na ich ocenę lub taki, od którego w dużej mierze zależeć będą ich późniejsze losy zawodowe. Jest to najważniejszy problem, przed którym staniemy w drugiej fazie badań — wzbudzenie odpowiedniej motywacji u osób wypełniających test.

Druga, znacznie bardziej zaawansowana próba budowania i zastosowania testów osiągnięć jest już w praktyce realizowana jako element reformy systemu edukacji. Mowa tu o tzw. Teście Kompetencji Szkolnych, który powstał w ramach eksperymentu wałbrzyskiego. Jest to narzędzie uwzględniające materiał objęty podstawą programową nauczania języka polskiego i matematyki, a także ogólną orientację w innych, po-

krewnych dziedzinach — humanistycznych i przyrodniczych. Test ten, sprawdzany od kilku lat na terenie dawnego województwa wałbrzyskiego, okazał się być bardzo obiecujący. Obecnie kolejne jego wersje tworzy Ogólnopolski Zespół nauczycieli języka polskiego i matematyki powołany przez MEN. Trzeba podkreślić, że test osiągnięć musi być ciągle tworzony od nowa — dlatego musi mieć ogromną liczbę przetestowanych empirycznie pytań w bazie. Instytut Studiów Społecznych jako instytucja zaproszona przez MEN do diagnozy metodologicznej tego narzędzia stwierdził, że TKS jest efektywną metodą egzaminacyjną, którą można zastosować przy przekraczaniu pewnego progu edukacyjnego — chodzi o egzamin, który wszyscy uczniowie będą zdawali na zakończenie gimnazjum (pierwotne badania dotyczyły egzaminu kończącego szkołę podstawową i będącego kryterium przyjęcia do szkoły wyższej). Na stosowaniu takich egzaminów opiera się w dużej mierze realizacja podstawowych celów wprowadzanej obecnie reformy oświaty, tj. wyrównywanie szans edukacyjnych, optymalizacja wyboru odpowiedniej ścieżki kształcenia i wprowadzanie zewnętrznego systemu oceniania jakości kształcenia.

Przystępując do współpracy z MEN nad projektem Testu Kompetencji Szkolnych, ISS argumentował, że test ten ma być nie tylko narzędziem **selekcyjnym** (a więc kierującym uczniów na ścieżki kształcenia odpowiednie do ich umiejętności i zdolności), lecz również ma spełniać funkcję **informacyjną** i **motywującą**. Aby było to możliwe, wynik testu musi być znany uczniowi **przed** momentem podjęcia decyzji do jakiej szkoły składa papiery (funkcja informacyjna), a także ósmoklasista musi mieć **możliwość poprawienia** uzyskanego wyniku (funkcja motywacyjna). W przypadku tradycyjnego egzaminu do szkół średnich uczeń znajduje się w niekorzystnej sytuacji — egzamin wiąże się z silnym stresem, a ósmoklasista składa papiery do wybranej szkoły, nie wiedząc jakie ma szanse dostania się do niej — nie ma wcześniejszej informacji (niezależnej od opinii szkoły, do której uczęszcza) o swoich zdolnościach i możliwościach, pozwalającej na podjęcie rozważnej decyzji, na dodatkowe powtórzenie, uzupełnienie czy wręcz zdobycie wymaganej przez daną, wybraną szkołę, wiedzy. Sugerowaliśmy, że aby egzamin spełnił swą rolę motywacyjną, należałoby przeprowadzić jego pierwszą część (której wynik stanowiłby 75% punktów) np. w styczniu, a drugą część — z materiału ostatniego półrocza — w czerwcu (co stanowiłoby 25% punktów). Wynik z pierwszej części dałby informacje o typie szkoły, o której może myśleć dany ósmoklasista. Egzamin testowy może również dać szansę uczniom niedocenianym przez szkołę, wskazując im ich zdolności i możliwości dostania się np. do technikum, a nie tylko szkoły zawodowej i motywując ich tym samym do wytężonej pracy w ostatnim półroczu szkoły podstawowej. W ten sposób uczeń dowiedziałby się, że ma zdolności pozwalające na zdawanie do technikum czy liceum, a nie do szkoły zawodowej. Mógłby zdawać egzamin powtórnie, jeśliby sądził, że ma wystarczające umiejętności i może uzyskać lepsze wyniki. Część druga powinna wzmóc motywację do pracy w drugim półroczu, tak by uzyskać maksymalnie dużo punktów i rzeczywiście dostać się do szkoły średniej, pozwalającej na zdanie matury, a w przyszłości o ubieganie się o przyjęcie na studia wyższe.

