

FUNKCJE KRAJOWYCH EGZAMINÓW W SYSTEMIE EDUKACJI

Henryk Szaleniec

Roman Dolata

Instytut Badań Edukacyjnych w Warszawie

WPROWADZENIE

Sprawdzanie osiągnięć szkolnych może przybierać różną postać i pełnić wiele funkcji, jednak nie ma systemu oświaty, w którym w jakiejś formie nie występuje. Może być planowane i realizowane przez nauczyciela w klasie, może mieć ujednoczoną postać w szkole lub być realizowane z wykorzystaniem krajowych, standaryzowanych testów szkolnych. Wyniki mogą pełnić funkcję kształtującą lub pod postacią egzaminów decydować o losach uczniów już na wczesnych etapach edukacji.

Niektóre sprawdziany osiągnięć są powszechne, a ich wyniki służą diagnozie i ewaluacji pracy szkół, inne przeprowadzane są na losowo wybranych próbach uczniów, by monitorować stan wykształcenia populacji. Jednak współczesna, masowa szkoła nie potrafi obyć się bez oceniania. Mimo że niemal od stulecia system oświaty jest krytykowany za ocenianie, to poza nielicznymi szkołami eksperymentalnymi, dotychczas nie udało się zrealizować idei szkoły bez oceniania. Co więcej, ostatnie dekady to intensywny rozwój systemów oceniania i ukonstytuowanie się nowego poziomu ewaluacji – oceny międzynarodowej, której przedmiotem stają się całe systemy szkolne. Takie programy badawcze, jak TIMSS, PIRLS, ICCS, czy najbardziej znany PISA, dawno przestały być wyłącznie przedmiotem zainteresowania spragnionych wiedzy badaczy i stały się narzędziem globalnej ewaluacji.

Choć ocenianie osiągnięć szkolnych jest stałym elementem współczesnej szkoły, to krajowe systemy egzaminacyjne wykorzystujące standaryzowane testy nie są normą w systemach oświaty. W Polsce krajowy system egzaminów zewnętrznych stał się elementem pakietu czterech reform społecznych końca XX wieku: oświatowej, służby zdrowia, emerytalnej i administracyjnej. System egzaminacyjny miał się stać swoistym zwornikiem programowym polskiej oświaty, który w sytuacji pluralizmu programowego i decentralizacji zarządzania szkołami miał zapobiegać zbyt daleko idącemu różnicowaniu się polskiej oświaty. Równocześnie miał być forpcztą nowego spojrzenia na oświatę, w którym mechanizmy kontroli przez normowanie praktyki nauczania byłyby stopniowo zastępowane przez regulację za pomocą definiowania efektów kształcenia i kontroli ich osiągnięcia.

Zaczątkiem polskiego systemu egzaminacyjnego był prowadzony w latach 90. w kilku województwach eksperyment wałbrzyski. Polegał on na zastąpieniu egzaminów wstępnych do szkoły ponadpodstawowej standaryzowanym testem diagnostycznym, zdawanym na zakończenie ósmej klasy. W 1993 r. ukazał się raport *Studium wstępne krajowego systemu oceniania w polskim szkolnictwie ponadpodstawowym* opracowany przez J. Marquand z Uniwersytetu Sheffield, w którym przedstawiono niedostatki obowiązującego wówczas egzaminu dojrzałości i zaproponowano szereg działań prowadzących do zmiany formuły matury. W 1994 r. powołano krajowy program *Nowa Matura*, mający na celu przygotowanie nowych egzaminów maturalnych.

W kolejnych latach (1995–1998) w wyniku działania tego programu następowała stopniowa modyfikacja tematów egzaminacyjnych. Podejmowane były próby ujednoczenia kryteriów oceniania oraz trwały prace nad rozwiązaniami organizacyjnymi nowej matury. Był to także czas testowania nowej formuły egzaminu dojrzałości i szkolenia nauczycieli.

W 1998 r. został powołany krajowy program *SMART – komponent 02*, który miał na celu przygotowanie podwalin pod instytucjonalizację systemu oceniania zewnętrznego. Rok później została utworzona Centralna Komisja Egzaminacyjna w Warszawie, a wkrótce minister edukacji narodowej powołał dyrektorów ośmiu okręgowych komisji egzaminacyjnych. Działaniom tym towarzyszyły prace nad podstawą programową i standardami wymagań egzaminacyjnych oraz przygotowaniem koniecznych rozwiązań legislacyjnych. 19 IV 1999 r. ukazało się rozporządzenie w sprawie zasad oceniania,

klasyfikowania i promowania uczniów i słuchaczy oraz przeprowadzania egzaminów i sprawdzianów w szkołach publicznych.

Kolejne lata to rozwój CKE i okręgowych komisji egzaminacyjnych oraz przygotowywanie szkół i komisji egzaminacyjnych do przeprowadzenia pierwszych zewnętrznych egzaminów. W 2002 r. rozpoczęto wdrażanie zewnętrznych egzaminów w całej populacji uczniów kończących szkołę podstawową i gimnazjum.

Jak na ironię, nowa matura, która była osią reformy egzaminów zewnętrznych (opartej na wynikach programów *Nowa Matura* i *SMART*), odbyła się w 2002 r. tylko w nielicznych szkołach i tylko dla chętnych maturzystów. W wyniku decyzji politycznych tuż przed wyborami do sejmu w 2001 r. wprowadzenie zewnętrznego egzaminu maturalnego zostało przesunięte na 2005 r.

