

**Dorota Jurdzińska**

Okręgowa Komisja Egzaminacyjna we Wrocławiu

**Elżbieta Modrzewska**

Okręgowa Komisja Egzaminacyjna w Jaworznie

## **E-ocenie jako innowacja technologiczna i jakościowa w systemie egzaminów zewnętrznych**

### **Wstęp**

Egzaminy zewnętrzne mają istotne znaczenie dla systemu oświaty w Polsce. Co roku uczniowie szkół podstawowych i gimnazjów oraz absolwenci szkół ponadgimnazjalnych przystępują do egzaminów zewnętrznych, które umożliwiają sprawdzenie poziomu ich osiągnięć na zakończenie danego etapu kształcenia. Wyniki egzaminów zewnętrznych pełnią więc ważną funkcję informacyjną, diagnostyczną i ewaluacyjną, dzięki czemu mogą stanowić podstawę doskonalenia systemu kształcenia na różnych poziomach zarządzania oświatą. Ponieważ system egzaminów oddziałuje na delikatną sferę życia społecznego, jaką jest ocena poziomu wykształcenia i ocena jakości pracy szkół, musi zapewniać rzetelność i porównywalność wyników oraz budzić zaufanie do ustalonych ocen i stosownych procedur.

System egzaminów zewnętrznych jest koordynowany przez Centralną Komisję Egzaminacyjną, ale za ocenianie prac odpowiadają okręgowe komisje egzaminacyjne. Można więc zadawać sobie pytanie, czy wobec tego w całym kraju są stosowane jednakowe procedury i jednakowe zasady sprawdzania. Czy egzaminatorzy z różnych komisji stosują schemat oceniania tak samo, niezależnie od tego, w którym województwie sprawdzają?

Na te i inne pytania dotyczące porównywalności oceniania chcemy poszukać odpowiedzi, analizując niezwykle ciekawe wyniki zastosowanego po raz pierwszy w naszym kraju podczas rzeczywistej sesji egzaminacyjnej systemu e-ocenia. E-ocenie zastosowaliśmy w tym roku do oceny prac egzaminu gimnazjalnego z zakresu matematyki z czterech komisji okręgowych: w Jaworznie, Krakowie, Łodzi i we Wrocławiu.

### **Czym jest e-ocenie**

E-ocenie (*electronic marking*) to system umożliwiający ocenianie prac egzaminacyjnych przez egzaminatorów pracujących w domach z wykorzystaniem komputera i internetu. Egzaminatorzy stosowali udostępnioną przez Centralną Komisję Egzaminacyjną aplikację *scoris*, firmy RM Education.

E-ocenie jest nowym rozwiązaniem zarówno pod względem organizacyjnym, jak i technologicznym. Istotna różnica pomiędzy stosowanym dotychczas ocenianiem tradycyjnym a e-oceniem polega przede wszystkim na

dostarczaniu egzaminatorowi rozwiązań zdających na innym niż papier nośniku. Zamiast papierowego oryginału egzaminatorzy otrzymują obrazy rozwiązań zadań, które oceniają na ekranie swojego komputera. W aplikacji mają do dyspozycji wiele narzędzi ułatwiających analizę rozwiązań oraz ich ocenę.

## E-ocenie a porównywalność wyników w skali kraju

Podczas e-ocenia zastosowano innowacyjne rozwiązania organizacyjne:

### 1. Losowy przydział zadań

Elektroniczny dostęp do ocenianych prac pozwolił na losowy przydział zadań zdających także egzaminatorom z poza zasięgu działania komisji danej egzaminacyjnej. Dotychczas prace zdających z każdej komisji były sprawdzane tylko przez egzaminatorów tej komisji. Mógł więc istnieć wpływ np. koordynatora komisji na sposób oceniania.

### 2. Ocenianie zadaniami

Dzięki aplikacji możliwe stało się przydzielenie egzaminatorom do sprawdzania tylko jednego zadania z trzech, a nie całych prac, jak w tradycyjnych ocenianiu. Wynik każdego ucznia był więc ustalany niezależnie przez trzech egzaminatorów.

Do sesji powołano 750 egzaminatorów. Liczba osób z poszczególnych komisji była proporcjonalna do liczby zdających. Ostatecznie w sesji pracowało 744 egzaminatorów. Rozkład egzaminatorów ilustruje poniższa tabela.