Tymczasem, w trakcie wprowadzania reformy edukacji, wokół Testu Kompetencji Szkolnych narosło wiele nieporozumień. Z przykrością obserwujemy jak podstawowe cele zastosowania tego testu gdzieś się rozmywają. Gorąco popieraliśmy wprowadzenie testowych egzaminów zewnętrznych, a teraz zmuszeni jesteśmy przed nimi przestrzegać. Dlaczego?

Po pierwsze, trafność i rzetelność **testów osiągnięć jest przeceniana**. To prawda, że stopnie stawiane przez nauczycieli nie są wystandaryzowane. Nie oznacza to jednak, że nie należy przywiązywać do ich ocen wagi. Nawet najlepszy test nie sprawdzi systematyczności pracy czy też twórczego rozwiązywania problemów (trudno oczekiwać, że twórcze pomysły muszą się pojawiać w sytuacji egzaminacyjnej). Opracowywane standardy egzaminacyjne należy podzielić na dwie grupy: (1) te, które ma obowiązek sprawdzić nauczyciel uczący i co powinno być podstawą jego oceny; (2) te, które się dają RZETELNIE zmierzyć za pomocą testu składającego się z pytań zamkniętych. Lepiej zacząć skromnie — i nie tworzyć standardów idealistycznych, bo nie będą miały one żadnej mocy regulacyjnej. Trzeba zdawać sobie sprawę z tego, że np. wielu uczniów polskich szkół nie potrafi czytać ze zrozumieniem. Sądzimy, że lepiej ustalić standardy minimalne i uznać, że do selekcjonowania talentów są konkursy i olimpiady. Egzamin ma wymusić „podniesienie" mediany w najistotniejszych dla edukacji wymiarach. Wymagania muszą być możliwe do spełnienia w warunkach, w jakich żyjemy. Możemy je później podnosić. Przykładowo, wystandaryzowane pytania doświadczalne z fizyki można

rozesłać nauczycielom z poleceniem, że mają przetestować swoich uczniów i na tej podstawie m.in. wystawiać oceny. Uważamy, że powinniśmy w ten sposób okazać nauczycielom zaufanie, bo żadna reforma bez odbudowania statusu nauczyciela nic nie da.

Czy można spełnienie wszystkich standardów egzaminacyjnych zamknąć w jednym teście? Nie. Lepiej nie popełniać grzechu pychy i określić, które standardy powinny być sprawdzane przez test, a które przez nauczyciela. Pozwoliłoby to nauczycielowi na budowanie standardowej oceny opisowej.

Zewnętrzna informacja testowa nie może być jedynym kryterium selekcyjnym. Powinno się ją traktować jedynie jako **ważny element**. Mogłaby być czynnikiem motywującym, gdyby szkoły wyższego szczebla postawiły warunek, że aby być przyjętym, konieczne jest uzyskanie X punktów na egzaminie zewnętrznym. Ale pewnie powinien być to warunek konieczny, ale nie wystarczający. Przykładowo, w USA selekcja na uniwersytet, do pracy jest dokonywana na **podstawie wielowymiarowych danych**. **Wynik w teście** powinien być przedstawiany dwojako: jako **wartość bezwzględna i wartość pozycyjna w danej szkole**. Ta druga może nam powiedzieć, że uczeń mimo uzyskania przeciętnego wyniku należy do najlepszych w swojej szkole. Warto takim uczniom dać szansę dalszego kształcenia, bo ich słaby wynik może być interpretowany jako problem środowiskowy.

