

Ścieżki rozwoju egzaminu gimnazjalnego

Dokąd zmierza egzamin gimnazjalny? Czy możliwa jest zamiana tradycyjnego, „papierowego” systemu oceniania na e-ocenianie? Czy młodzież gimnazjalna już całkiem niedługo będzie zdawać swój egzamin na ekranach komputerów? Artykuł pomoże odpowiedzieć na te i inne pytania oraz skłoni do refleksji nad przyszłością egzaminu gimnazjalnego (i w ogóle – egzaminów zewnętrznych i ich oceniania) zarówno entuzjastów wykorzystania systemów elektronicznych, jak i tych, którzy są do nich nastawieni sceptycznie.

Pierwszy dziesięcioletni cykl egzaminu gimnazjalnego, który rozpoczął się wraz z powstaniem gimnazjum i egzaminów zewnętrznych, zakończył się w 2011 roku. W tym okresie (2002–2011) egzamin był przygotowywany i przeprowadzany w myśl standardów wynikających z obowiązującej wówczas podstawy programowej, według tych samych założeń merytorycznych, pomiarowych i statystycznych dla każdego rocznika zdających. Warto zauważyć, że do tej pory egzamin w zasadzie nie zmieniał się w obrębie jego dwóch części – humanistycznej i matematyczno-przyrodniczej. Przeobrażenia zaczęły się w 2009 roku, gdy rozszerzono go o język obcy nowożytny zdawany na jednym poziomie, natomiast w kwietniu 2012 gimnazjaliści, kształceni już zgodnie z nową podstawą programową, przystąpili po raz pierwszy do egzaminu w nowej formule. W jego części humanistycznej zostały wyodrębnione **język polski** (arkusz egzaminacyjny złożony z zadań otwartych i zamkniętych) i **historia z wiedzą o społeczeństwie** (arkusz z zadaniami

zamkniętymi). Z kolei w części matematyczno-przyrodniczej wydzielono **matematykę** (arkusz z zadaniami otwartymi i zamkniętymi) oraz **przedmioty przyrodnicze** (arkusz z zadaniami zamkniętymi). Formułę egzaminu z języka obcego nowożytnego wzbogacono o część zaawansowaną na poziomie rozszerzonym, która w odróżnieniu od egzaminu na poziomie podstawowym zawiera nie tylko zadania zamknięte, lecz także otwarte.

Rezultaty ze wszystkich sześciu zakresów nowego egzaminu, które są brane pod uwagę podczas rekrutacji absolwentów gimnazjum do szkoły ponadgimnazjalnej, trafią do szkół pod koniec czerwca. Do oceny wyników egzaminu w nowej formie należy podejść z rozwagą, konieczna jest refleksja. Zmieniła się nie tylko podstawa programowa, lecz także formuła egzaminu. Przekształcono cały konstrukt (ang. *construct*), na którym bazują przygotowanie i ocena wyników nowego egzaminu – ideę, typy zadań, relacje zadań egzaminacyjnych

w stosunku do umiejętności kształconych zgodnie z nową podstawą.

Na początku kolejnego rozdziału w historii egzaminu gimnazjalnego nasuwa się kilka pytań. Czy podobnie jak w poprzedniej dekadzie czeka nas długoterminowy stabilny okres tego doniosłego już egzaminu? W jakim zakresie będzie on doskonalony? Możemy przewidywać różne warianty jego rozwoju, a prawdopodobieństwo wystąpienia określonych scenariuszy zależy od działań związanych m.in. z odpowiedziami na zamieszczone poniżej pytania.

Czy możemy oczekiwać zmiany skali komunikowania wyników, pomimo zapisu w Rozporządzeniu Ministra Edukacji Narodowej z dnia 30 kwietnia 2007 r. w sprawie warunków i sposobu oceniania, klasyfikowania i promowania uczniów i słuchaczy oraz przeprowadzania sprawdzianów i egzaminów w szkołach publicznych głoszącego, że wyniki komunikowane są w skali procentowej pochodnej od wyników surowych ▶



▶ (podobnie jak dla egzaminu maturalnego) i dodatkowo w skali centylowej (w której przedstawiane jest ułożenie wyniku danego ucznia w rozkładzie wyników wszystkich uczniów piszących ten sam test)?

Czy w najbliższych latach możemy się spodziewać kotwiczenia, zrównywania i kalibrowania wyników egzaminów uzyskiwanych z zastosowaniem różnych arkuszy egzaminacyjnych tak, aby niwelowane były różnice wynikające z nieuniknionych wahań trudności arkuszy egzaminacyjnych pomiędzy latami a także edycjami egzaminu w tej samej sesji (termin główny i termin dodatkowy)?

Czy w przyszłym roku lub w najbliższych latach czeka nas zmiana technologii oceniania zadań otwartych, czyli przejście od tradycyjnego oceniania na oryginalnych pracach uczniów do oceniania na ekranie monitora na podstawie skanów zadań (tzw. e-ocenianie)?

Czy w naszym systemie egzaminacyjnym możliwe jest przejście od „papierowego” egzaminu do egzaminu zdawanego online przy komputerze (e-egzamin) z wykorzystaniem szerokiej gamy typów zadań, które nie są dostępne na tradycyjnym egzaminie?

