

dr Henryk Szaleniec

Instytut Badań Edukacyjnych w Warszawie

Czy egzaminy zewnętrzne wpływają na wewnątrzszkolne ocenianie i politykę edukacyjną?

Wstęp

Funkcjonuje obiegowa opinia na temat obniżania się umiejętności (wymaganych zgodnie z podstawą programową) kolejnych populacji kończących edukację szkolną od czasu reformy edukacji w 1999 roku. Nierzadko można się też spotkać się ze stwierdzeniami, że to wprowadzenie gimnazjum i egzaminów zewnętrznych ma ujemny wpływ na proces nauczania i spowodowało zubożenie wewnątrzszkolnego oceniania. Także w krajach, w których egzaminy zewnętrzne funkcjonują o wiele dłużej niż w Polsce, obserwuje się narzekania nauczycieli na negatywny wpływ testów standaryzowanych stosowanych na egzaminach na proces nauczania. Jako przyczynę tego zjawiska podaje się efekt zwrotny polegający na wpływie egzaminów na codzienną praktykę szkolną. Bolesław Niemierko w artykule zatytułowanym *Czy egzaminy zakłócają pracę nauczyciela* (Niemierko, 2008) wymienia aż 10 zarzutów, wśród których na drugim miejscu jest zaniedbywanie przez nauczycieli w procesie dydaktycznym materiału pozaegzaminacyjnego.

W idealnym przypadku egzaminy zewnętrzne mogą mieć jednak wielce pozytywny wpływ na to, co się dzieje w klasie zarówno w zakresie nauczania, jak i oceniania. Każdego roku nauczyciele pozyskują informację zarówno o mocnych, jak i słabych aspektach wyników, z którymi ich uczniowie opuścili szkołę. Informację, która może być w refleksyjny sposób wykorzystana do doskonalenia środowiska uczenia się w klasie, w tym oceniania kształtującego. W praktyce często jest jednak inaczej, o czym świadczy wiele opracowań na ten temat (Pellegrino i in., 2001). Sharped (Sharped L.A., 2006) zwraca uwagę na dwa czynniki, które powodują zawężenie programu nauczania do umiejętności będących przedmiotem egzaminów zewnętrznych:

1. zakres, w jakim egzaminy zewnętrzne pokrywają się z celami kształcenia (w naszym wypadku z całym spektrum wymagań określonych w podstawie programowej);
2. poziom umiejętności nauczycieli do zogniskowania procesu dydaktycznego na rzeczywistym uczeniu się w szkolnej klasie, a więc także umiejętności oceniania kształtującego.

Po pierwsze, te dwa czynniki określają jeden z ważnych celów kształcenia i doskonalenia nauczycieli:

- nabycie umiejętności rozkładania akcentów w procesie dydaktycznym tak, aby stwarzać okazje do skutecznego opanowania umiejętności, które specyfikuje podstawa programowa, a z pewnością nie pojawią się na egzaminach zewnętrznych;

- opanowanie umiejętności pełniejszego niż dotychczas wykorzystywania możliwości oceniania kształtującego do kierowania procesem uczenia się.

Po drugie, wyznaczają strategię doskonalenia egzaminów zewnętrznych i krajowych badań z wykorzystaniem standaryzowanych testów. Ocenianie wewnątrzszkolne, a szczególnie ocenianie kształtujące i ocenianie poprzez egzaminy zewnętrzne służą odrębnym celom, które się dopełniają, a nie wykluczają. Ocenianie wewnątrzszkolne wspomaga uczenie się, natomiast ocenianie na egzaminach dokumentuje poziom osiągnięć na progach edukacyjnych i to tylko w zakresie, który poddaje się zewnętrznemu ocenianiu technikami stosowanymi na egzaminie (egzamin pisemny, egzamin ustny).