Dzisiaj, gdy egzaminy zewnętrzne w Polsce wchodzą w drugą dekadę funkcjonowania, warto zastanowić się nad ich istotą – do czego nam, państwu i społeczeństwu, potrzebne są krajowe egzaminy szkolne?

EGZAMINY SZKOLNE W KRAJACH UE

W większości krajów Unii Europejskiej (Eurydice 2010a) szkoła średnia II stopnia kończy się krajowym egzaminem, którego wysoki wynik stanowi także warunek konieczny dostania się na wyższą uczelnię. W niektórych krajach, podobnie jak w Polsce, organizuje się także egzaminy na zakończenie kolejnych etapów kształcenia lub na zakończenie okresu obowiązkowej nauki. Różnią się one w poszczególnych krajach liczbą przedmiotów egzaminacyjnych, konstrukcją testów, strukturą (istnieniem podziału na część wewnętrzną i zewnętrzną), sposobem przeprowadzania egzaminu (w szkołach lub w zewnętrznych placówkach), a nawet organizacją oceniania (przez szkolnych egzaminatorów lub zewnętrznych).

Egzaminy przeprowadzane na zakończenie etapów edukacji można podzielić również ze względu na znaczenie informacji o jego wyniku, czyli na egzaminy niskiej i wysokiej stawki. Jeżeli znaczenie wyniku egzaminowania jest istotniejsze niż znaczenie komentarza dydaktycznego odnośnie do tego wyniku, to egzamin decyduje o karierze zdających i nazywany jest egzaminem doniosłym (*high stakes*). Są nimi wszystkie egzaminy, których wyniki wykorzystywane są do rekrutacji do szkół wyższego szczebla. Egzaminy doniosłe na zakończenie edukacji obowiązkowej lub szkoły średniej I stopnia oraz na zakończenie szkoły średniej II stopnia są organizowane w Estonii, Danii, Holandii, Irlandii, Norwegii, Polsce, Portugalii, Słowenii, na Malcie i w Zjednoczonym Królestwie. Natomiast na Cyprze, w Grecji, Francji oraz na Litwie i Węgrzech egzaminy wysokiej stawki są przeprowadzane na zakończenie szkoły średniej II stopnia (Eurydice 2010b).

Na Słowacji uczniowie kończący naukę w szkołach podstawowych (klasa IX) piszą ogólnokrajowy test z języka ojczystego i matematyki. Wyniki testu zazwyczaj stanowią część kryteriów przyjęcia do szkoły średniej. Natomiast ukończenie szkoły średniej uwarunkowane jest między innymi zdaniem egzaminu maturalnego, który składa się z części wewnętrznej i zewnętrznej. Pisemna część egzaminu, jednolita dla całego kraju, przygotowuje Narodowy Instytut Certyfikowanych Pomiarów Kształcenia. Egzamin ten stanowi warunek konieczny ubiegania się o przyjęcie na studia wyższe.

W Holandii w ostatniej klasie szkoły podstawowej przeprowadzany jest test osiągnięć uczniów kończących szkołę, przygotowywany przez Narodowy Instytut Pomiarów Kształcenia – CITO. Z tego testu korzysta ponad 85% holenderskich szkół podstawowych. Warto podkreślić, że na podstawie wyników testu CITO oraz stopni szkolnych, zainteresowań i motywacji ucznia szkoły podstawowe doradzają rodzicom, który typ szkoły średniej będzie najbardziej odpowiedni dla ich dziecka.

Z kolei egzaminy końcowe w średnich szkołach ogólnokształcących składają się z dwóch części: egzaminu wewnętrznego (szkolnego) przygotowywanego i ocenianego przez nauczycieli danej szkoły i egzaminu państwowego przygotowywanego przez CITO, ale ocenianego przez szkolnych nauczycieli (ze szkoły ucznia i ze szkoły obcej), zgodnie ze standardami wyznaczonymi przez CITO, po ocenie próbki prac egzaminacyjnych. Z niektórych przedmiotów przeprowadza się tylko egzamin wewnętrzny.

W Austrii od 2005 r. prowadzone są prace przygotowujące wprowadzenie zewnętrznego systemu egzaminów. Pełne wdrożenie planowano na 2012 r. Także w Republice Czeskiej ma miejsce reforma, mająca na celu zbudowanie zewnętrznego systemu egzaminów. Pełne wdrożenie planowane na 2012 r. zostało przesunięte na 2013 r. Do 2011 r. wszystkie szkoły średnie II stopnia samodzielnie przeprowadzały maturalne egzaminy końcowe.

We wszystkich krajach UE na zakończenie nauki w szkole średniej absolwenci otrzymują świadectwa potwierdzające ukończenie tego etapu edukacji. Zwykle stanowią one wymóg formalny ubiegania się o indeks uczelni wyższej. Poza dwoma krajami (Hiszpania i Szwecja) uzyskanie certyfikatu kończącego edukację na poziomie szkoły średniej wiąże się z egzaminami – w większości przypadków zewnętrznymi lub mieszanymi. We Francji, Irlandii, na Malcie, w Rumunii i w Słowenii świadectwo obejmuje tylko wyniki egzaminów wewnętrznych. W niektórych krajach, podobnie jak w Polsce, wydawane są dwa świadectwa – ukończenia szkoły i potwierdzające zdanie egzaminów. Oprócz Polski takie rozwiązania stosowane są w Estonii, Finlandii, na Węgrzech i Słowacji.