**Tabela 1. Liczba egzaminatorów z poszczególnych komisji oceniających przydzielone zadania**

	OKE Jaworzno	OKE Kraków	OKE Łódź	OKE Wrocław	Razem
Zadanie 21	63	101	52	53	269
Zadanie 22	45	109	56	50	260
Zadanie 23	56	87	40	32	215
Razem	164	297	148	135	744

### 3. Monitoring procesu oceniania

Oprogramowanie scoris assessor umożliwia prowadzenie nieprzerwanej, bieżącej kontroli pracy zarówno poszczególnych egzaminatorów, jak i Przewodniczących Zespołów Egzaminatorów (PZE). Osoby te podlegają procesowi wstępnej kwalifikacji, a w trakcie oceniania system przesyła egzaminatorom losowo określony procent zadań kontrolnych, których poprawne ocenienie warunkuje dalsze wykonywanie pracy. Oprogramowanie umożliwiło weryfikację standardu oceniania nie tylko na poziomie pojedynczego egzaminatora, poszczególnych komisji okręgowych, ale także w skali połowy kraju, co nie jest bez znaczenia dla porównywalności wyników i obiektywnej oceny jakości systemu egzaminacyjnego.

## **Weryfikacja jakości oceniania**

Dostępne w aplikacji scoris narzędzia statystyczne i raporty z zakresu kontroli jakości oceniania pozwoliły na zebranie wielu informacji o porównywalności pracy egzaminatorów, a zwłaszcza popełnianych przez nich błędów.

Jak pisze Popham, są trzy potencjalne źródła błędów, które powinny być monitorowane podczas oceniania: schemat oceniania, procedura oceniania i oceniający (egzaminator).

Różnice między procedurami stosowanymi w bieżącym roku i w latach poprzednich dostarczają interesujących materiałów, m.in. do weryfikacji wątpliwości:

1. Czy różnice w wynikach egzaminów w różnych regionach kraju rzeczywiście odzwierciedlają różnice w poziomie osiągnięć uczniów, a nie wynikają np. z odmiennej interpretacji zasad oceniania w różnych OKE?
2. Czy da się ograniczyć subiektywizm oceniania egzaminatorów, biorąc pod uwagę fakt, że każdy z nich ma inne doświadczenia zawodowe?

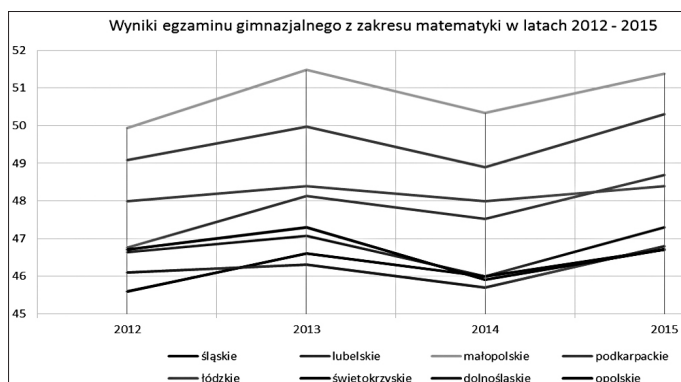
Można założyć, że jeśli istniał jakiś istotny wpływ komisji okręgowych na sposób oceniania prac, a w 2015 r. został on zniwelowany, to zmieni się pozycja wyników poszczególnych województw.

W celu weryfikacji tych założeń porównamy wyniki egzaminu gimnazjalnego w poszczególnych regionach w latach 2012-2014, czyli w latach, gdy każda OKE oceniała prace swoich uczniów niezależnie, z rokiem 2015, w którym prace były losowo oceniane przez egzaminatorów wszystkich komisji biorących udział w projekcie. Opierając się na tym zestawieniu, ustalimy, czy wspólne ocenianie prac zmieniło proporcje między średnimi wynikami uzyskanymi w poszczególnych regionach. Wynik pozytywny świadczyłby o występowaniu w ubiegłych latach efektu egzaminatora na poziomie całych OKE.

E-ocenianie pozwala na dokładniejszą analizę procesu oceniania prac przez egzaminatorów niż ocenianie tradycyjne. Możemy porównać częstość popełniania błędów (w tzw. zadaniach kontrolnych) przez egzaminatorów z poszczególnych OKE. Celem tej analizy jest uzyskanie dokładniejszych informacji o jakości szkolenia egzaminatorów z regionów różnych OKE i o ich ewentualnych tendencjach do zawyżania bądź zaniżania ocen za poszczególne zadania.

## **Wyniki egzaminu gimnazjalnego z zakresu matematyki w latach 2012-2015**

Co roku wyniki egzaminów budzą duże emocje, głównie za sprawą różnic w wynikach uczniów pomiędzy poszczególnymi województwami. Przyjrzyjmy się więc wynikom egzaminu gimnazjalnego w latach 2012-2015 w ośmiu województwach czterech komisji egzaminacyjnych: Jaworzno, Kraków, Łódź, Wrocław.