Rezultaty porównawczych badań międzynarodowych, takich jak PISA, PIRLS, TEDS-M¹, a także egzaminy zewnętrzne dostarczają informacji przydatnych dla osób odpowiedzialnych za politykę oświatową w skali przekraczającej szkolną klasę, konkretną szkołę czy uczelnię. Przy czym, o ile badania międzynarodowe pozwalają na przedstawienie rezultatów naszych uczniów i studentów na tle osiągnięć ich rówieśników w wielu krajach, to dodatkowo umożliwiają na śledzenie zmian w zakresie poszczególnych umiejętności w szerokim horyzoncie czasowym. Niestety, jak dotychczas wyniki egzaminacyjne bez dodatkowych zabiegów nie nadają się do śledzenia osiągnięć uczniowskich w przedziałach czasowych wykraczających poza jeden rok. Dzieje się tak dlatego, że nasze egzaminy już na samym początku nie zostały zaprojektowane tak, by umożliwić porównanie wyników z roku na rok. Wszystkie zadania zastosowane na egzaminie w danym roku są oryginalnie nowe i po egzaminie publikowane. Nie ma więc możliwości zastosowania próbki tych samych zadań w kolejnym roku jako zadania kotwiczące dla całej populacji. Ponieważ są to egzaminy doniosłe, nie praktykuje się też zastosowania zadań planowanych na przyszły rok na próbie zdających egzamin w danym roku. Wróćmy na chwilę do genezy polskiego systemu egzaminów zewnętrznych.

W jakim zakresie system egzaminów zewnętrznych rozminął się z założeniami przyjętymi na etapie projektowania?

Zanim powstał obecny system egzaminów zewnętrznych, w 1993 roku Judith Marquand (Marquand J., 1993) w *Studium wstępnym krajowego systemu oceniania w polskim szkolnictwie ponadpodstawowym* zwróciła uwagę, że system oceniania stanowi kluczową część gwarancji jakości zapewniającej prawidłowe funkcjonowanie systemu zarządzania edukacją oraz kryterium, które powinno pozwolić na badanie, czy uzyskane wyniki odpowiadają zamierzonym celom. Mówiąc *system oceniania* mamy na myśli zarówno ocenianie wewnątrzszkolne (kształtujące i sumujące), jak i egzaminy zewnętrzne. Jeżeli system oceniania ma stymulować wszechstronny rozwój uczniów, to jak píše Judith Marquand:

¹ Międzynarodowe badania porównawcze prowadzone przez IEA na temat umiejętności matematycznych i umiejętności dydaktycznych (w tym oceniania) studentów przyszłych nauczycieli matematyki w szkole podstawowej

Można to osiągnąć przez:

- *przekazywanie czytelnych sygnałów, iż wszechstronny rozwój, zgodny z wysokimi standardami jest ceniony;*
- *stymulowanie systemu, w którym uczący się jest centralnym punktem sygnalizując poprzez to, co i jak jest oceniane, że indywidualny (i zindywidualizowany) rozwój rozumienia i umiejętności jest rzeczą ważną;*
- *informację zwrotną dla uczniów o ich postępach, tak by zrozumieli, co muszą zrobić, aby osiągnąć lepsze wyniki;*
- *rozwój samooceny;*
- *dostarczenie na właściwych etapach nauki (...) stosownej informacji [dla ucznia i - przyp. aut.] zainteresowanym z zewnątrz².*

Właściwie tylko ostatni punkt jednoznacznie odnosi się do egzaminów na progach edukacyjnych. W 1993 roku obowiązujący wówczas system oceniania na egzaminach na zakończenie szkolnej edukacji (egzamin dojrzałości) cechowała słabość w zakresie czterech obszarów:

1. ustalania kryteriów oceny,
2. nadania uprawnień ocenającym,
3. zapewnienia porównywalności kryteriów oceniania,
4. oceniania rezultatów i zapewnienia informacji zwrotnej zarówno w skali lokalnej, jak i regionalnej oraz w skali całego kraju³.