W przeszłości państwa UE wyniki egzaminów doniosłych są wykorzystywane do oceny stanu krajowego systemu edukacji.

W dalszej części artykułu zajmiemy się tylko jednym z typów oceniania szkolnego – krajowym systemem egzaminacyjnym wykorzystującym jako główne narzędzie oceny standaryzowane testy osiągnięć szkolnych. Uwaga nasza skupi się na egzaminowaniu w zakresie kształcenia ogólnego. Na zakończenie krótko zajmiemy się najczęściej podnoszonym argumentem krytycznym wobec egzaminów – tzw. nauczaniem pod testy.

GŁÓWNE FUNKCJE SYSTEMU EGZAMINOWANIA

Krajowe systemy egzaminowania w zakresie kształcenia ogólnego mogą służyć realizacji różnych funkcji. Najważniejsze z nich to:

- 1) ocenianie stopnia opanowania przez uczniów wymagań programowych dla celów selekcyjnych;
- 2) przekazywanie w obrębie systemu oświaty profesjonalnej informacji o osiągnięciach szkolnych poszczególnych uczniów (diagnostyka osiągnięć);
- 3) ewaluacja, w tym autoewaluacja pracy szkół;
- 4) monitorowanie procesów edukacyjnych w skali kraju;
- 5) mobilizacja uczniów i dostarczanie informacji zwrotnej o efektach uczenia się.

Polskie egzaminy w różnym stopniu wypełniają te funkcje. Dokonanie nawet częściowego przeglądu tych funkcji daje możliwość zarysowania obszarów, w których warto i należy doskonalić system egzaminowania.

OCENIANIE STOPNIA SPEŁNIENIA PRZEZ UCZNIÓW WYMAGAŃ PROGRAMOWYCH DLA CELÓW SELEKCYJNYCH

We współczesnych demokracjach w myśleniu o dobrej edukacji akceptowana jest ideologia merytokratyczna. Wizja sprawiedliwego społeczeństwa nakłada na oświatę wymóg równości szans edukacyjnych, co prowadzi do idei jednolitej szkoły obowiązkowej. Ewentualna selekcja po każdym cyklu kształcenia powinna być oparta na merytokratycznym kryterium, czyli ocenie dotychczasowych osiągnięć szkolnych. Powinny one być obiektywnie i precyzyjnie wymierzone. Oczywiście związek merytokratycznej wizji ładu społecznego z wymogiem dokładnego pomiaru osiągnięć szkolnych nie oznacza, że egzaminy szkolne nie mogą być elementem systemów oświatowych w państwach od demokracji bardzo odległych. Egzamin szkolny może być jawnym lub tylko lekko skrywanym narzędziem społecznego eli-

taryzmu i ochrony grupowych przywilejów. Z drugiej strony państwa demokratyczne realizujące idee równości szans edukacyjnych mogą stawiać na inne sposoby oceny szkolnych osiągnięć, niż krajowe systemy egzaminacyjne, wykorzystujące standaryzowane testy osiągnięć.

Dla dobrego pomiaru osiągnięć uczniów na potrzeby selekcji kluczowy jest problem rzetelności i unormowania ilościowego wyniku testowego. Oczywiście pozostałe kryteria dobrego testowania też muszą być spełnione, ale rzetelność i porównywalność ma niewątpliwie znaczenie. Przyjmuje się, że testy egzaminacyjne służące selekcji powinny mieć bardzo wysoką rzetelność. Dla jej maksymalizacji kluczowe są trzy aspekty procesu budowy testu (arkusza egzaminacyjnego): dobre oszacowanie parametrów zadań testowych w badaniu pilotażowym, duża liczba i reprezentatywność programowa zadań w teście oraz wiarygodne zakodowanie zadań otwartych.

Na osobne podkreślenie zasługuje jedna z procedur tworzenia rzetelnych testów – sprawdzanie stroniczości zadań testowych ze względu na społeczne cechy ucznia, np.: płeć, status społeczny, przynależność do mniejszości czy miejsce zamieszkania. Do tej pory w polskim systemie egzaminacyjnym na etapie budowania arkusza w żaden sposób nie sprawdzano funkcjonowania zadań testowych w takich podgrupach. Jest to poważne zaniedbanie metodologiczne, którego skutki są istotne również z punktu widzenia polityki równych szans edukacyjnych.

Jeżeli w procedurach rekrutacyjnych do szkoły wyższego szczebla wykorzystuje się wyniki z różnych edycji danego egzaminu, nie mniej ważna od rzetelności jest ich porównywalność między latami. Do tego celu wykorzystuje się na świecie różne procedury zrównywania lub szerzej – wiązania (*linking*) wyników z różnych edycji egzaminu.

W zależności od typu testu, poza ogólną rzetelnością ważna może być też niepewność pomiaru w okolicy wartości kryterialnych. Gdy egzamin ma wyznaczone kryterium zaliczenia, test egzaminacyjny musi się charakteryzować szczególnie wysoką precyzją na tym właśnie poziomie. Gdy test ma sprawdzić spełnianie wymagań programowych na poziomie minimalnym, należy rozpatrzyć, czy nie warto zrezygnować z testu szerokiego spektrum na rzecz testu minimum kompetencji. Test taki składa się z dużej liczby zadań będących dobrymi wskaźnikami spełnienia kryterium i bardzo precyzyjnie dzieli populację na grupę spełniającą i niespełniającą wymagania minimum (takie testy można by zastosować na egzaminie maturalnym na poziomie podstawowym). W przypadku testów służących wyselekcjonowaniu grupy uczniów o najwyższych osiągnięciach, wskazane byłoby komponowanie arkusza z zadań trudnych i bardzo trudnych (takie testy mogą być przydatne na egzaminie maturalnym w wersji rozszerzonej).