Zanim przejdziemy do rozważenia czterech problemów zasygnalizowanych w py-

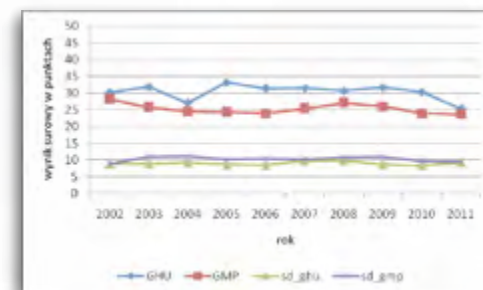
taniach, zatrzymajmy się przez chwilę na dziesięcioletniej przeszłości egzaminu gimnazjalnego. Nie będę wnikał w zagadnienia dotyczące oceny egzaminu z punktu widzenia jego roli i wyników, poziomu poszczególnych grup umiejętności przewidzianych do sprawdzania i sprawdzanych na egzaminie, gdyż nie stanowi to przedmiotu tego artykułu. Jest to bowiem zadanie adresowane do dydaktyków przedmiotów, które obejmuje egzamin gimnazjalny, i wymaga ono rzetelnego, starannie zaplanowanego studium.

Komunikowanie wyników egzaminu gimnazjalnego w latach 2002–2011

Dotychczas dla egzaminu gimnazjalnego wyniki komunikowane były uczniom jako suma punktów uzyskanych za rozwiązanie poszczególnych zadań. Taką skalę nazywamy **skalą wyników surowych**. Jej istotnym mankamentem jest to, że przedziały pomiędzy poszczególnymi jej stopniami nie są równe. Różnica pomiędzy 15. a 16. punktem z egzaminu gimnazjalnego w części humanistycznej niekoniecznie równa się różnicy jednego punktu pomiędzy 48. i 49. Ponadto wynik np. 35 punktów w 2004 roku może oznaczać całkiem co innego niż ten sam wynik uzyskany na egzaminie rok później, a to z powodu nieuniknionej fluktuacji trudności arkuszy egzaminacyjnych zastoso-

wanych w kolejnych latach. Na **rysunku 1** przedstawiono średnie dla wyników surowych w kolejnych latach uzyskane przez wszystkich gimnazjalistów piszących egzamin z wykorzystaniem arkusza standardowego dla części humanistycznej i matematyczno-przyrodniczej oraz miary rozrzutu (dyspersji) wyników, czyli odchylenia standardowe.

Niestety, wyniki przedstawione na **rysunku 1** nie uprawniają nas do wyciągania jakichkolwiek wniosków na temat zmian w poziomie osiągnięć gimnazjalistów na przestrzeni dziesięciu lat. Każdy z tych wyników zależy nie tylko od poziomu umiejętności gimnazjalistów piszących egzamin i od towarzyszącego pomiarowi losowego błędu pomiaru, lecz także od trudności zastosowanego w danym roku arkusza egzaminacyjnego. Nawet przy bardzo starannej procedurze tworzenia arkuszy wahania trudności są nieuniknione. Tak więc na podstawie posiadanych danych nie jesteśmy w stanie stwierdzić, czy np. w 2005 roku średni poziom osiągnięć – mierzony wskaźnikiem, którym jest wynik egzaminu gimnazjalnego w części humanistycznej – jest wyższy niż w roku 2004. Wyciągnięcie jednoznacznych wniosków uniemożliwia brak wiedzy o tym, czy arkusz z 2004 roku był trudniejszy od tego z 2005 roku, a jeśli tak, to o ile. Dlatego konieczne jest wpro-



Rysunek 1. Średnie wyniki surowe i miary różnicowania wyników (odchylenia standardowe) części humanistycznej i matematyczno-przyrodniczej w latach 2002–2011

► wadzenie rozwiązań, które pozwoliłyby na kalibrowanie wyników egzaminów uzyskiwanych z zastosowaniem różnych arkuszy egzaminacyjnych tak, aby niwelowane były różnice wynikające z nieuniknionych wahań trudności arkuszy egzaminacyjnych pomiędzy latami. Aby możliwe było takie porównywanie, konieczne jest zasygnalizowane w pytaniu drugim zrównywanie i kalibrowanie wyników.

Do szerszej komunikacji średnich wyników szkół wykorzystywana była dodatkowo dziewięciostopniowa skala staninowa – skala standardowa o średniej 5 i odchyleniu standardowym 2, kalibrowana w każdym roku za pomocą danego arkusza egzaminacyjnego na populacji zdających egzamin. W każdym roku tworzona była osobna skala dla egzaminu pisanego na podstawie arkuszy standardowych (dla uczniów bez dysfunkcji i z dysleksją rozwojową), osobno dla uczniów niedosłyszących, niewidomych, z upośledzeniem lekkim. W praktyce oznacza to, że w każdym roku średnia z egzaminu, którego wyniki wyrażone są w tej skali, wynosiła 5, a odchylenie standardowe 2. Nawet gdyby poziom osiągnięć zdających

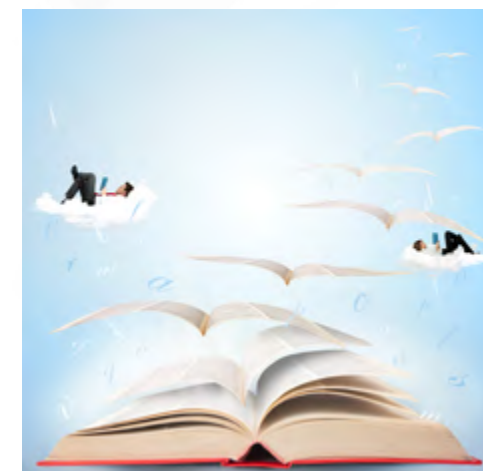
egzamin zmieniał się znacznie w kolejnych latach, to średnia zawsze wynosiłaby 5.

Skala ta, choć wygodna dla komunikowania średnich wyników szkół, jest zbyt krótka (dziewięć stopni) w przypadku wykorzystywania rezultatów egzaminów do rekrutacji uczniów do szkoły programowo wyższej. Do celów rekrutacyjnych bardziej przydatna jest dłuższa skala pozwalająca na większe zróżnicowanie uczniów ze względu na wynik egzaminu.