Drugim powszechnie pokutującym grzechem w myśleniu o testach osiągnięć jest **nieznajomość poprawnej procedury budowania testów**. Nieprawdą jest, że jesteśmy w stanie zbudować narzędzia, które będą rzetelne i trafne już w pierwszym roku stosowania. **Tworzenie testu osiągnięć to długi, kosztowny i nie pozbawiony błędów proces**. Amerykański ETS zatrudnia ponad 2000 pracowników i posiada w bazie danych dla jednego tylko testu (GRE) ponad 2 miliony pytań. Pojedyncze pytanie zanim zostanie włączone do właściwego testu egzaminacyjnego przechodzi procedurę weryfikacyjną trwającą prawie dwa lata. Nawet najlepszy ekspert nie może zagwarantować, że pytanie jest dobre bez jego przetestowania. **Własności psychometryczne pytania można ocenić jedynie na podstawie badań empirycznych.**

W Polsce powstało 8 okręgowych komisji egzaminacyjnych, z których każda dopuszcza możliwość używania własnych regionalnych testów. Oznacza to, że być może będziemy mieli w kraju 8 zupełnie innych testów na każdym poziomie edukacyjnym. Czemu ma służyć regionalizm ocen? Na pewno nie jest to dobry krok w kierunku budowy jednego, ogólnopolskiego narzędzia spełniającego wszystkie kryteria metodologiczne i pozwalającego porównywać osiągnięcia uczniów i szkół z różnych regionów Polski.

Po trzecie, warto zwrócić uwagę na **mnożenie niepotrzebnych kosztów**. Nie ma żadnej merytorycznej potrzeby, aby konstruowaniem tych samych narzędzi zajmowało się kilka niezależnych ośrodków. Czy Polskę stać na takie wydatki? Tworzenie bazy pytań testowych jest zajęciem bardzo kosztownym i powinno być **dziełem metodyków z całego kraju**, którzy przysyłają pytania do ośrodka tworzącego dany test. Każdy z okręgów może odpowiadać za gromadzenie bazy pytań z innego przedmiotu.

Czy stać nas na pytania otwarte? Koszt przeprowadzenia egzaminu z pytaniami otwartymi jest 10 razy wyższy niż z pytaniami zamkniętymi (dane z USA). Czy nie lepiej pieniądze wydawane na szkolenie kwalifikatorów przeznaczyć na podniesienie pensji nauczycieli, tak aby więcej czasu poświęcali na sprawdzanie prac domowych? Nawet najlepszy test nie może zastąpić kompetentnej oceny wypracowania czy rozwiązania złożonego problemu matematycznego. Ponadto, wprowadzenie do egzaminów zewnętrznych pytań otwartych **uniemożliwia automatyczne sprawdzanie prac** przez skaner i nie tylko zwiększa koszty, ale dodatkowo dramatycznie wydłuża czas oceny. Eliminuje to szansę na powtarzanie egzaminu.

I na koniec czwarta, może najważniejsza, uwaga. Egzamin testowy w obecnej formie **zamiast otwierać raczej zamyka szanse uczniów** słabszych lub gorzej znoszących stres egzaminacyjny. Przypomnijmy, że aby wyrównywać szanse, powinien on spełniać funkcję informacyjną i motywującą. Do tego konieczne są dwie rzeczy — wynik testu musi być znany uczniowi przed momentem podjęcia decyzji, do jakiej szkoły składa papiery (funkcja informacyjna), a także musi on mieć możliwość poprawienia uzyskanego wyniku (funkcja motywacyjna). W przeciwnym wypadku egzamin zewnętrzny staje się **dniem jedynej szansy**, kiedy to zdający uczeń wie, że koszt jego gorszej dyspozycji intelektualnej będzie ogromny. Z uporem maniaka powtarzaliśmy, że zdjęcie z egzaminu odium jednorazowości jest bardzo ważne. Każdy kto uważa, że wypadł poniżej swoich możliwości, powinien mieć szansę powtórzenia testu. Koszty powtarzania powinien ponosić

zdający, a w wypadku dzieci osób najgorzej sytuowanych materialnie, organy finansujące oświatę. Motywowanie do nauki jest przecież podstawowym zadaniem władz oświatowych. Zostało dobrze udokumentowane w badaniach, że stosunek nauki zależy od tego, czy uczeń traktuje zadanie jako informację o rozwoju (musi wtedy widzieć szansę poprawy) czy też jako ocenę jego osoby. Ta druga sytuacja rodzi niepotrzebny stres.