W wypadku rekrutacji na studia wyższe warto poddać dyskusji możliwość rezygnacji z testu sprawdzającego spełnianie wymagań programowych na rzecz testu o maksymalnej wartości prognostycznej. Przegląd narzędzi stosowanych w innych krajach prowadzi do wniosku, że polskie rozwiązania maturalne, choć bliskie tradycji angielskiej i francuskiej, nie są jedyną metodą rekrutacji na studia.

PRZEKAZYWANIE W OBRĘBIE SYSTEMU OŚWIATY PROFESJONALNEJ INFORMACJI O OSIĄGNIĘCIACH SZKOLNYCH POSZCZEGÓLNYCH UCZNIÓW

Egzaminy mogą być swoistym pasem transmisyjnym informacji o uprzednich osiągnięciach szkolnych uczniów. Na podstawie informacji o wynikach egzaminów można podejmować decyzje o indywidualizacji programów nauczania czy działaniach wyrównawczych. Egzaminy mogą służyć tak rozumianej diagnostyce, ale trzeba wyraźnie powiedzieć, że cel ten może być znacznie lepiej i taniej realizowany za pomocą standaryzowanych, unormowanych testów diagnostycznych stosowanych przez nauczycieli lub personel psychologiczno-pedagogiczny szkoły. Dobremu opisowi szerokiego profilu osiągnięć szkolnych ucznia rozpoczynającego naukę w danej szkole (gimnazjum, szkole ponadgimnazjalnej), lepiej niż egzamin kończący szkołę niższego szczebla może służyć test w rodzaju *Stanford Achievement Test Series*¹. Niestety, tego typu testów w polskim systemie oświaty brakuje.

Diagnostyce osiągnięć szkolnych w polskim systemie egzaminacyjnym służyć ma sprawdzian przeprowadzany na zakończenie VI klasy

szkoły podstawowej. W jakim stopniu jego wyniki są wykorzystywane do wstępnej oceny osiągnięć ucznia na progu edukacji gimnazjalnej? W roku szkolnym 2009/10 działający przy Centralnej Komisji Egzaminacyjnej zespół zajmujący się rozwojem metody edukacyjnej wartości dodanej przeprowadził badanie wykorzystania przez gimnazja wyników sprawdzianu na zakończenie szkoły podstawowej. Badanie było prowadzone na ogólnopolskiej próbie szkół (150 gimnazjów; błąd standardowy oszacowań procentowych w wypadku próby nauczycieli wahał się od 0,5 do 1,5 pkt. proc.). Poniżej przedstawiono niektóre jego wyniki.

Na pytanie: *Czy w tym roku szkolnym w Państwa szkole analizowano wyniki sprawdzianu po szkole podstawowej uczniów rozpoczynających naukę w gimnazjum?* 80% dyrektorów gimnazjów odpowiedziało twierdząco. Dyrektorów, którzy stwierdzili, że w szkole analizuje się wyniki sprawdzianu, zapytano: *Jakie działania podjęto w szkole na podstawie tej analizy?* Odpowiedzi można podzielić na trzy grupy.

Pierwszy typ odpowiedzi wskazuje na grupowanie uczniów i działania wyrównawcze. Charakterystyczna dla tego typu wypowiedź dyrektora brzmiała: *Ponieważ wynik szkoły znalazł się poniżej średniej wojewódzkiej, należało wprowadzić efektywniejsze metody uczenia. Podzielono więc uczniów na zespoły według poziomu umiejętności i dostosowano do tych grup zajęcia wyrównawcze lub zajęcia poszerzające umiejętności.*

Drugi rodzaj odpowiedzi stanowiły zdawkowe, ogólnikowe i raczej mało realistyczne deklaracje typu: *Każdy nauczyciel uczący w klasie dostosowuje treści i metody pracy do potencjału uczniów* [na podstawie wyników sprawdzianu].

W trzeciej grupie odpowiedzi znalazły się takie, które wskazywały – wbrew deklaracjom – na niewykorzystanie wyników sprawdzianu. Poniżej przytoczone zostały trzy przykładowe odpowiedzi pochodzące z ankiet:

1. *Nauczyciele języka polskiego, matematyki przeanalizowali wyniki sprawdzianów. Zostały przygotowane testy startowe dla klas pierwszych z takich przedmiotów, jak: język polski, matematyka, historia, biologia, geografia.*

2. *Nauczyciele dokonują analizy na początku roku szkolnego, pracując w zespołach przedmiotowych. Przeprowadzają na pierwszych zajęciach z uczniami testy diagnostyczne.*

3. *Corocznie w naszym gimnazjum przeprowadzamy na początku września badanie umiejętności pierwszoklasistów z języka polskiego, matematyki, języka niemieckiego i historii. Następnie w zespołach przedmiotowych są one porównywane z wynikami sprawdzianów.*

Odpowiedzi ostatniego typu są spójne z niską oceną wartości diagnostycznej sprawdzianu dokonaną przez dyrektorów gimnazjów. Na pytanie: *Czy Pani/Pana zdaniem wyniki sprawdzianu są wiarygodną informacją o umiejętnościach szkolnych absolwentów szkół podstawowych trafiających do Pani/Pana szkoły?* odpowiedzi zdecydowanie tak udzieliło 5%, a raczej tak – 53% dyrektorów. Jeszcze bardziej krytyczni byli nauczyciele. Na analogiczne pytanie skierowane do nich zdecydowanie tak odpowiedziało 3%, a raczej tak 37% respondentów.