System egzaminów zewnętrznych, który powstał 4 lata później niż prognozowała Judith Marquand, miał wyeliminować te braki. Rzeczywiście egzaminy zewnętrzne, które na poziomie gimnazjalnym rozpoczęły się w 2002 roku, a w przypadku matury skutecznie dopiero w 2005 roku, nie we wszystkich aspektach sprostały oczekiwaniom zawartym w wymienionych powyżej czterech punktach. W dzisiejszej praktyce egzaminacyjnej rozwiązania zadań egzaminacyjnych punktowane są według ustalonych na każdą sesję kryteriów. W sesji egzaminacyjnej do tych prac zatrudniani są tylko przeszkoleni egzaminatorzy, którzy uzyskali uprawnienia do oceniania. Ale o ile możemy być w znacznym stopniu zadowoleni ze stopnia porównywalności kryteriów oceniania w danej sesji w skali całego kraju, to już z taką samą pewnością nie możemy twierdzić, że zarówno kryteria, jak i narzędzia egzaminacyjne są w zadawalającym stopniu porównywalne pomiędzy latami, a nawet sesjami egzaminacyjnymi w danym roku (sesja główna i np. poprawkowa lub w terminie dodatkowym).

Prof. Konarzewski w tekście dla „Gazety Szkolnej” zatytułowanym *Egzaminy zewnętrzne nie alarmują, a powinny* pisze, że twórcy egzaminów zewnętrznych obiecywali, że system egzaminacyjny pozwoli śledzić zmiany zachodzące z roku na rok w wykształceniu zarówno poszczególnych uczniów, jak i całych roczników. Obietnica ta nie znalazła jednak pokrycia w praktyce egzaminacyjnej. Konarzewski stwierdza, że: *Stało się tak za sprawą procedury zwanej standaryzacją testu [stosowaną przez CKE – przyp. aut.]. Wyniki testowania uczniów zależą od tego, jak wiele trudnych zadań znajdzie się w teście. (...) Jest jasne, że standaryzacja*

² Marquand Judith, (1993) Studium wstępne krajowego systemu oceniania w polskim szkolnictwie ponadpodstawowym, Departament Kształcenia Ogólnego, Departament Współpracy z Zagranicą MEN.

³ Ibid. str. III.

utrzymująca średnią i inne parametry testu na stałym poziomie przekreśla nadzieje na ujawnianie zmian w wykształceniu uczniów. Z podobieństwa średnich wyników w 2002 r. i 2008 r. nie można wnioskować, że młodzież z obu tych roczników miała wykształcenie podobnej jakości. Z różnic średnich w 2003 r. i 2004 r. nie wynika, że załamało się kształcenie humanistyczne w kraju⁴.

Jest to więc obszar, w którym konieczne są skuteczne działania. Działania, które pozwolą nie tylko na bieżąco śledzić zachodzące z roku na rok zmiany w wykształceniu całych roczników, ale także umożliwią opisanie ich z wykorzystaniem wyników egzaminu od początku funkcjonowania w Polsce systemu egzaminów zewnętrznych. Podejmując takie działania, warto jednak mieć na uwadze zachowanie równowagi pomiędzy patrzeniem na szkołę przez pryzmat efektów uwidocznionych w wynikach egzaminacyjnych i oglądem bliższym, który bierze pod uwagę działania bezpośrednio sprzyjające rozwojowi uczenia się. Zresztą już cytowanym *Studium* Judith Marquand podkreśla: *Z rozwojem porównywalności w systemie oceniania związane są trudności praktyczne. Niezbędnym jest znalezienie równowagi pomiędzy całkowitą porównywalnością z jednej strony oraz popieraniem indywidualnej odpowiedzialności za naukę i nauczanie z drugiej strony. To, co jest oceniane oraz sposób, w jaki się dokonuje oceny, może mieć ogromny wpływ na zakres i sposób nauczania, a tym samym na zakres i sposób przyswajania wiedzy. Metody oceniania uczniów są więc niezmiernie ważnymi instrumentami wpływu na to, co się dzieje w klasie i w całym systemie*⁵.

Jaki wpływ na to, co się dzieje w klasie i w całym systemie edukacyjnym, mają przeprowadzane od 2002 roku egzaminy zewnętrzne? Jaki wpływ na ocenianie wewnętrzzszkolne, a szczególnie na ocenianie kształtujące, mają egzaminy zewnętrzne? Jaki wpływ na poziom umiejętności kolejnych roczników absolwentów opuszczających szkołę mają egzaminy zewnętrzne?