**Wykres 1. Wyniki egzaminu gimnazjalnego z zakresu matematyki 2012-2015 w ośmiu województwach (dla 2015 roku pierwsza linia od góry małopolskie, kolejne: podkarpackie, lubelskie, łódzkie, świętokrzyskie, śląskie, dolnośląskie i opolskie)**

Jak widać, w kolejnych latach, niezależnie od trudności zadań, różnice w wynikach pomiędzy poszczególnymi województwami są bardzo podobne. Najwyższe wyniki osiągają uczniowie z województw małopolskiego i podkarpackiego, a najniższe uczniowie z województw świętokrzyskiego i dolnośląskiego. Wynik e-oceniania z 2015 r. nie zmienił tej pozycji, co oznacza, że to, skąd są egzaminatorzy oceniający prace uczniów, nie ma istotnego wpływu na średni wynik danego województwa. Przyjrzyjmy się bliżej wynikom uczniów z tych województw w bieżącym roku, zwłaszcza w zakresie zadań otwartych, czyli zadań ocenianych przez egzaminatorów.

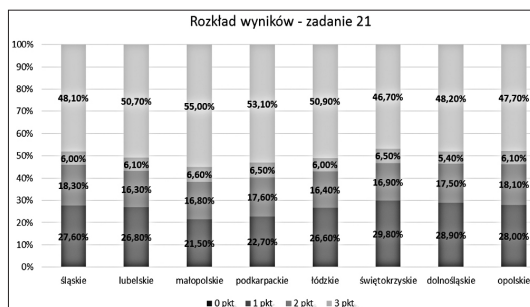
**Tabela 2. Wykonanie zadań z matematyki w 2015 roku w czterech okręgowych komisjach egzaminacyjnych**

Okręgowe Komisje Egzaminacyjne	Województwo	Zadanie 21	Zadanie 22	Zadanie 23
		3 pkt	2 pkt	4 pkt
Wyniki w % punktów				
OKE w Jaworznie	śląskie	58	28	23
	lubelskie	60	30	26
OKE w Krakowie	małopolskie	65	34	29
	podkarpackie	63	32	27
OKE w Łodzi	łódzkie	60	29	26
	świętokrzyskie	57	25	23
OKE we Wrocławiu	dolnośląskie	58	28	23
	opolskie	58	28	23

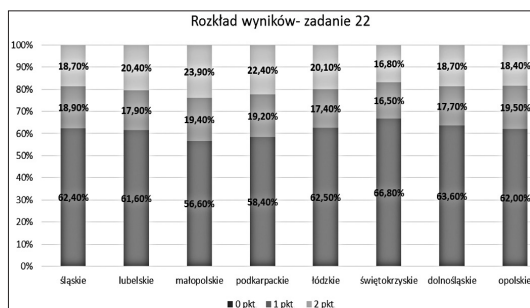
Podobnie jak w całym teście, najwyższe wyniki za poszczególne zadania osiągnęli uczniowie z województw małopolskiego i podkarpackiego. To pokazuje, że miejsce skąd pochodzi egzaminator nie ma istotnego wpływu na wyniki uczniów. Potwierdza to także szczegółowy rozkład liczby uzyskanych punktów w poszczególnych zadaniach oraz bardzo wysoki współczynnik korelacji między liczbą punktów uzyskaną za zadania zamknięte i otwarte.

Tabela 3. Współczynniki korelacji

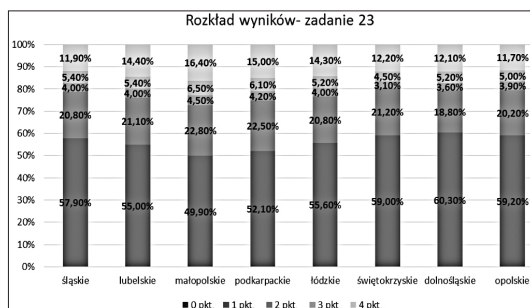
	Współczynnik korelacji między zadaniami zamkniętych i otwartymi
Kraj	0,735
OKE Wrocław	0,739
OKE Łódź	0,721
OKE Kraków	0,730
OKE Jaworzno	0,737



Wykres 2. Procentowy rozkład punktów za zadanie 21 (zdający mógł otrzymać maksymalnie 3 punkty)



Wykres 3. Procentowy rozkład punktów za zadanie 22 (zdający mógł otrzymać maksymalnie 2 punkty)



Wykres 4. Procentowy rozkład punktów za zadanie 23 (zdający mógł otrzymać maksymalnie 4 punkty)

## E-ocenie a sposoby zapewniania rzetelności i porównywalności oceniania w skali kraju

W 2015 r. dążono do tego, aby procedury stosowane podczas procesu oceniania były maksymalnie ujednolicone we wszystkich komisjach.

### 1. Szkolenie online

W celu przygotowania egzaminatorów do oceniania na uruchomionej dla czterech komisji platformie Moodle zorganizowano trzy cykle szkoleniowe, podczas których egzaminatorzy wszystkich komisji oceniali te same rozwiązania zadań. Rozwiązania były wybrane i ocenione przez koordynatorów z wszystkich komisji. Szkolenia te pozwoliły także na weryfikację umiejętności egzaminatorów i wybór do sesji najlepiej przygotowanych osób.

### 2. Szkolenie stacjonarne

W ramach e-ocenia egzaminatorzy z czterech OKE po raz pierwszy odbyli jednakowe szkolenie, w trakcie którego prezentowany był film oraz wspólna dla wszystkich OKE prezentacja dotycząca stosowania kryteriów oceniania.