Skala staninowa także była budowana dla średnich wyników dla szkół. Centralna Komisja Egzaminacyjna co roku przygotowywała krajowe sprawozdanie z egzaminu, w którym omawiane były wyniki poszczególnych egzaminów, charakterystyki zadań i parametry statystyczne danego egzaminu, takie jak wskaźniki tendencji centralnej (średnia arytmetyczna wyników, mediana i modalna), wskaźniki dyspersji (odchylenie standardowe), moc różnicująca zadań, wskaźnik rzetelności egzaminu przeprowadzonego za pomocą poszczególnych arkuszy egzaminacyjnych. Sprawozdania są dostępne na stronie internetowej [CKE](#).

W pierwszych latach przeprowadzania egzaminów zewnętrznych również okręgowe komisje egzaminacyjne publikowały dodatkowo raporty regionalne.

Jako ewidentny sukces w zakresie komunikowania szkołom informacji związanej z wynikami egzaminu należy uznać opracowanie metodologii i wdrożenie szacowania wskaźnika edukacyjnej wartości dodanej dla każdej szkoły gimnazjalnej i maturalnej¹. Prace nad EWD rozpoczęły się w 2005 roku w ramach projektu CKE „Badania dotyczące rozwoju metodologii szacowania wskaźnika edukacyjnej wartości dodanej” współfinansowanego przez EFS. W ramach projektu rozwinięto i wdrożono metodologię szacowania wskaźników EWD gimnazjów i szkół ponadgimnazjalnych kończących się maturą. Towarzyszyły temu starannie zaplanowane szkolenia różnych grup użytkowników sięgających do wyników egzaminacyjnych i wskaźnika EWD. Na uwagę zasługuje fakt, że wówczas została przybliżona nauczycielom także standardowa skala o średniej 100 i odchyleniu standardowym 15, w której podawany jest wskaźnik edukacyjnej wartości dodanej i wyniki średnie z egzaminu dla szkoły. ►



¹ Począwszy od 2005 roku zespół kierowany przez dr. hab. R. Dolatę w ramach projektu CKE „Badania dotyczące rozwoju metodologii szacowania wskaźnika edukacyjnej wartości dodanej” współfinansowanego przez EFS rozwinął i wdrożył metodę szacowania wskaźników EWD gimnazjów i szkół ponadgimnazjalnych kończących się maturą. Szczegółowe informacje i dostęp do wskaźników EWD dla szkół znajdują się na stronie internetowej [projektu](#).

► **Czy możemy oczekiwać kolejnej zmiany skali komunikowania wyników egzaminu gimnazjalnego?**

Począwszy od sesji egzaminacyjnej 2012 roku, każdy uczeń zdający egzamin gimnazjalny otrzyma zaświadczenie o szczegółowych wynikach swojego egzaminu. Dla każdego z zakresów egzaminu – język polski, historia i wiedza o społeczeństwie, matematyka, przedmioty przyrodnicze, język obcy nowożytny na poziomie podstawowym, język obcy nowożytny na poziomie rozszerzonym – będą podane dwie liczby: wynik procentowy oraz wynik centylowy.

Do komunikowania rezultatów w postaci procentów wyniku maksymalnego przyzwyczailiśmy się już w przypadku egzaminu maturalnego i egzaminu potwierdzającego kwalifikacje zawodowe. Wynik w skali centylowej wprowadzany jest w naszych egzaminach po raz pierwszy. Informuje, jaki jest odsetek liczby gimnazjalistów (zaokrąglony do liczby całkowitej), którzy uzyskali z danego zakresu wynik niższy niż dany zdający. Np. uczeń, którego wynik centylowy w zakresie matematyki wynosi 85, dowie się, że 85% wszystkich gimnazjalistów uzyskało

za zadania matematyczne wynik taki sam lub niższy niż on, a 15% – wynik wyższy², natomiast jeśli uzyskał wynik centylowy 50 (mediana), dowie się, że 50% wszystkich zdających uzyskało za zadania matematyczne wynik taki sam lub niższy niż on, a 50% – wynik wyższy niż on. Mamy tu na myśli gimnazjalistów egzaminowanych za pomocą tego samego arkusza. Rozkład wyników w skali centylowej jest prostokątny – niezależnie od tego, jaki był rozkład wyników surowych (por. Rysunek 2).

W skali centylowej szerokość przedziałów skali od C1 (centyl 1) do C99 (centyl 99) wynosi 1% wyników. Skrajne centyle oznaczone na rysunku C1⁻ i C99⁺ zawierają po 0,5% wyników³. Tak więc wynik centylowy egzaminu danego gimnazjalisty określa usytuowanie jego wyniku w populacji wszystkich wyników z danego egzaminu. Informacja ta może być szczególnie cenna przy określaniu szansy dostania się do wybranej szkoły ponadgimnazjalnej.

Równoległe z obowiązującymi urzędowymi skalami do komunikowania wyników gimnazjalnych będą funkcjonowały dwie stan-

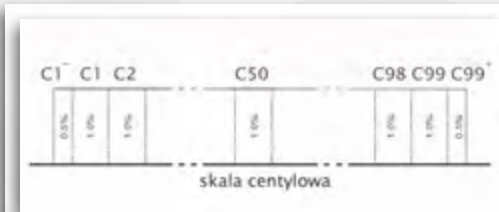
dardowe skale: coraz powszechniej znana w środowisku nauczycieli szkół gimnazjalnych **skala standardowa o średniej 100 i odchyleniu standardowym 15**, w której komunikowane są wyniki i wskaźniki edukacyjnej wartości dodanej oraz **skala staninowa o średniej 5 i odchyleniu standardowym 2**, która od 2002 roku była używana w raportowaniu wyników szkół w sprawozdaniach CKE i OKE prezentujących rezultaty sprawdzianu i egzaminu gimnazjalnego. Obydwie skale można wyprowadzić, przekształcając liniowo podstawową skalę standardową „Z” o średniej 0 i odchyleniu standardowym 1. Rysunek 3. przedstawia porównanie kilku skal z tej rodziny⁴.