Z punktu widzenia diagnozy kluczowa jest ilość informacji, czyli jak najszerszy zakres sprawdzanych osiągnięć szkolnych i dostateczna precyzja pomiaru. Jak już wspomniano, zwykle do tego celu używa się baterii testów diagnostycznych. Im szerszy profil osiągnięć pozwala opisać dana bateria testów, tym lepiej. Oczywiście warunkiem użyteczności wszelkiej informacji jest jej rzetelność. Jeżeli będziemy opisywać umiejętność czytania na podstawie kilku pytań do jednego tekstu, narazimy odbiorcę na duże ryzyko błędu. Ale wymagania dotyczące wskaźników diagnostycznych mogą być mniej restrykcyjne, niż w wypadku wykorzystywania wyników do rekrutacji do szkoły wyższego szczebla. Dla diagnozy niezwykle ważne jest unormowanie testu w stosunku do wymagań programowych. Wyróżnienie poziomów wykonania testów opisanych w kategoriach wymagań programowych pozwala na właściwą interpretację dydaktyczną wyniku i ułatwia planowanie pracy z uczniem.

Należy wyraźnie powiedzieć, że sprawdzian szóstoklasistów w dotychczasowej formie jest z punktu widzenia diagnozy osiągnięć szkolnych słabym narzędziem pomiarowym. Przede wszystkim zawiera zbyt mało zadań, jest niewystarczająco zróżnicowany pod względem treści, a wynik nie jest unormowany ze względu na wymagania pro-

gramowe. Dodatkowo wskaźniki liczone dla poszczególnych obszarów standardów wymagań egzaminacyjnych mają niską trafność i rzetelność.

EWALUACJA I AUTOEWALUACJA PRACY SZKÓŁ

W krajach demokratycznych obserwuje się postęp decentralizacji oświaty. Rozszerzanie autonomii szkoły wynika z wielu przesłanek. Najważniejszy jest jednak fakt, że jesteśmy coraz bardziej świadomi, że nie można zwiększać efektywności edukacji przez odgórne zarządzanie. Szkoły na podstawie wiedzy profesjonalnej, rozpoznania lokalnych czynników efektywności powinny na bieżąco korygować swoje działania. W trosce o to, aby system szkolny nie stał się luźną korporacją szkół oraz aby zapewnić mu odpowiednie kryteria oceny efektywności, stosuje się dwa podstawowe instrumenty polityki oświatowej: krajową podstawę programową, na której bazują programy nauczania, i krajowe egzaminy.

Jakie kryteria muszą spełniać testy egzaminacyjne, by mogły dobrze służyć ewaluacji? Kluczowe wydają się wymagania analogiczne, jak w wypadku funkcji diagnostycznej, czyli szeroki zakres sprawdzanych celów nauczania oraz unormowanie testów z punktu widzenia wymagań programowych. Arkusze egzaminacyjne powinny być testami badającymi wystarczająco precyzyjnie szerokie spektrum umiejętności uczniów.

Ważne dla ewaluacji szkół jest zastosowanie do analizy danych egzaminacyjnych metody edukacyjnej wartości dodanej (OECD 2008). Dobre szacowanie EWD wymaga kilku pomiarów w trakcie ocenianego etapu edukacji; absolutne minimum to pomiar na wejściu i wyjściu. W Polsce od 2006 r. trwają prace nad zastosowaniem metody edukacyjnej wartości dodanej do analizy wyników egzaminacyjnych na potrzeby ewaluacji wewnątrzszkolnej i zewnętrznej. Wypracowano już wiele rozwiązań, a prace pilotażowe zaplanowane są do 2013 r.²

Dobrej ewaluacji służy też dostęp do wiarygodnej informacji o statusie społeczno-ekonomicznym społeczności lokalnej, w której funkcjonuje szkoła. Przykładem wykorzystania informacji o statusie społeczno-ekonomicznym społeczności lokalnej, w której funkcjonuje szkoła, jest ewaluacja szkół w Australii. Rezultaty prezentowane są na portalu *My School*³, prowadzonym przez ACARA (*Australian Curriculum, Assessment and Reporting Authority*)⁴. Portal umożliwia przeszukiwanie profili przeszło 9000 australijskich szkół i ma spełniać dwie główne funkcje. Po pierwsze, ma informować rodziców i uczniów o szkole, jej misji, personelu, zasobach, źródłach finansowania, charakterystyce uczniów i ich wynikach w długofalowych okresach. Po drugie, ma zapewnić szkołom i społecznościom lokalnym porównywalność wyników nauczania w zakresie czterech obszarów nauczania ze szkołami, do których uczęszczają uczniowie o podobnym profilu indywidualnym i społeczno-ekonomicznym. Porównania mają na celu dostarczenie informacji stymulujących działania w kierunku poprawy wyników nauczania. Natomiast zarówno nauczyciele, jak i uczniowie szkół o wysokich osiągnięciach i podobnych charakterystykach społeczno-ekonomicznych mogą być źródłem informacji na temat metod dydaktycznych, które dają korzystniejsze rezultaty.