### 3. Sesja kwalifikacyjna

Każdy egzaminator przed rozpoczęciem pracy musiał zakwalifikować się do sesji. W tym celu ocenił 6 zadań, które posiadały już ocenę ekspercką. Każdy błąd w ocenie skutkował koniecznością kontaktu z PZE i weryfikacją poprawności oceniania.

### 4. System kontroli jakości

Aplikacja scoris zawiera wbudowany system monitorowania jakości oceniania. Zadaniem tego systemu jest dostarczanie egzaminatorom losowo zadań kontrolnych, w liczbie stanowiącej około 6% liczby sprawdzanych prac. W tym celu krajowy zespół koordynatorów przygotował po 80 zadań kontrolnych dla każdego z 3 zadań stanowiących punkt odniesienia (wzorcową ocenę) do wyników oceny egzaminatorów. Rozwiązania te posłużyły do weryfikacji poprawności oceniania przez egzaminatorów. Egzaminator był zatrzymywany zawsze, kiedy w dwóch kolejnych zadaniach kontrolnych, które otrzymywał, przyznawał inną liczbę punktów niż eksperci.

### 5. Ujednolicona procedura

Oceniania zadaniami, umożliwiło egzaminatorom przejście od punktowania całych prac do specjalizacji w ocenianiu poszczególnych zadań.

**Tabela 4. Liczba zadań kontrolnych przydzielonych egzaminatorom w zależności od liczby ocenionych zadań**

	Rozpiętość liczby zadań kontrolnych dla jednego egzaminatora	
	od	do
Zadanie 21	36	47
Zadanie 22	38	47
Zadanie 23	33	44

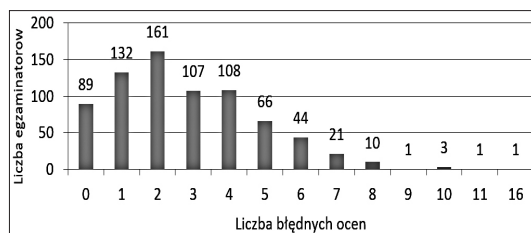
## Analiza błędów popełnianych przez egzaminatorów

System monitorowania jakości oceniania dostarczył nowych informacji, dzięki którym mamy możliwość analizy błędów popełnianych przez egzaminatorów w trakcie oceniania. Dotyczyły one głównie dokładności punktowania – łagodności bądź surowości.

Dane te posłużą nam do porównania trudności w ocenianiu poszczególnych zadań w różnych komisjach.

**Tabela 5. Zgodność oceniania zadań kontrolnych**

	Zadanie 21	Zadanie 22	Zadanie 23
Liczba egzaminatorów	269	260	215
Liczba egzaminatorów, którzy niepoprawnie ocenili co najmniej 1 zadanie	237	232	186
Odsetek egzaminatorów, którzy niepoprawnie ocenili co najmniej 1 zadanie	88,10%	89,23%	86,51%
Liczba zastosowanych zadań kontrolnych w trakcie przy egzaminatorów	11 176	10 383	8 411
Liczba poprawnie ocenionych zadań kontrolnych	10 112	9 686	8 079
Odsetek poprawnie ocenionych zadań kontrolnych	90,48%	93,29%	96,05%
Liczba niepoprawnie ocenionych zadań kontrolnych	1 063	697	332
Odsetek niepoprawnie ocenionych zadań kontrolnych	9,51%	6,71%	3,95%



**Wykres 5. Liczba zadań kontrolnych ocenionych niepoprawnie (łącznie liczba zadań kontrolnych: 29 970)**

## Porównanie pracy egzaminatorów z różnych OKE

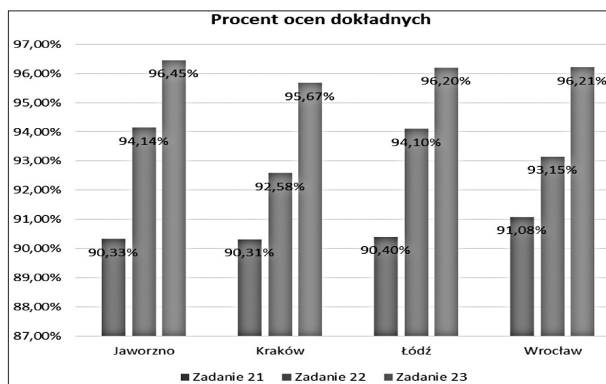
Aby zapewnić wysoką jakość i zaufanie do egzaminów zewnętrznych, prace egzaminacyjne powinny być oceniane spójnie w całym kraju. W tym celu krajowy zespół koordynatorów opracowuje kryteria oceniania, na podstawie których są szkoleni egzaminatorzy w całym kraju. Ewentualne rozbieżności w ocenianiu mogą pojawić się z różnych powodów. Dotychczas szkolenia egzaminatorów były realizowane samodzielnie w ramach każdej komisji egzaminacyjnej. Każda z komisji dysponowała tym samym materiałem szkoleniowym wypracowanym wspólnie z Centralną Komisją Egzaminacyjną bezpośrednio po egzaminie.