Aby odpowiedzieć na pierwsze dwa pytania, konieczne są badania. To duże pole do popisu dla doktorantów kierunków pedagogicznych. Poszukiwanie odpowiedzi na trzecie pytanie kieruje nas do danych gromadzonych przez system egzaminów zewnętrznych.

Pomimo że system egzaminów zewnętrznych zgromadził obszerną bazę wyników egzaminacyjnych kolejnych populacji, zdających egzaminy począwszy od 2002 roku, to ich analiza nie umożliwia odpowiedzi na tak postawione pytanie.

Konieczne jest zrównanie wyników, czyli określenie wyników równoważnych uzyskanych za pomocą arkuszy egzaminacyjnych zastosowanych w kolejnych latach (sesjach egzaminacyjnych) i odniesienie ich do tzw. roku bazowego, który może być arbitralnie ustalony oddzielnie dla każdego egzaminu. Innymi słowy, zrównywanie wyników prowadzi to takiego *wyjustowania* różnych narzędzi pomiaru, które pozwala założyć, że cechuje je ten sam poziom trudności.

⁴ K. Konarzewski, *Egzaminy zewnętrzne nie alarmują, a powinny*, „Gazeta Szkolna”, 2008 nr 17-18, s. 1, 19.

⁵ Ibid.

Zrównanie wyników (*equating*) uzyskujemy poprzez kotwiczenie (*anchoring*) arkuszy egzaminacyjnych przez zewnętrzne lub wewnętrzne dla danego testu zadania, grupy zadań lub całe testy.

W Polsce pierwsza próba porównania rezultatów sprawdzianu z kolejnych sesji egzaminacyjnych z zastosowaniem zrównywania wyników (*equating*) z wykorzystaniem do kalibrowania testki kotwiczącej wiążącej narzędzia stosowane na egzaminie w kolejnych latach (*linking*) (2003-2004) została przeprowadzona metodą ekwicyntylową (Niemierko, 2004) w ramach badań koordynowanych przez CKE. Z zastosowaniem takiego samego planu gromadzenia danych (model zrównywania) zostało przeprowadzone również zrównanie wyników sprawdzianu 2004 i 2005 do rezultatów z roku 2003, przy wykorzystaniu dwóch metod statystycznych bazujących na teorii wyniku zadania (IRT) z wykorzystaniem modelu i programu OPLM (*One Parametr Logistic Model*) (Szaleniec H., 2009) oraz modelu Rasha i programu Winsteps (Smolik M., 2009). Próby te dostarczyły wstępnych doświadczeń w dziedzinie zrównywania, ale nie zapewniły jednak rozwiązań do systematycznego zastosowania względem wszystkich egzaminów zewnętrznych. Słabym punktem tamtych badań był brak analiz symulacyjnych, które pozwoliłyby wybrać odpowiednio takie rozwiązania, które zapewniłyby najmniejszy błąd standardowy zrównywania. Kontynuacją prac w tym zakresie jest projekt realizowany przez Instytut Badań Edukacyjnych.

Pierwszym etapem tego projektu, który kończy się w pierwszym kwartale 2011 roku, jest dopracowanie koncepcji metodologicznej badań umożliwiających długookresową porównywalność wyników egzaminów zewnętrznych z wykorzystaniem najnowszych rozwiązań na świecie w tej dziedzinie. W tym samym roku w marcu zostaną przeprowadzone badania na reprezentatywnej próbie gimnazjów, mające na celu dostarczenie danych koniecznych do skali-browania wyników egzaminu gimnazjalnego z 10 lat (2002-2011) w ten sposób, aby mogły być one traktowane jako wyniki równoważne – funkcjonalnie równorzędne.

Dlaczego gimnazjum zostało wybrane jako pierwszy egzamin do zrównywania wyników?