W porównaniach szkół prezentowanych na portalu *My School* kluczowe jest uwzględnienie ich różnicowania pod względem pochodzenia społeczno-ekonomicznego uczniów, mierzonego indeksem ICSEA (*Index of Advantage*). Indeks ten bierze pod uwagę charakterystyki społeczne uczniów danej szkoły, a dokładnie – wykształcenie i zawód rodziców. Informacje te zbierane są w momencie zapisywania ucznia do szkoły oraz uzupełniane charakterystykami szacowanymi dla danej okolicy przez australijskie urzędy statystyczne.

MONITOROWANIE PROCESÓW EDUKACYJNYCH W SKALI KRAJU

Krajowy system egzaminów dostarcza niezwykle bogatego zbioru danych. Dzięki nim można monitorować skutki wprowadzanych reform oraz zmiany zachodzące spontanicznie w systemie. Przykładowe pytania to: *Czy poziom osiągnięć szkolnych rośnie?, Czy zmienia się zróżnicowanie osiągnięć?, Co się dzieje ze zróżnicowaniem międzyszkolnym?*

Jeżeli polskie egzaminy mają być dobrym narzędziem monitorowania, muszą być spełnione dwa warunki. Po pierwsze, systemowi edukacyjnemu należy zapewnić stabilność w czasie. Po drugie, należy na bieżąco prowadzić zrównywanie i kalibrowanie wyników, które zapewni wyeliminowanie losowych wahań trudności arkuszy egzaminacyjnych stosowanych w kolejnych sesjach (latach).

Dla funkcji monitorującej wskazane byłoby, by egzaminom zewnętrznym towarzyszyło (np. co 2 lata) krótkie badanie ankietowe (np. na reprezentatywnej próbie 5000 uczniów), dostarczające podstawowych informacji o zdających (m.in. wykształcenie i zawód rodziców, miejsce zamieszkania). Dzięki temu stosunkowo tanemu badaniu można by monitorować takie kluczowe dla polityki spójności społecznej parametry, jak zróżnicowanie osiągnięć szkolnych ze względu na sytuację społeczno-ekonomiczną rodziny ucznia, czy zależność wyników od miejsca zamieszkania.

Jeśli w systemie monitorowania oświaty chce się wykorzystywać wyniki egzaminów z kolejnych lat, to bardzo ważne jest, by wyniki testów egzaminacyjnych były porównywalne pomiędzy latami. Zrównywanie wyników testowania w celu zapewnienia ich porównywalności jest codzienną praktyką w wielu systemach zajmujących się testowaniem i egzaminami na dużych populacjach. Można tu wymienić m.in.:

- amerykańskie testy SAT (*Scholastic Assessment Test*) oraz ACT (*American College Testing*),
- izraelski PET (*Psychometric Entrance Test*),
- szwedzki SweSAT (*Swedish Scholastic Assessment Test*),
- kanadyjskie testy przygotowywane i administrowane przez EQAO (*Education Equality and Accountability Office*),
- australijski NAPLAN (*National Assessment Program – Literacy and Numeracy*).

W Japonii prowadzone są badania i próby zrównywania *post factum* wyników egzaminów z języka angielskiego IPET (*Ibaraki Prefecture English Test for High School Students*).

Monitorowanie procesów edukacyjnych można prowadzić z wykorzystaniem wyników egzaminacyjnych lub za pomocą badań sondażowych na losowych, krajowych próbach reprezentatywnych. Na wiele pytań można odpowiedzieć na podstawie badań reprezentatywnych typu amerykańskiego *The National Assessment of Educational Progress* (NAEP; *nces.ed.gov*). Współcześnie na znaczeniu zyskują także międzynarodowe badania osiągnięć szkolnych (PISA, PIRLS, TIMSS, ICCS). W bardzo dużym stopniu mogą one zastępować krajowe badania monitorujące.

W 2010 r. Instytut Badań Edukacyjnych podjął wieloetapowe studium, mające na celu zrównanie *post factum* wyników egzaminacyjnych kilku rodzajów egzaminów. Pierwszym egzaminem, dla którego w 2011 r. przeprowadzono zrównanie i wykalibrowanie wyników dla lat 2002–2011, był egzamin gimnazjalny w części humanistycznej i matematyczno-przyrodniczej⁵.

W 2012 r. rozpoczęto prace nad zrównaniem wyników sprawdzianu szóstoklasisty z lat 2002–2012. W kolejnych etapach przewidziano zrównanie wyników egzaminu maturalnego z części matematycznej dla lat 2010–2013 (2013) oraz z języka polskiego i języka angielskiego dla lat 2010–2014 (2014).

Aby zrównanie wyników egzaminu gimnazjalnego z lat 2002–2011 można było wykorzystać do monitorowania procesów edukacyjnych w skali kraju, a także do diagnozy szkolnej, w czwartym kwartale 2012 r. został uruchomiony internetowy serwis komunikujący rezultaty tych badań. Opracowana w IBE i zastosowana podczas badań metodologia zrównywania to zainicjowanie działań, które być może zaowocują przygotowaniem projektu wdrożenia w systemie polskich krajowych egzaminów systematycznego zrównywania wyników na etapie tworzenia narzędzi i przeprowadzania egzaminu.