Podstawą wielu skal jest wynik standardowy „z”, który jest odchyleniem od średniej podzielonym przez odchylenie standardowe (Hornowska, 2001, s. 134) –

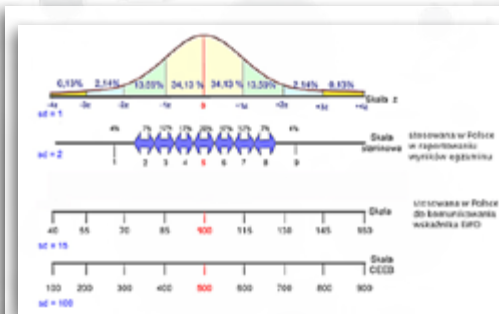
$$Z = \frac{X - X_{\text{średnie}}}{sd}$$

gdzie:

z – wynik w skali standardowej,
x – wynik surowy otrzymany przez zdającego na egzaminie,



Rysunek 2. Rozkład wyników w skali centylowej



Rysunek 3. Porównanie czterech przykładowych skal standardowych

² Informator o egzaminie gimnazjalnym od roku szkolnego 2011/2012, CKE i OKE.

³ E. Hornowska, cytując A. Matuszewskiego, nazywa skrajne centyle odpowiednio centylem 0 i centylem 100.

⁴ Empiryczne rozkłady wyników uzyskiwane w naszych egzaminach w różnym stopniu odbiegają od rozkładu normalnego. Aby otrzymać rozkład wyników w nowej skali zbliżony do normalnego, musimy dodatkowo dokonać przekształcenia rozkładu. Jest to już przekształcenie nieliniowe.

► $x_{\text{średnie}}$ – wynik średni dla ogółu zdających, sd – odchylenie standardowe.

Jak już wspomniano, w skali standardowej „z” wynik średni $z_{\text{średnie}}$ równa się 0, a odchylenie standardowe sd jest równe 1. Cały zakres wyników najczęściej mieści się od -3 do $+3$. Przekształcenie wyników surowych na wyniki standardowe nie zmienia kształtu rozkładu wyników, czyli skośność i wypiętrzenie (kurtoza) rozkładu pozostają niezmienione.

Wynik w skali standardowej informuje, o ile odchylen standardowych powyżej (lub poniżej) wartości średniej znajduje się dany wynik (dla wyników poniżej średniej „z” jest ujemne). I tak $z = -0,1$ oznacza, że dany wynik określa wartość równą $0,1$ odchylenia standardowego poniżej średniej, $z = -2,5$ odpowiednio $2,5$ odchylenia standardowego poniżej średniej, a $z = 2,5$ oznacza wynik $2,5$ odchylenia standardowego powyżej średniej. Chcąc z powrotem przejść do wyników surowych, wystarczy pomnożyć wynik w skali „z” przez odchylenie standardowe dla wyników surowych i dodać wartość średnią dla wyników surowych –

$$X = X_{\text{średnie}} + z * sd$$

Na rysunku 3. jako drugą od góry przedstawiono skalę staninową. Wynik otrzymujemy,

mnożąc wynik w skali „z” przez nowe odchylenie standardowe (w tym przypadku 2) i dodając nową średnią, która tutaj wynosi 5 –

$$W = Z * 2 + 5$$

gdzie:

w – wynik w nowej skali,

z – wynik w standardowej skali „z”.

W podobny sposób uzyskujemy wyniki w skali o średniej 100 i odchyleniu standardowym 15, stosowanej w Polsce do komunikowania wskaźnika EWD –

$$W = Z * 15 + 100$$

Na ostatniej pozycji na rysunku znajduje się skala, w której średnia wynosi 500, a odchylenie standardowe 100. Skala ta stosowana jest przez The College Entrance Examination Board (skala CEEB). Warto o niej pamiętać, gdyż jest powszechnie stosowana w wielu systemach egzaminacyjnych i w międzynarodowych badaniach porównawczych wyników osiągnięć (np. PISA).

Perspektywa zrównywania wyników egzaminacyjnych

Jednym z celów wprowadzenia egzaminów zewnętrznych było oczekiwanie, że ich wy-

niki będą przydatne do śledzenia zmian edukacyjnych w skali całego kraju na przestrzeni wielu lat. Obecnie nie jest to możliwe z powodu nieuniknionej w praktyce fluktuacji trudności arkuszy egzaminacyjnych w kolejnych latach. Dlatego konieczne jest wspomniane zrównywanie wyników. Stosuje się je w wielu systemach egzaminacyjnych, np. w USA (SAT – Scholastic Assessment Test oraz ACT – American College Testing), w Izraelu (Psychometric Entrance Test), w Szwecji (Swedish Scholastic Assessment Test) i w międzynarodowych badaniach porównawczych, np. TIMSS (Trends in International Mathematics and Science Study), PIRLS (Progress in International Reading Literacy Study) oraz PISA (Programme for International Student Assessment), w których uczestniczy także nasz kraj (Pokropek, 2011).