W 2002 roku do pierwszych egzaminów zewnętrznych przystąpili pierwsi uczniowie kończący edukację w gimnazjum i zreformowaną szkołę podstawową. Przez kolejne lata sprawdzian na zakończenie szkoły podstawowej i egzamin gimnazjalny prowadzone były zgodnie z wymaganiami określonymi przez Podstawę Programową wprowadzoną w 1999 roku⁶ i informatory do każdego egzaminu. W 2012 roku odbędzie się pierwszy egzamin gimnazjalny uczniów, których edukacja oparta jest na nowej Podstawie Programowej⁷. Rok 2011 jest ostatnim rokiem, który pozwala przeprowadzić zakotwiczenie rezultatów egzaminów gimnazjalnych w minionych latach bez konieczności wprowadzania

⁶ Rozporządzenie Ministra Edukacji Narodowej z dnia 30 czerwca 1999 r. zmieniające rozporządzenie w sprawie podstawy programowej kształcenia ogólnego. (Dz. U. z dnia 12 lipca 1999 r.).

⁷ Rozporządzenie Ministra Edukacji Narodowej z dnia 23 grudnia 2008 r. w sprawie podstawy programowej wychowania przedszkolnego oraz kształcenia ogólnego w poszczególnych typach szkół (Dz. U. z dnia 15 stycznia 2009 r. Nr 4, poz. 17).

poprawki ze względu na zmianę podstawy programowej i zmiany w strukturze egzaminu. Jednocześnie badania te dostarczą danych do zakotwiczenia egzaminów 2002-2011 z egzaminem 2012.

Pierwszy zmieniony Sprawdzian na zakończenie szkoły podstawowej, bazujący na nowej podstawie programowej, odbędzie się dopiero w 2015 roku, podobnie jak egzamin maturalny.

W 2011 roku będziemy dysponować wynikami egzaminów na zakończenie szkoły podstawowej i gimnazjum z 10 kolejnych lat (2002-2011). To dostatecznie długi okres, w którym mogą być już obserwowalne zmiany wywołane reformami wprowadzonymi w edukacji pod koniec minionego wieku. Musimy także uwzględnić jeszcze jeden warunek – stabilność w czasie formuły poszczególnych egzaminów, szczególnie, że po dziewięciu latach ich funkcjonowania (2002-2010) w związku z wejściem nowej podstawy programowej wprowadzane są istotne zmiany w tym zakresie.

Pozostając przy egzaminie gimnazjalnym, warto podkreślić, że pierwsza istotna zmiana nastąpiła już w 2009 roku, kiedy został on rozszerzony o język obcy.

Od roku szkolnego 2011/2012 egzamin gimnazjalny będzie przeprowadzany na nowych zasadach.

- Egzamin będzie sprawdzał opanowanie przez uczniów wiadomości i umiejętności określonych w wymaganiach ogólnych i szczegółowych zawartych w nowej podstawie programowej kształcenia ogólnego.
- Egzamin będzie się składał z takich samych części jak dotychczasowy, ale każda część będzie miała inną strukturę. Inaczej też będzie wyglądać zaświadczenie o szczegółowych wynikach egzaminu gimnazjalnego.
- Część egzaminu z języka obcego nowożytnego będzie podzielona na dwa poziomy: podstawowy i rozszerzony. Do sześciu języków obcych, z których obecnie można zdawać egzamin, od sesji egzaminacyjnej 2013 dojdzie język ukraiński.

Planowane w marcu badania mają na celu nie tylko zrównanie (skalibrowanie) wyników przeszłych egzaminów, które mają szczególne znaczenie dla opisu osiągnięć uczniów za okres 10 lat, ale są także próbą wprowadzenia systematycznego kotwiczenia corocznych egzaminów do wyników roku bazowego. Wybór roku bazowego może być arbitralny i dla każdego egzaminu inny. Dla egzaminu gimnazjalnego rozsądnym wydaje się wybór roku 2012 szczególnie, że jest to pierwszy rok egzaminu uczniów kształconych zgodnie z nową podstawą programową. Ponadto jest szansa, że zmieniona od 2012 roku struktura egzaminu będzie stabilna co najmniej przez kilka lat. Warto też zwrócić uwagę na dodatkowy aspekt tego przedsięwzięcia. Testy zawierające zadania kotwiczące uzupełnione będą o nowe zadania, nieznanne uczniom biorącym udział w badaniach. Powstanie w ten sposób wykalibrowany załączek przyszłego banku zadań.