W tym roku po raz pierwszy wszyscy egzaminatorzy czterech komisji egzaminacyjnych obejrżeli ten sam film szkoleniowy, w trakcie którego omówiono schematy oceniania wraz z próbą ich zastosowania do oceny autentycznych zadań egzaminacyjnych rozwiązanych przez uczniów. Wyłoniono trzech koordynatorów zadań spośród egzaminatorów różnych komisji egzaminacyjnych. Po szkoleniu bezpośrednim (w zespołach) zmieniono zasadniczo warunki pracy egzaminatorów (praca w mieszkaniu na własnym komputerze). Tym samym zwiększyła się trudność polegająca na indywidualnej odpowiedzialności za ocenę zadania. Mając wątpliwość w ocenie zadania egzaminatorzy, musieli korzystać z pomocy przewodniczącego zespołu, a nie koleżanki czy kolegi siedzącego obok.

W naszym artykule skoncentrujemy się na dwóch istotnych aspektach. **Po pierwsze**, zbadamy, czy i jak duże są różnice w wynikach spowodowane odmienną interpretacją schematów omawianych przez koordynatorów z poszczególnych OKE. **Po drugie**, jak na różnice w wynikach może wpływać sposób przekazania schematów oceniania na szkoleniach egzaminatorów, co może prowadzić do bardziej „surowego” bądź „łagodnego” oceniania prac. Aby zbadać wyżej wymienione aspekty, naturalne jest porównanie zgodności punktowania zadań przez egzaminatorów z różnych OKE.

Do wykonania tego zadania wykorzystano dane dotyczące oceniania zadań kontrolnych przez egzaminatorów w trakcie całego procesu oceniania.

Na wykresie 6 prezentujemy odsetek poprawnie ocenionych zadań kontrolnych przez egzaminatorów czterech komisji egzaminacyjnych z podziałem na poszczególne zadania.

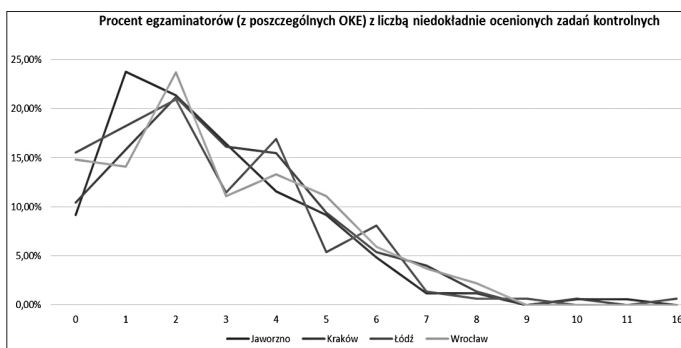


**Wykres 6. Zgodność (dokładność) oceniania zadań kontrolnych przez egzaminatorów**

Powyższy wykres pokazuje, że w każdym z zadań, jakość oceniania egzaminatorów z poszczególnych OKE była podobna.

Porównajmy więc odsetek egzaminatorów z poszczególnych OKE, którzy popełnili odpowiednio 1 błąd, 2 błędy itd. w ocenie rozwiązań kontrolnych. Jak widać na wykresie 7, występuje tutaj dość duże zróżnicowanie. Niemniej w każdej komisji około 80% egzaminatorów błędnie oceniło nie więcej niż 4 rozwiązania kontrolne.





**Wykres 7. Odsetek egzaminatorów według liczby błędnie ocenionych zadań kontrolnych**

Dane z powyższych dwóch wykresów nie dostarczają jednak informacji o zakresie popełnianych błędów, tzn. jak często ocena egzaminatora była zawyżona, a jak często była zaniżona oraz jak duże było odchylenie oceny egzaminatora od oceny wzorcowej. Aby dowiedzieć się, czy istnieje u egzaminatorów z poszczególnych OKE tendencja do bycia „surowym” czy „łagodnym” w punktowaniu, porównamy wyniki oceniania zadań przez egzaminatorów z różnych komisji.

### Porównanie wyników oceniania różnych (typów) zadań

W analizie błędów skupimy się wyłącznie na niepoprawnych ocenach egzaminatorów (przypomnijmy, że ponad 90% ocen egzaminatorów z każdej OKE było poprawnych).

W poniższej tabeli prezentujemy średnią wartość różnicy między oceną egzaminatorów a oceną sędziów kompetentnych przyjętą jako wzorcową.