Zrównywanie wyników egzaminu jest procesem statystycznym, w którym wykorzystuje się powiązanie arkuszy egzaminacyjnych (ang. *linking*) poprzez wprowadzanie wspólnych zadań lub poddawanie badaniom osób rozwiązujących zadania z tych samych arkuszy. Zrównanie wyników ma na celu umożliwienie zamiennego posługiwania się nimi – wyniki zrównane możemy traktować jako równoważne. Mislavy (1992) podkreśla, że proces zrównywania wyników testu jest ściśle związany z procesem tworzenia na-►

► rzędzi pomiaru. Najczęściej zrównanie funkcjonuje dobrze, dlatego że przestrzegano procedury dobrego konstruowania testu. W tym kontekście warunkiem koniecznym do przeprowadzenia procesu zrównywania wyników danego egzaminu z dwóch lub więcej kolejnych sesji jest zastosowanie na egzaminach arkuszy egzaminacyjnych zbudowanych do pomiaru tego samego zakresu umiejętności i według tych samych specyfikacji statystycznych. Oznacza to, że zrównanie wyników danego egzaminu w kolejnych latach jest możliwe, jeżeli do mierzenia wykorzystano ten sam konstrukt i zastosowano te same specyfikacje tworzenia arkuszy egzaminacyjnych. Możemy przyjąć, że ten warunek dla egzaminu gimnazjalnego w latach 2002–2011 był spełniony w wystarczającym stopniu.

Problem zrównywania wyników egzaminacyjnych był podejmowany pilotażowo przez CKE w latach 2004 i 2005 (Szaleniec, 2005; Niemierko, 2004). Studium zrównania wyników obejmowało trzy populacje uczniów piszących sprawdzian: P_{03} , P_{04} i P_{05} z których każda rozwiązywała odpowiednio arkusz sprawdzianu: T_{03} , T_{04} i T_{05} , oraz dwie populacje piątoklasistów, z których zostały pobrane próby celowe uczniów S_{V04}^1 i S_{V05}^1 . Ci ostatni rozwiązywali testy kotwiczące zbudowane w połowie z zadań z poprzedniego roku i w połowie z następnego.

Obecnie w Instytucie Badań Edukacyjnych trwają prace mające na celu zrównanie wyników sprawdzianu i egzaminu gimnazjalnego w części humanistycznej i matematyczno-przyrodniczej począwszy od 2002 roku oraz wyników matury z matematyki, języka polskiego i języka angielskiego począwszy od roku 2010, kiedy to matematyka stała się obowiązkowym przedmiotem na tym egzaminie. Badania zrównujące wyniki egzaminu gimnazjalnego zostały przeprowadzone w 2011 roku na krajowej losowej próbie 442 szkół. Na początku czwartego kwartału 2012 roku zostaną przedstawione ich wyniki, czyli zrównane wyniki z dziesięciu lat wykalibrowane w skali wyników z rocznika 2003, który został wybrany przez zespół badawczy jako rok odniesienia. Wyniki dla szkół zostaną przedstawione w takiej samej skali, w jakiej komunikowane jest EWD, czyli w skali standardowej o średniej 100 i odchyleniu standardowym 15. W bieżącym roku prowadzone są badania zrównujące dla sprawdzianu. W 2013 roku planowane są badania dotyczące egzaminu maturalnego z matematyki, a w 2014 – z języka polskiego i języka angielskiego.

Opracowana i zastosowana podczas sygnalizowanych powyżej badań metodologia zrównywania może stanowić podstawę do przygotowania wdrożenia zrównywania wyników egzaminacyjnych w systemie

polskich egzaminów zewnętrznych. Najlepszym rozwiązaniem byłoby wprowadzenie zadań kotwiczących do arkuszy egzaminacyjnych stosowanych podczas każdej sesji – zadań, które nie byłyby publikowane, a więc możliwe do zastosowania do kotwiczenia w kolejnej sesji. Takie rozwiązanie pozwoliłoby na komunikowanie uczniom zrównanych wyników na zakończenie sesji egzaminacyjnej. Problem ten jest szczególnie ważny w przypadku egzaminu maturalnego z powodu ubiegania się o indeks w danym roku absolwentów z różnych lat. Jest to wyzwanie, które niewątpliwie powinno być podjęte przez system egzaminacyjny i to w najbliższym czasie.

Zmiana technologii oceniania zadań otwartych

15 maja 2012 roku Minister Edukacji Narodowej podpisała decyzję o uruchomieniu projektu „Wdrożenie systemu informatycznego do e-oceniania”. Projekt jest realizowany przez CKE i współfinansowany z Europejskiego Funduszu Społecznego w ramach III Priorytetu Programu Operacyjnego Kapitał Ludzki 2007–2013, Działanie 3.2 Rozwój systemu egzaminów zewnętrznych. Projekt stanowi kontynuację wcześniejszych działań, które zainicjowano w październiku 2007 roku (Szaleniec, 2010). W pierwszym etapie trującym do końca 2009 roku przy współ-

OŚWIATOWE ABC

Oświatowe ABC. Egzaminy

Na stronie CKE został zamieszczony dodatek informacyjno-promocyjny do „Gazety Wyborczej” z serii „Oświatowe ABC” poświęcony egzaminom zewnętrznym.

Zapraszamy do lektury!

► pracy z angielską firmą DRS przeprowadzono pilotaż e-oceniania z wykorzystaniem oprogramowania e-marker® dla sprawdzianu, egzaminu gimnazjalnego w części humanistycznej i matematyczno-przyrodniczej oraz matematyki na poziomie maturalnym, a także studium wykonalności w zakresie wdrożenia e-oceniania. W pilotażu nowej technologii oceniania uczestniczyło ponad 330 egzaminatorów ze wszystkich okręgowych komisji egzaminacyjnych.