Model badań zrównujących

Zrównywanie wyników egzaminów jest procedurą statystyczną, która bazuje na analizie danych. Dlatego też, aby zrównać wyniki np. egzaminu gimnazjalnego z lat 2002-2011, konieczne jest podjęcie dwóch decyzji: jaki zastosujemy plan gromadzenia danych – czyli model zrównywania oraz jakie metody statystyczne zastosujemy do analizy tych danych. Pierwsza decyzja jest kluczowa, ponieważ generuje największą część kosztów. Dla tych samych danych można zastosować różne metody analiz statystycznych, aby w efekcie wybrać te, które zapewnią najmniejszy błąd zrównywania.

Do przeprowadzenia badań zrównujących zostanie wylosowana próba reprezentatywna dla szkół. Ze względów organizacyjnych badania zostaną przeprowadzone dla całych klas w szkołach, które wejdą do próby w terminie poprzedzającym o miesiąc sesję egzaminacyjną w gimnazjach. Zeszyty testowe zostaną wcześniej przypisane do uczniów w taki sposób, aby z jednej strony zapewnić rzetelność testowania, a z drugiej umożliwić kotwiczenie wszystkich zadań tak, aby utworzyły jeden megatest.

Do zaobserwowania ogólnego trendu zmian dla całej populacji wystarczyłoby zrobić 3 pomiary kotwiczące np. arkusze 2003, 2006 i 2010. Dlatego dla arkuszy z tych lat przyjęto większą liczbę badanych rozwiązujących każde zadanie włączone do arkusza kotwiczącego z danego roku. Powinno to zapewnić większą dokładność (mniejszy błąd zrównania pomiędzy tymi latami). Zrównane wyniki egzaminów będą zapewne przydatne także do opisanie ewentualnych zmian w poziomie osiągnięć w wyróżnionych warstwach populacji (lokalizacja szkół) i także w krótszych okresach niż 10 lat. Dlatego złożono kotwiczenie arkuszy ze wszystkich lat. Ponadto rezultaty kotwiczenia zostaną wykorzystane do wykalibrowania ze względu na trudność wszystkich zadań na tej samej skali, co stanowi wkład do prac nad bankami zadań.

Jak można zauważyć w tabeli 1. (trzeci wiersz), nowy test kotwiczący będzie częścią składową wszystkich zeszytów testowych. Takie rozwiązanie pozwoli oszacować, jaki wpływ na wyniki będzie miał fakt, że arkusze z minionych lat są dostępne dla szkół i uczniów na stronie CKE. Wyjaśnienia wymaga też fakt, że do oszacowania trendu ewentualnych zmian został wybrany rok 2003, a nie pierwszy rok egzaminów 2002. Decyzja taka została podjęta, gdyż przypuszcza się, że wyniki z 2002 roku obciążone są ze względu na pierwsze doświadczenia szkół w zakresie egzaminów zewnętrznych oraz brak obserwatorów zewnętrznych, którzy zostali wprowadzeni dopiero od 2003 roku.

Tabela 1. Model zrównywania planowany do zastosowania dla egzaminu gimnazjalnego w części humanistycznej i matematyczno-przyrodniczej. Każdy zeszyt będzie zawierał 1/3 zadań z testu kotwiczącego i po 1/3 zadań z arkuszy egzaminacyjnych z zaznaczonych lat

Rok/próba	1	2	3	4	5	6	7	8	9	10	11	Liczba rozwiązujących zadania z danego testu
Egzamin 2011												9000
Nowy test kotwiczący												9000
2002												818
2003												3273
2004												1636
2005												1636
2006												3273
2007												1636
2008												1636
2009												1636
2010												2455
Próba/zeszyt	818	818	818	818	818	818	818	818	818	818	818	

Jak już wcześniej wspomniano, badania zostaną przeprowadzone 14 i 15 marca 2011 roku. Wybór terminu testowania w bezpośredniej bliskości sesji właściwej został podyktowany koniecznością zapewnienia możliwie jak najwyższej motywacji uczniów do pracy z zadaniami testu.