**Tabela 6. Różnica między średnią oceną egzaminatorów a oceną sędziów kompetentnych (wzorcową)**

Okręgowe Komisje Egzaminacyjne	Zadanie 21	Zadanie 22	Zadanie 23
OKE Wrocław	-0,020	-0,088	-0,308
OKE Łódź	-0,122	-0,023	-0,683
OKE Kraków	-0,108	-0,043	-0,592
OKE Jaworzno	-0,163	-0,096	-0,628

Wynika z niej, że w ocenianiu występuje nieco większa tendencja do zaniżania oceny niż do jej zawyżania. Niemniej, różnice nie są wielkie. Zauważmy też, że różnice między poszczególnymi komisjami również nie są znaczące. Największą rozpiętość w ocenie widzimy w zadaniu 23. Należy jednak podkreślić, że uwzględniając poprawne oceny egzaminatorów (około 90% w każdej OKE), odchylenia te są znikome.

Przejdźmy do analizy wyników oceniania zadań kontrolnych.

### Zadanie 21

Przyjrzyjmy się danym w tabeli 7. Prezentujemy w niej dane ilustrujące oceny egzaminatorów w zadaniu 21. Wartości w poszczególnych kolumnach odnoszą się do wielkości błędów (różnica między oceną wzorcową sędziów kompetentnych i oceną egzaminatorów).

**Tabela 7. Różnica między średnią oceną egzaminatorów a oceną sędziów kompetentnych (wzorcową) w zadaniu 21**

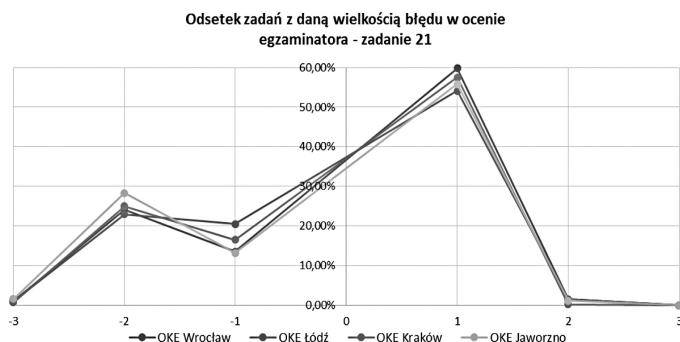
Okręgowe Komisje Egzaminacyjne	Średnia wartość bezwzględna	Odchylenie standardowe	Średnie odchylenie bezwzględne
OKE Wrocław	1,276	1,849	1,281
OKE Łódź	1,248	1,795	1,277
OKE Kraków	1,264	1,801	1,281
OKE Jaworzno	1,326	1,981	1,348

Powyższe wyniki nie wskazują na istotne różnice w poprawności stosowania schematów punktowania przez egzaminatorów z poszczególnych OKE. Uwzględniając fakt, że za zadanie można było zdobyć 3 punkty oraz to, że w naszych zestawieniu nie uwzględniamy poprawnie ocenionych zadań kontrolnych, średnie odchylenie i odchylenie standardowe nie są duże.

**Tabela 8. Różnica między średnią oceną egzaminatorów a oceną sędziów kompetentnych (wzorcową) w zadaniu 21**

Błąd w ocenie	Odsetek zadań z daną wielkością błędu w ocenie egzaminatora – zadanie 21			
	OKE Wrocław	OKE Łódź	OKE Kraków	OKE Jaworzno
-3	1,01%	0,98%	0,75%	1,55%
-2	24,12%	22,93%	25,00%	28,29%
-1	13,57%	20,49%	16,50%	13,18%
1	59,80%	54,15%	57,50%	55,81%
2	1,51%	1,46%	0,25%	1,16%
3	0,00%	0,00%	0,00%	0,00%

W zadaniu 21 niespełna 10% zadań kontrolnych zostało niepoprawnie ocenione przez egzaminatorów. Średnia wartość błędu nie jest jednak duża, a skrajne wartości ( $\pm 3$  punkty odchylenia od oceny wzorcowej) prawie się nie zdarzały. Pomimo, że średni błąd egzaminatora jest ujemny (dla każdej OKE), w ponad 55% niepoprawnych ocen błąd popełniano na korzyść zdającego, zazwyczaj o +1 punkt. Dane z powyższej tabeli zostały zilustrowane na wykresie 8.



**Wykres 8. Rozkład błędów w ocenie zadań kontrolnych zadania 21**

Powyższe wyniki wskazują, iż egzaminatorzy w miarę dobrze poradzi sobie z oceną tego zadania. Podobnie, dość dobrze poradzi sobie uczniowie z rozwiązaniem tego zadania (57% – 65%).

### Zadanie 22

W tabeli 9 prezentujemy dane statystyczne dotyczące błędnych ocen egzaminatorów w zadaniu 22. Tak jak wcześniej wartości w poszczególnych kolumnach odnoszą się do wielkości błędów w zadaniach kontrolnych niepoprawnie ocenionych.