Pilotaż pokazał przede wszystkim, że w naszym systemie egzaminów zewnętrznych możliwe jest wprowadzenie oceniania odpowiedzi na zadania otwarte z wykorzystaniem kopii rozwiązań widocznych na monitorze komputera zamiast na oryginalnej papierowej pracy i że takie rozwiązanie może być zaakceptowane przez tych, którzy od lat odpowiadają za organizację tradycyjnego oceniania w OKE. Potwierdziły to opinie egzaminatorów uczestniczących w pilotażu. Ich zdaniem badanie wykazało, że zmiana technologii oceniania nie sprawia poważnych kłopotów. Nawet ci egzaminatorzy, którzy początkowo widzieli wiele przeszkód uniemożliwiających wprowadzenie tej technologii do oceniania dłuższych wypowiedzi w humanistycznej części egzaminu gimnazjalnego, po pilotażu przyznali, że jest to bardzo przyjazna forma zapewniająca większą niż dotych-

czas rzetelność oceniania. Ponadto bardzo istotnym argumentem za wykorzystaniem elektronicznego systemu oceniania jest według nich możliwość skoncentrowania się na zagadnieniach merytorycznych z pominięciem konieczności zarządzania pracami, szczególnie w przypadku, gdy egzaminator chce specjalizować się w ocenianiu poszczególnych zadań lub wiązek zadań.

W kolejnym etapie, realizowanym w latach 2010–2011 w ramach projektu „Przygotowanie systemu informatycznego do e-oceniania”, system egzaminacyjny pozyskał polską wersję aplikacji do e-oceniania Scoris Assessor zbudowaną przez angielską firmę RM Assessment, która jest częścią RM Education. Testowanie nowego rozwiązania przeprowadzono w ramach pilotażu obejmującego wszystkie ogniwa i etapy procesu oceniania, które mają miejsce w sesji egzaminacyjnej w takiej formie, jaka jest wymagana dla procesu e-oceniania. Została też przygotowana dokumentacja techniczna obejmująca m.in. standaryzację procedur oceniania, standaryzację i kwalifikowanie (dopuszczenie do oceniania) egzaminatorów, hosting, pomoc merytoryczną online dla egzaminatorów, pomoc techniczną, zasady i narzędzia zapewniania jakości w trakcie procesu oceniania poprzez stosowanie losowo przydzielanych zadań weryfikacyj-

nych (ocenionych prac przez zespół ekspertów). W tych działaniach uczestniczyło około 350 osób: egzaminatorów, specjalistów IT z okręgowych komisji egzaminacyjnych, przeszkolonych przez RM ekspertów pomocy dla egzaminatora (pierwsza linia wsparcia helpdesk).

Celem głównym nowego projektu koordynowanego przez Jadwigę Brzdąk (wicedyrektor OKE Jaworzno) jest wdrożenie e-oceniania do systemu egzaminów zewnętrznych. Działania projektu będą się koncentrować na trzech głównych zagadnieniach:

1. Przygotowanie kadry wdrażającej ocenianie prac z wykorzystaniem aplikacji Scoris Assessor w egzaminach, które stopniowo będą coraz szerzej wykorzystywać technologię e-oceniania. Konieczna kadra posiadająca nowe umiejętności będzie składać się z egzaminatorów, przewodniczących zespołów egzaminatorów, koordynatorów egzaminów, osób pełniących funkcje menedżerów sesji oceniania i ekspertów IT odpowiedzialnych za poprawne funkcjonowanie aplikacji.
2. Przeprowadzenie e-oceniania prac egzaminacyjnych z egzaminu maturalnego z matematyki w sesji poprawkowej w 2012 roku. ►



XVIII Ogólnopolska Konferencja Diagnostyki Edukacyjnej

Polskie Towarzystwo Diagnostyki Edukacyjnej w imieniu wszystkich organizatorów ma zaszczyt zaprosić na XVIII Ogólnopolską Konferencję Diagnostyki Edukacyjnej, która odbędzie się we Wrocławiu w dniach 21–23.09.2012 r.

Współorganizatorami konferencji są Okręgowa Komisja Egzaminacyjna we Wrocławiu, Dolnośląska Szkoła Wyższa we Wrocławiu oraz Urząd Marszałkowski Województwa Dolnośląskiego. Tematem tegorocznej konferencji będą regionalne i lokalne diagnozy edukacyjne.

Szczegółowe informacje dostępne na stronie [Polskiego Towarzystwa Diagnostyki Edukacyjnej](http://www.pide.org.pl).

▶ 3. Przeprowadzenie e-oceniania prac egzaminacyjnych z egzaminu gimnazjalnego z zakresu matematyki w sesji głównej w kwietniu 2013 roku.

4. Przygotowanie warunków technicznych i organizacyjnych systemowego wdrożenia e-oceniania w kolejnych egzaminach.

W kontekście planowanych działań w ramach tego projektu, można się spodziewać, że e-ocenianie będzie zastosowane bardzo szeroko już w następnej sesji egzaminu gimnazjalnego. Jest to zmiana bardzo złożona pod względem logistycznym, która już teraz powinna zaangażować system egzaminacyjny w obszarach wykraczających poza ramy samego projektu „Wdrożenie systemu informatycznego do e-oceniania”. Jednocześnie jest to także działanie komplementarne z próbami e-egzaminu, w szczególności jeżeli dąży się do tego, aby zawierał on także zadania otwarte rozszerzonej odpowiedzi.