WPLYW EGZAMINÓW NA NAUCZANIE

Po omówieniu wybranych funkcji krajowego systemu egzaminacyjnego przyjrzymy się skutkom ubocznym stosowania testów. Najczęściej przytaczane w dyskusji publicznej i literaturze naukowej jest tzw. nauczanie pod testy. Nie ulega wątpliwości, że nawet najbardziej rzetelne i trafne testy egzaminacyjne same z siebie nie mogą przyczynić się do rozwoju krytycznego i twórczego myślenia czy innych złożonych umiejętności bez jasno określonych strategii nauczania. Dla

tego też nadzieje na reformę systemu edukacji, opartą jedynie na zmianie typu egzaminu, okazały się w wielu krajach pienne (Goldstein, Lewis 1996). Egzaminy odciskają swoje piętno na procesie nauczania. Funkcjonuje przekonanie poparte wynikami wielu badań, że egzaminy zewnętrzne stymulują uczenie pod dany egzamin. Takie nauczanie koncentruje się na umiejętnościach łatwych do zweryfikowania, i co do których istnieje duże prawdopodobieństwo, że będą sprawdzone podczas zewnętrznych egzaminów.

N. Fredriksen już przeszło 35 lat temu podkreślał, że jest wysoce prawdopodobne, że egzamin zewnętrzny ma wpływ na nauczanie i uczenie się, jeżeli nauczycielom i uczniom będą znane przykłady arkuszy egzaminacyjnych podobnych do tych, które stosowane są na egzaminach (Fredriksen 1984). Nic dziwnego, że nauczyciele w procesie dydaktycznym często sięgają po arkusze egzaminacyjne z poprzednich lat. Zarówno uczniowie, jak i nauczyciele chcą być przecież skuteczni. Z drugiej strony, ponieważ na nauczanie danego przedmiotu przeznaczona jest stała liczba godzin lekcyjnych, istnieje duże prawdopodobieństwo, że nauczyciele poświęcą więcej czasu na nauczanie umiejętności zazwyczaj sprawdzanych na egzaminie końcowym, zaniedbując kształtowanie tych, które nie są na nim uwzględniane.

Ponieważ nie wszystkie umiejętności jednakowo łatwo poddają się sprawdzaniu zewnętrznemu, to głównie sprawdzane i nauczane mogą być te z nich, które są łatwe i ekonomiczne w testowaniu na egzaminach zewnętrznych. W praktyce tanie testy wyboru wypierają egzaminy bardziej pracochłonne. Wiele ważnych umiejętności nie sprawdza się na krajowych egzaminach i w efekcie także się ich nie naucza (Fredriksen 1984).

G. Madaus argumentuje, że w krajach, w których egzaminy krajowe są organizowane od bardzo wielu lat, profesjonalizm nauczycieli jest mierzony rezultatami uczniów uzyskanymi na egzaminach zewnętrznych. W związku z tym nauczyciele ćwiczą z uczniami te umiejętności, które najprawdopodobniej będą przedmiotem pomiaru w trakcie najbliższej sesji (Madaus 1988). Zatem w perspektywie efektów długofalowych egzaminy zewnętrzne mogą mieć ogromny wpływ na proces nauczania, *de facto* nie tyle na planowany, ile na realizowany program nauczania.

L. Shepard (1992) na podstawie swoich badań doszedł do wniosku, że jeżeli egzaminom nadaje się wysoką rangę w wyniku nacisków politycznych, dużego zainteresowania mediów tymi egzaminami, to wyniki zwykle są wyższe niż rzeczywiste osiągnięcia i dostarczają fałszywych informacji o opanowaniu umiejętnościach. W Polsce mamy niewiele badań na ten temat. Jednym z nielicznych jest studium K. Konarzewskiego (2008). Już sam podtytuł – *Pokusa łatwego zysku* – przedstawia podstawą tezę autora.

Przed instytucjami odpowiedzialnymi za przeprowadzanie krajowych egzaminów oraz autorami zadań egzaminacyjnych stoi bardzo ważne wyzwanie tworzenia takich narzędzi, które pozwoliłyby sprawdzić opanowanie kluczowych umiejętności ujętych w programie nauczania, ale na podstawie zmiennej i szerokiego spektrum treści. Aby egzaminy zewnętrzne nie prowadziły do zawężenia programu nauczania do treści sprawdzanych w poprzednich sesjach, nauczyciel musi wiedzieć, że weryfikują one jedynie część wymagań programowych i że rozsądniej jest planować nauczanie zgodne z programem nauczania, a nie z tym, co znalazło się na poprzednich egzaminach.

PODSUMOWANIE

Żaden system egzaminacyjny nie jest doskonały i każdy ma swoje specyficzne problemy. Realizacja założonych funkcji przez krajowy system egzaminacyjny nie może być rozpatrywana w kategoriach dychotomicznych, a raczej jako kontinuum określające stopień ich realizacji. Wprowadzenie standaryzowanych, powszechnych egzaminów zewnętrznych było krokiem umożliwiającym rozpoczęcie realizacji wszystkich wymienionych wyżej funkcji, choć sednem systemu egzaminacyjnego jest racjonalizacja procesów selekcyjnych na progach szkolnych oraz ewaluacja szkół.