**Tabela 9. Różnica między średnią oceną egzaminatorów a oceną sędziów kompetentnych (wzorcową) w zadaniu 22**

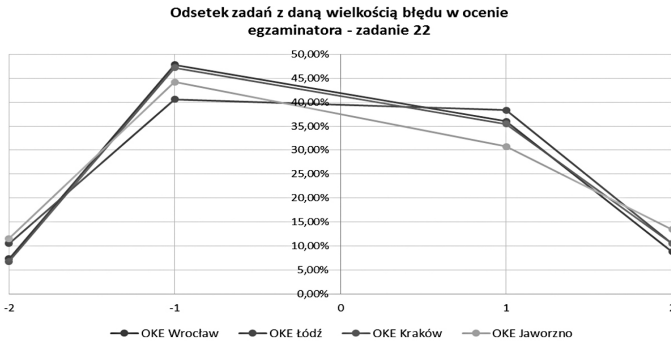
Okręgowe Komisje Egzaminacyjne	Średnia wartość bezwzględna	Odchylenie standardowe	Średnie odchylenie bezwzględne
OKE Wrocław	1,162	1,478	1,153
OKE Łódź	1,211	1,631	1,210
OKE Kraków	1,173	1,517	1,169
OKE Jaworzno	1,250	1,741	1,239

**Tabela 10. Różnica między średnią oceną egzaminatorów a oceną sędziów kompetentnych (wzorcową) w zadaniu 22**

Błąd w ocenie	Odsetek zadań z daną wielkością błędów w ocenie egzaminatora – zadanie 22			
	OKE Wrocław	OKE Łódź	OKE Kraków	OKE Jaworzno
-2	7,35%	10,53%	6,79%	11,54%
-1	47,79%	40,60%	47,22%	44,23%
1	36,03%	38,35%	35,49%	30,77%
2	8,82%	10,53%	10,49%	13,46%

Za zadanie 22 uczeń mógł otrzymać maksymalnie 2 punkty. Zatem wysokie wartości odchylenia standardowego (w każdej OKE), wskazują, że choć błędnych ocen nie było dużo (6,71%), to często występowały błędy skrajne w ocenie ( $\pm 2$  punkty). Szczególnie widać to w porównaniu z pozostałymi zadaniami,

gdzie odsetek ocen o maksymalnym błędzie jest niewielki (kilkakrotnie mniejszy niż w zadaniu 22). Sądzymy, że świadczy to o problemach egzaminatorów z ocenianiem tego typu zadania (dowód matematyczny). Wynika z tego potrzeba wyłaniania w trakcie szkoleń e-learningowych takich egzaminatorów, którzy lepiej radzą sobie z oceną tego typu zadań



**Wykres 9. Rozkład błędów w ocenie zadań kontrolnych zadania 22**

W wyniku analizy częstości błędów popełnianych przez poszczególnych egzaminatorów i kierunku ich popełniania (zawyżanie lub zaniżanie punktacji) nie stwierdzono wyraźnych tendencji ani u pojedynczych egzaminatorów ani u grup egzaminatorów jednej czy dwóch komisji egzaminacyjnych.

### Zadanie 23

W poniższej tabeli 11 prezentujemy dane dotyczące błędnych ocen zadań kontrolnych w zadaniu 23, analogiczne do zestawień dla zadań poprzednich.

**Tabela 11. Różnica między średnią oceną egzaminatorów a oceną sędziów kompetentnych (wzorcową) w zadaniu 23.**

Okręgowe Komisje Egzaminacyjne	Średnia wartość bezwzględna	Odchylenie standardowe	Średnie odchylenie bezwzględne
OKE Wrocław	0,962	0,982	0,891
OKE Łódź	1,017	0,583	0,812
OKE Kraków	1,054	0,827	0,881
OKE Jaworzno	1,038	0,721	0,849

Pomimo tego, że za zadanie to można było uzyskać najwięcej punktów (4 punkty), zarówno błędy popełniane przez egzaminatorów, jak i rozproszenie ocen (odchylenie standardowe) są dużo mniejsze niż w poprzednich zadaniach. W zadaniu 23. stwierdzono największy odsetek poprawnych ocen zadań kontrolnych. Tak wysoka poprawność oceniania tego zadania może wynikać z różnych powodów. Może to być na przykład:

- dobra znajomość tematyki zadania (duża praktyka w nauczaniu tego typu zagadnień),
- wysoka skuteczność szkolenia egzaminatorów,
- bardzo dobrze przygotowany schemat oceniania.