Tabela 2. Statystyki opisujące arkusze egzaminacyjne planowane do zrównania (część matematycznoprzyrodnicza)

	2002	2003	2004	2005	2006	2007	2008	2009	2010	Nowy test
Liczba zdających	558257	548716	539591	526462	522285	505759	473683	451574	434120	3564
Liczba zadań otwartych	9	11	11	13	6	10	12	9	9	9
Surowy wynik średni	28.2	25.8	24.5	24.3	23.9	25.3	27.1	26.0	23.9	25.1
Średnia łatwość zadań	56.3	51.5	50.0	48.5	47.8	50.7	54.2	52.0	47.8	50.0
Odchylenie standardowe	8.89	10.91	11.03	10.15	10.3	10.2	10.7	11.0	9.63	10.4
Skośność rozkładu	0.05	0.17	0.29	0.30	0.34	0.47	0.25	0.25	0.40	0.16
Wskaźnik rzetelności	0.86	0.87	0.88	0.88	0.88	0.88	0.87	0.90	0.87	0.92
Błąd standardowy	3.29	3.88	3.81	3.5	3.6	3.6	3.9	3.50	3.50	2.97
Średnia moc różnicująca zadań	0.42	0.46	0.47	0.46	0.46	0.46	0.46	0.48	0.42	0.46
Wyniki egzaminu w poszczególnych latach opisane z wykorzystaniem teorii wyniku zadania (IRT)										
Miara poziomu osiągnięć w jednostkach logit względem średniej łatwości arkusza unormowanego do wartości 0	0.89	0.48	0.28	0.42	0.38	0.42	0.57	0.58	0.22	
Odchylenie standardowe	1.03	1.03	1.05	1.08	1.03	1.03	1.03	1.14	0.95	
INFIT	0.96	0.99	0.98	1.02	0.97	1.00	1.02	1.00	0.95	
OUTFIT	0.97	0.99	1.00	0.99	0.98	0.99	1.02	1.03	1.02	

Źródło: opracowanie własne na podstawie wyników egzaminów dostarczonych przez CKE. Oszacowanie parametrów zgodnie z IRT - Dorota Węziak-Białowska

Za cały egzamin zdający mógł uzyskać maksymalnie 50 punktów, z czego 25 punktów za 25 zadań zamkniętych i 25 punktów za zadania otwarte, których liczba podlegała fluktuacjom w ciągu 9 lat (por. drugi wiersz tabeli). Miara poziomu osiągnięć wyrażona w jednostkach logit (czwarty wiersz od dołu) jest odniesiona do średniej łatwości testu unormowanej w każdym roku oddzielnie do wartości 0. W ostatnich dwóch wierszach podane są statystyki dopasowania wyników do modelu Rascha. Współczynniki te są średnimi kwadratów w formie statystyki χ^2 podzielonymi przez stopnie swobody tak, że wyrażane są one w skali stosunkowej z oczekiwaną wartością 1 (Bond T.G., Fox Ch. M., 2001; Smolik M., 2009). Jeżeli dla wyników 2005 wartość wskaźnika infit = 1,02, to oznacza, że zaobserwowane wyniki odpowiedzi są o 2 procent bardziej zróżnicowane niż przewiduje model.

W ostatniej kolumnie przedstawione są parametry testu, który był zastosowany jako kotwiczący na próbie losowej 3564 uczniów w badaniach pilotażowych zrównywania.

Podsumowanie

Ze zrównywaniem wyników po kilku latach wiążą się dwa istotne problemy:

1. zapewnienie motywacji podobnej do motywacji na egzaminie,
2. uwzględnienie faktu, że arkusze egzaminacyjne z ubiegłych lat są dostępne na stronie www Centralnej Komisji Egzaminacyjnej i w różnym stopniu ich zadania są wykorzystywane do ćwiczeń przez nauczycieli.

Praktycznie nie ma możliwości zapewnienia takiej samej motywacji na badaniach, jak na egzaminie. Można tylko przy odpowiedniej organizacji badań zminimalizować tę różnicę. Można też podjąć próbę oszacowania różnicy poziomu motywacji na egzaminie i podczas badań. W tym celu badania zostały zaplanowane na miesiąc przed egzaminem i dla uczniów będą stanowić egzamin próbny. Oznacza to jednak, że wyniki, jakie uczniowie osiągną w testach, muszą do nich wrócić w rozsądnym czasie przed egzaminem – około 2 tygodni. Przy takim założeniu zespół badawczy musi w ciągu dwóch tygodni od przeprowadzenia badań ocenić prace, przetworzyć wyniki i zakomunikować je uczniom i szkołom.