W kierunku e-egzaminu

W jakim kierunku pójdą zmiany w ocenianiu w obecnym dziesięcioleciu? Ronald K. Hambleton (2006) wymienia trzy kierunki zmian, których jesteśmy świadkami w zakresie teorii i praktyki oceniania. Po pierwsze – od strony analitycznej będącej podstawą zarówno tworzenia narzędzi (szczególnie bu-

dowy banków zadań), jak i analizy wyników egzaminacyjnych – możemy się spodziewać dalszego przechodzenia od klasycznej teorii testu KTT do probabilistycznej teorii wyniku zadania testowego IRT (Item Response Theory). Kolejnym kierunkiem może być coraz większe wykorzystanie komputerów do projektowania i przeprowadzania egzaminów oraz wprowadzenie zindywidualizowanych e-egzaminów dopasowanych do poziomu umiejętności zdającego. Wreszcie możliwe jest także stosowanie w praktyce oceniania nowych typów zadań, które nie mają zastosowania w egzaminach pisanych długopisem na papierze.

Przenikanie technologii informatycznej do egzaminowania (nieformalne i formalne) odbywa się już na naszych oczach. Przykładem tego są e-testy przeprowadzane w ramach wielu projektów o zasięgu lokalnym i krajowym. Pilotaże próbnych e-egzaminów przeprowadza obecnie już wiele podmiotów, np. wydawnictwa i uczelnie wyższe (w tym próba symulacji e-egzaminu na poziomie gimnazjalnym w części matematyczno-przyrodniczej czy z matematyki na poziomie maturalnym).

W ramach projektu CKE „Modernizacja egzaminów zawodowych” współfinansowanego z EFS w 2010 roku odbył się pilotaż etapu pisemnego egzaminów zawodowych (dla

dwóch zawodów – technik ekonomista i kucharz małej gastronomii) z wykorzystaniem systemów elektronicznych (e-egzamin). W 2012 roku odbędzie się kolejny tego typu pilotaż części pisemnej dla nowych egzaminów z pięciu kwalifikacji zawodowych. W to przedsięwzięcie prowadzone w szkołach na terenie całej Polski zaangażowane są wszystkie okręgowe komisje egzaminacyjne.

E-ocenianie i egzaminowanie z wykorzystaniem systemów elektronicznych jest powszechne w zdalnym kształceniu (e-learning), które coraz pewniej znajduje swoje miejsce w kursach doskonalących, studiach podyplomowych, ale też w regularnym kształceniu akademickim i coraz szerzej w szkolnym – w ramach różnych projektów. Przygotowywanie egzaminów dopasowanych do poziomu indywidualnych umiejętności zdających pozwoli w przyszłości na zredukowanie prawie o połowę liczby zadań na egzaminie. Ta dziedzina pomiaru edukacyjnego rozwija się na świecie bardzo dynamicznie. Warto także zauważyć, że rozwój technik multimedialnych ogromnie rozszerzył gamę typów zadań stosowanych zarówno w procesie uczenia się, jak i podczas sprawdzania umiejętności. Przypominają one, w większym stopniu niż te stosowane dotychczas w papierowej wersji egzaminu, działanie konieczne przy rozwiązywaniu nie sztucznych, ale rzeczywistych problemów. ▶

► Jak będzie wyglądać egzamin gimnazjalny w 2021 roku, czyli dziesięć lat od wprowadzenia formuły, która obowiązuje dziś? Spójrzmy na to zagadnienie mniej lub bardziej futurologicznie. Być może różnice – w porównaniu do stanu obecnego – będą jedynie niewielkie, kosmetyczne, polegające na zmianie głównie w zakresie procedur budowania arkuszy egzaminacyjnych. Taki

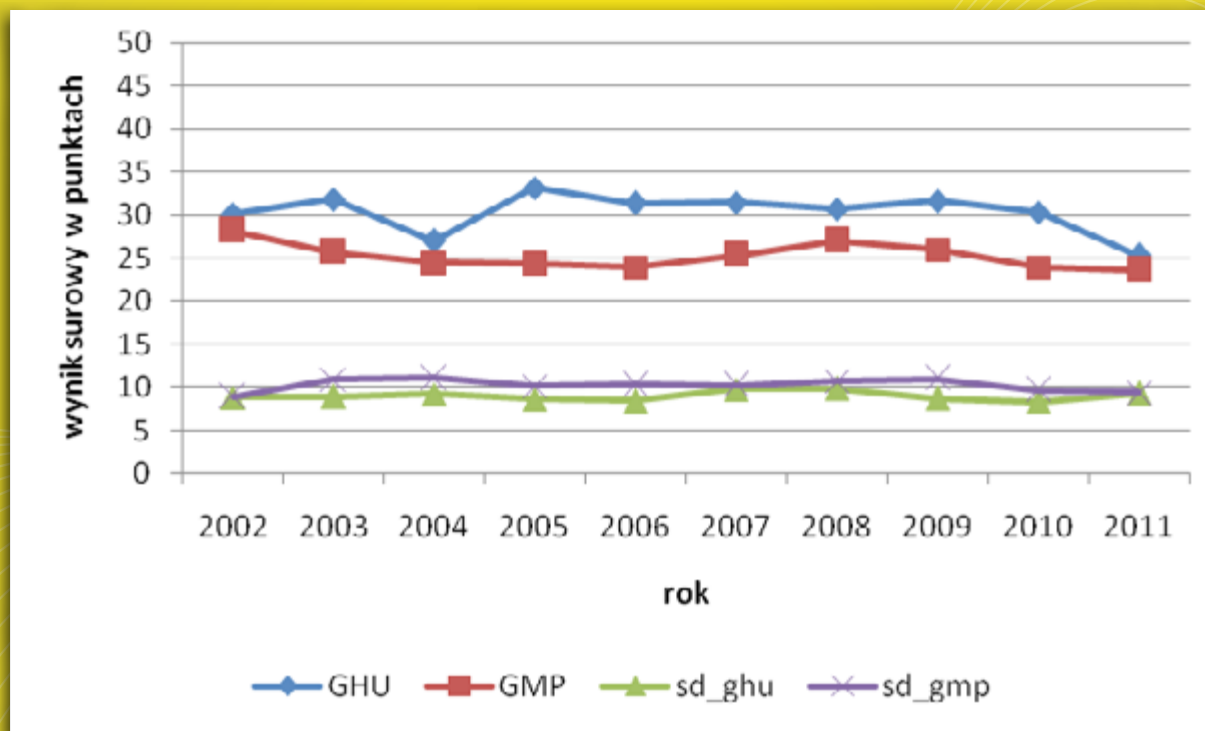
scenariusz moim zdaniem jest mało prawdopodobny. Z drugiej strony możemy sobie wyobrazić e-egzamin (który nie może funkcjonować bez profesjonalnego banku zadań, którego budowa trwa wiele lat) – wydaje mi się, że wizja egzaminu przeprowadzanego z wykorzystaniem systemów elektronicznych z zadaniami zamkniętymi z historii i wiedzy o społeczeństwie a także