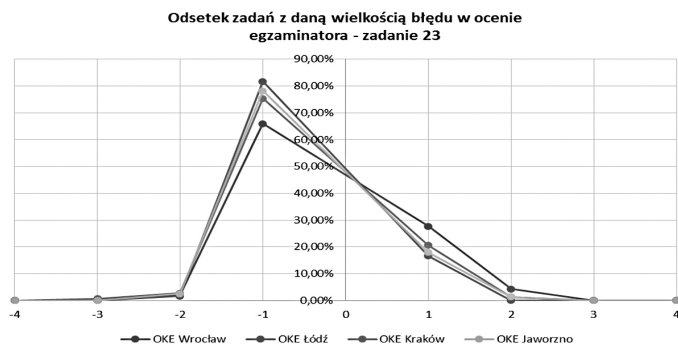
Analizowane tutaj dane nie są jednak wystarczające do jednoznacznego ustalenia przyczyn tak dobrego wyniku. Sposób zapisania schematu oceniania w tym zadaniu warto wykorzystać jako dobry przykład możliwości zachowania porównywalności oceniania przez wielu egzaminatorów mimo różnych metod rozwiązywania tego zadania przez uczniów.

Warto jednak zwrócić uwagę, że zadania 21 i 22 mogły być rozwiązywane przez uczniów na wiele sposobów. Ta „wielość” metod rozwiązania obu zadań oraz „zakamuflowany” sposób rozumowania uczniów mógł powodować trudności w analizie rozwiązania. W przypadku zadania 23 jego rozwiązanie miało wyraźnie odrębne etapy (sposób obliczenia  $h$ , sposób obliczenia  $r$ , sposób obliczenia  $V$  i ostatni – poprawny wynik), zatem określenie reguł przydzielania punktów było stosunkowo łatwe.

W poniższej tabeli widzimy, że ponad 90% błędnych ocen różni się od oceny wzorcowej o +1 lub -1 punkt (co stanowi tylko 25% maksymalnej liczby punktów, zaś w zadaniu 21 było to 33%). Jednak, w przeciwieństwie do zadania 21, przeważają oceny niższe niż ustalone przez ekspertów. Zauważamy, że skłonność do zawyżania bądź zaniżania ocen w dużej mierze zależy od typu zadania. Nie zauważono istotnych różnic w surowości bądź łagodności ocen egzaminatorów z czterech komisji egzaminacyjnych.

**Tabela 12. Różnica między średnią oceną egzaminatorów a oceną sędziów kompetentnych (wzorcową) w zadaniu 23**

Błąd w ocenie	Odsetek zadań z daną wielkością błędu w ocenie egzaminatora – zadanie 23			
	OKE Wrocław	OKE Łódź	OKE Kraków	OKE Jaworzno
-4	0,00%	0,00%	0,00%	0,00%
-3	0,00%	0,00%	0,68%	0,00%
-2	2,13%	1,67%	2,74%	2,56%
-1	65,96%	81,67%	75,34%	78,21%
1	27,66%	16,67%	20,55%	17,95%
2	4,26%	0,00%	1,37%	1,28%
3	0,00%	0,00%	0,00%	0,00%
4	0,00%	0,00%	0,00%	0,00%



**Wykres 10. Rozkład błędów w ocenie zadań kontrolnych zadania 23**

## Wnioski

Przedstawione wyniki wskazują na brak istotnych różnic w jakości pracy egzaminatorów różnych komisji egzaminacyjnych. Tezę tę potwierdzamy dwoma argumentami. Po pierwsze, średnie wyniki uzyskane przez uczniów objętych działaniem komisji egzaminacyjnych występujące w latach poprzednich potwierdziły się (np. najlepsze wyniki w OKE Kraków, najłabsze w OKE Wrocław).

Także względne pozycje ośmiu województw ustalone na podstawie średnich wyników uczniów nie zmieniły się w stosunku do lat wcześniejszych. Po drugie, nie zauważono istotnych różnic w surowości ocen egzaminatorów w czterech komisjach egzaminacyjnych, na co wskazuje analiza rozkładu błędów popełnianych przez egzaminatorów w ocenie poszczególnych typów zadań. Widoczne niewielkie odchylenia utrzymują się w granicach błędu statystycznego.

Kluczowe znaczenie dla uzyskania jak najrzetelniejszych wyników pomiaru umiejętności uczniów ma konstrukcja schematu oceniania. Ustrukturyzowany schemat oceniania ułatwia uzyskanie wysokiej porównywalności oceniania zadania przez wielu egzaminatorów. Jednocześnie widac, że dobrym kierunkiem dla zapewnienia porównywalności oceniania są ogólnopolskie szkolenia uczące wspólnego rozumienia schematu oceniania oraz jednolite procedury oceniania. Egzaminatorzy wysoko ocenili możliwość oceniania jednego zadania podczas całej sesji, co przyspiesza wykonanie pracy dzięki zauważeniu powtarzalności stosowanych metod rozwiązania zadania i niweluje liczbę błędów popełnianych przez egzaminatorów. Kluczowym zadaniem każdej komisji jest dobór najlepszych egzaminatorów zarówno pod względem merytorycznym, jak i sprawności posługiwania się nową technologią IT.

## Bibliografia

1. Popham W. J., *Modern Educational Measurement [w:] A practitioner's perspective*. Englewood Cliffs, New York, Prentice Hall, 1990, za: Dubiecka A., Szalencik H., Węziak D., *Efekt egzaminatora*, Warszawa 2006.