Drugi problem wiąże się z koniecznością oszacowania wpływu dostępności testów na łatwość zadań, która będzie oszacowana podczas badań. Tutaj zostały podjęte zabiegi:

1. każdy test rozwiązywany przez badanych będzie składał się z zadań ze znanych już arkuszy z lat ubiegłych i z zadań, które nie mogły być znane uczniom
2. w kwestionariuszu ankiety dla nauczyciela zostaną zamieszczone pytania dotyczące wykorzystywania w procesie dydaktycznym arkuszy egzaminacyjnych z ubiegłych lat do ćwiczeń
3. w kwestionariuszu ankiety dla ucznia zostaną zamieszczone pytania, mające umożliwić oszacowanie, w jakim stopniu zadania występujące w arkuszu były już uczniom znane.

Jest to także próba zebrania informacji do oszacowania wpływu egzaminów ze-wnętrznych na ocenianie wewnątrzszkolne. Ocenianie zewnętrzne jest tylko częścią systemu oceniania – pomimo że ważną częścią, to nie powinno być ono postrzegane jako dominujące i nie powinno wpływać na zawężenie procesu oceniania wewnątrzszkolnego. Szczególnie że, jak już wcześniej podkreślono, ocenianie wewnątrzszkolne i ocenianie poprzez egzaminy zewnętrzne służą odrębnym celom, które się dopełniają, a nie wykluczają. Warto zadbać, aby została zachowana rozsądna równowaga. Skuteczne i pełne wewnątrzszkolne ocenianie (ocenianie kształtujące, orientujące i sumujące) jest warunkiem koniecznym zapewnienia rezultatów, które są przedmiotem pomiaru na egzaminie zewnętrznym.

Bibliografia:

1. Bond T. G, Fox Ch. M., *Applying The Rasch Model. Fundamental Measurement In Human Sciences*, Lawrwnce Erlbaum Associates, Publischers, London, 2001.
2. Marquand J., *Studium wstępne krajowego systemu oceniania w polskim szkolnictwie ponadpodstawowym*, Departament Kształcenia Ogólnego, Departament Współpracy z Zagranicą MEN, 1993.
3. Niemierko B., *Zrównywanie wyników sprawdzianu 2004 do wyników sprawdzianu 2003* [w:] Niemierko B. Szaleniec H. (red.) *Standardy wymagań i normy testowe w diagnostyce edukacyjnej*, PTDE 2004.
4. Niemierko B., *Czy egzaminy zakłócają pracę nauczyciela*, [w:] Niemierko B. Szmigel M. K. (red.) *Uczenie się i egzamin w oczach nauczyciela*, PTDE 2008.
5. Saida Ch., Hattori T., *Post-hoc IRT equating of previously administered English tests for comparison of test stores*, *Language Testing* April 2008 vol. 25 no. 2 187-210, SAGE <http://ltj.sagepub.com/content/25/2/187.refs.html>.
6. Shepard L.A., *Classroom Assessment*, [w:] Brennan R.L., (red.) *Educational Measurement*, Praeger Publishers, 2006.
7. Smolik M., *Zrównywanie wyników Sprawdzianu 2004 i 2005 do wyników Sprawdzianu 2003 z wykorzystaniem modelu Rascha (program Winsteps)*, [w:] Szaleniec H., (red.) *Teoria wyniku zadania IRT. Zastosowania w polskim systemie egzaminów zewnętrznych*, CKE, Warszawa 2009.
8. Szaleniec H., *Wykorzystanie probabilistycznych modeli zadania testowego do zrównywania wyników sprawdzianu 2003-2005 i budowa banku zadań* [w:] Szaleniec H., (red.) *Teoria wyniku zadania IRT. Zastosowania w polskim systemie egzaminów zewnętrznych*, CKE, Warszawa 2009.