przedmiotów przyrodniczych oraz z zadaniami otwartymi ocenianymi przez egzaminatorów online w technologii e-oceniania w przypadku języka polskiego, języka obcego nowożytnego i matematyki jest i optymistyczna, i realistyczna.

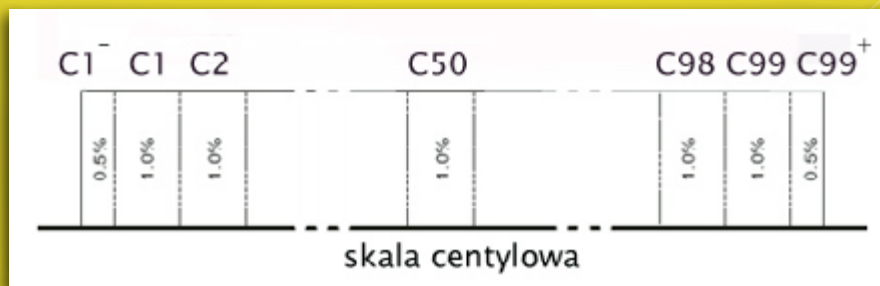
Henryk Szaleniec

Bibliografia

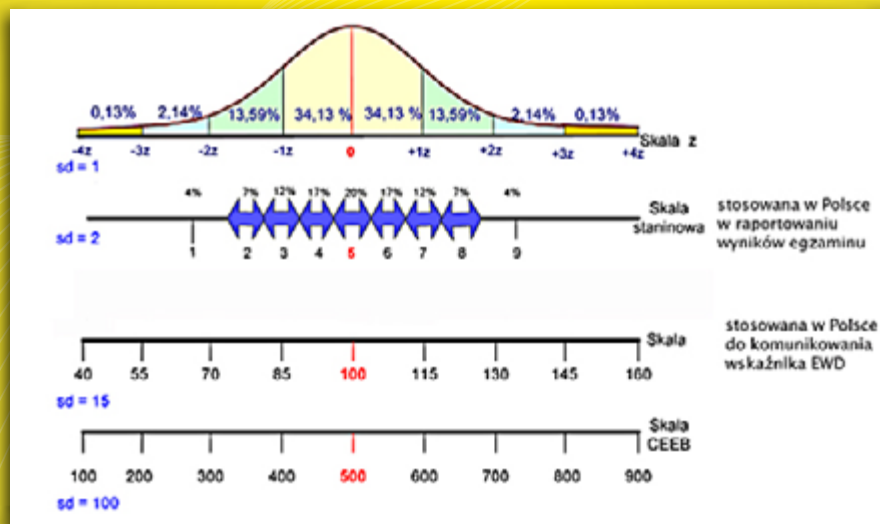
Hambleton R.K., (2006), *Psychometric Model, Test Designs and Item Types for the next Generation of Educational and Psychological Tests*, [w:] *Computer Based Testing and the Internet Issues and Advances*, Bartram D., Hambleton R.K., (red.), Hoboken, NJ: John Wiley & Sons. | Hornowska E., (2001), *Testy psychologiczne. Teoria i praktyka*, Warszawa: Wydawnictwo Naukowe Scholar. | Mislevy R.J., (1992), *Linking educational assessments: concepts, issues, methods, and prospects*, Princeton, NJ: ETS Policy Information Center. | Niemierko B., (2004), *Zrównywanie wyników sprawdzianu 2004 do wyników sprawdzianu 2003*, [w:] Niemierko B., Szaleniec H. (red.), *Standardy wymagań i normy testowe w diagnostyce edukacyjnej*, Kraków: PTDE. | Pokropek A., (2011), *Zrównywanie wyników egzaminów zewnętrznych w kontekście międzynarodowym*, [w:] Niemierko B., Szmigiel M.K. (red.), *Ewaluacja w edukacji: koncepcje, metody, perspektywy*, Kraków: PTDE. | Szaleniec H., (2005), *Wykorzystanie probabilistycznych modeli zadania testowego do zrównania wyników sprawdzianu 2003–2005 i budowania banku zadań*, [w:] Niemierko B., Szyling G. (red.), *Holistyczne i analityczne metody diagnostyki edukacyjnej perspektywy informatyczne egzaminów szkolnych*, Gdańsk: Wydawnictwo Fundacji Rozwoju Uniwersytetu Gdańskiego. | Szaleniec H., (2010), *E-ocenianie to nie tylko zmiana technologii*, [w:] „Edukacja. Studia. Badania. Innowacje”, nr 1(109).



Rysunek 1. Średnie wyniki surowe i miary zróżnicowania wyników (odchylenia standardowe) części humanistycznej i matematyczno-przyrodniczej w latach 2002–2011



Rysunek 2. Rozkład wyników w skali centylowej



Rysunek 3. Porównanie czterech przykładowych skal standardowych