Diagnoza osiągnięć szkolnych i monitorowanie procesów edukacyjnych mogą być realizowane innymi metodami, choć warto, by krajowy system egzaminów dobrze realizował także i te cele.

System egzaminacyjny na początku swojego funkcjonowania wydawał się nowatorski i dobrze dopasowany do polskich potrzeb. Kolej-

ne jego reformy miały charakter raczej kosmetyczny i nie zmieniały fundamentów, które były tworzone pod koniec ubiegłego wieku. Aktualnie najważniejsze problemy systemu egzaminacyjnego, które trzeba rozwiązać, to:

- słabość procedur i standardów konstrukcji testów egzaminacyjnych. Trzeba wdrożyć procedury zgodne ze współczesnym stanem wiedzy w zakresie testowania osiągnięć szkolnych. Szczególnie dotkliwy jest brak banków zadań, sprawnie działających procedur standaryzacyjnych, procedur rzetelnego kodowania zadań otwartych, procedur zrównywania i skalowania wyników egzaminacyjnych już w trakcie budowy testów;

- zbyt mało wyrazista polityka w zakresie ewaluacyjnej funkcji egzaminów;

- brak polityki budowania zaufania wobec systemu egzaminacyjnego i związana z tym słaba legitymizacja systemu egzaminacyjnego.

Ważnym ograniczeniem rozwoju egzaminów jest słabość zewnętrznych, wobec systemu egzaminacyjnego, mechanizmów kontroli jakości egzaminów pod względem psychometrycznym i dydaktycznym. Przyczyną jest słabość polskiego środowiska naukowego zajmującego się szkolnym ocenianiem i związane z tym mizerne zaplecze badawcze mogące monitorować i planować zmiany.

¹ *Stanford Achievement Test Series* jest stosowany w USA i szkołach amerykańskich na całym świecie. Dla każdej klasy przygotowywane są testy mierzące umiejętności w takich obszarach, jak: czytanie ze zrozumieniem, rozwiązywanie problemów matematycznych, język ojczysty, ortografia, słuchanie ze zrozumieniem, nauki przyrodnicze i społeczne.

² Patrz: www.ewd.edu.pl

³ Patrz: www.myschool.edu.au

⁴ ACARA to niezależna australijska instytucja odpowiedzialna za opracowanie krajowych programów nauczania i krajowych programów oceniania z zastosowaniem standaryzowanych testów. Gromadzi i analizuje dane oraz raportuje wyniki testowania. ACARA w ramach programu NAPLAN (*National Assessment Program Literacy and Numeracy*) przeprowadza standaryzowane testy czytania i umiejętności matematycznych dla całej populacji, a także w trzyletnim cyklu badania na reprezentatywnej próbie uczniów z zastosowaniem krajowych standaryzowanych testów (NAP – *National Assessment Program*) w zakresie umiejętności przyrodniczych, wiedzy obywatelskiej oraz umiejętności TIK.

⁵ Wyniki badań są prezentowane w artykule Szalaniec i in. (2012).

LITERATURA

- Eurydice (2010a), *Egzaminy ogólnokrajowe dla uczniów w Europie: cele, organizacja oraz wykorzystanie*, Warszawa: Fundacja Rozwoju Systemu Edukacji.
- Eurydice (2010b), *Kluczowe dane o edukacji w Europie 2009*, Warszawa: Fundacja Rozwoju Systemu Edukacji.
- Eurydice (2010c), *Systemy edukacji w Europie – stan obecny i planowane reformy*. Słowacja, Warszawa: Fundacja Rozwoju Systemu Edukacji.
- Fredriksen N. (1984), *The real test bias: influence of testing, on teaching and learning*, „American Psychologist” 39(3).
- Goldstein H., Lewis T., red. (1996), *Assessment, problems, developments and statistical issues*, New York: John Wiley and Sons.
- Konarzewski K. (2008), *Przygotowanie uczniów do egzaminu: pokusa łatwego zysku*, Warszawa: Instytut Spraw Publicznych.
- Madaus G. (1988), *The influence of testing on the curriculum*, w: Tanner (red.), *Critical issues in curriculum*, (87th Yearbook of NSSE Part 1), Chicago, IL: University of Chicago Press.
- OECD (2008), *Measuring improvements in learning outcomes: best practices to assess the value-added of schools*, Paris.
- Shepard L. (1992), *Will national tests improve student learning?*, [CSE, Technical Report CRESST], Boulder: University of Colorado.
- Szalaniec H., Grudniewska M., Kondrątek B., Kulon F., Pokropek A. (2012), *Wyniki egzaminu gimnazjalnego 2002–2010 na wspólnej skali*, „Edukacja” nr 119(3).

SUMMARY

In addition to inter-school student assessment both informative and cumulative, national examinations also function in many educational systems, whose intensive development is observed since the early 21st C. Development is also observed in extensive international school achievement tests, such as PISA, PIRLS, TIMSS or ICSS, where entire school systems are compared and evaluated. These tests may support or even replace in some areas the functions fulfilled by national exams. Among the many goals of national examination systems, it is particularly worth pointing out the following: evaluation the degree to which pupils meet program requirements (used in the recruitment to higher level schools), mobilization of pupils and supplying feedback on the results of learning, diagnosis of achievements, evaluation, in this self-evaluation of the work of schools, as well as monitoring educational processes on the national scale. The article presents the above-mentioned functions in the context of conditions Polish exams should comply with in order to achieve these goals.