

**Elżbieta Modrzewska**

Okręgowa Komisja Egzaminacyjna w Jaworznie

**Hanna Wylęzek**

Okręgowa Komisja Egzaminacyjna w Jaworznie

## **Czy gimnazjaliści rzeczywiście coraz gorzej piszą? Obserwacje na podstawie analizy prac egzaminacyjnych i wyników egzaminu gimnazjalnego w latach 2009 i 2011**

Egzaminy zewnętrzne są swoistym podsumowaniem efektów kształcenia po kolejnych etapach edukacji. Dlatego nie dziwi fakt, że ich wyniki stają się coraz częściej przedmiotem dyskusji publicznej.

Analizy statystyczne wyników egzaminów zewnętrznych powinny umożliwiać prowadzenie obserwacji, jak kształtują się osiągnięcia uczniów oraz na jakim poziomie opanowują oni poszczególne umiejętności. Jest to poważne wyzwanie, gdyż wnioski wyprowadzane z analiz mogą służyć nie tylko do oceny pracy szkół i nauczycieli, ewaluacji programów nauczania czy systemu kształcenia przyszłych nauczycieli, ale także do podejmowania decyzji o kształceniu edukacji na poziomie samorządów i całego kraju.

Analiza wyników części humanistycznej egzaminu gimnazjalnego w województwie śląskim w 2011 roku wzbudziła niepokój środowisk zainteresowanych edukacją. Średni wynik egzaminu 25,41 oraz łatwość zestawu 0,51 okazały się w porównaniu z wynikami z lat ubiegłych wyraźnie niższe.

**Tabela 1. Wyniki średnie i wskaźniki łatwości zestawów części humanistycznej egzaminu gimnazjalnego w województwie śląskim w latach 2002-2011**

Wskaźniki	Rok				
	2002	2003	2004	2005	2006
Wynik średni	31	32,8	28,1	33,6	31,7
Wskaźnik łatwości zestawu	0,62	0,66	0,56	0,67	0,63
Wskaźniki	Rok				
	2007	2008	2009	2010	2011
Wynik średni	31,8	31,1	31,9	30,9	25,4
Wskaźnik łatwości zestawu	0,64	0,62	0,64	0,62	0,51

Również dziennikarze bezpośrednio po ogłoszeniu wyników zadawali pytania o przyczynę takiego stanu rzeczy. Podjęli w mediach dyskusję na temat niskich osiągnięć uczniów, zastanawiali się, dlaczego gimnazjaliści tak słabo napisali część humanistyczną egzaminu, wyrokowali, że wyniki nie nastroją optymistycznie.

Te i podobne stwierdzenia środowisk dziennikarskich zdecydowanie nie budują dobrego klimatu wokół gimnazjów, które i tak nie cieszą się najlepszą opinią.

Wydaje się, że niższy wynik jest potwierdzeniem tego, że absolwenci gimnazjów umieją coraz mniej.

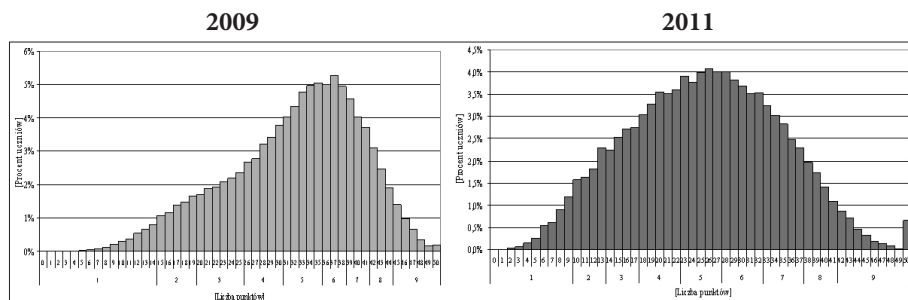
Wobec tak niesprzyjającego klimatu wokół niskich wyników tegorocznych absolwentów gimnazjów rodzi się potrzeba przeprowadzenia szczegółowych, rzetelnych analiz, które pozwolą zbadać ten problem, wskazać przyczyny i podjąć działania wspierające dyrektorów szkół i nauczycieli, aby mogli zbudować skuteczne programy naprawcze oraz zaplanować pracę dydaktyczną.

## W poszukiwaniu przyczyn niższych wyników

Aby ustalić przyczyny niższych wyników, przeprowadzono badanie porównawcze poziomu osiągnięć uczniów w zakresie umiejętności tworzenia własnego tekstu na przykładzie zadań rozszerzonej odpowiedzi. W tym celu wybrano do badania arkusze z 2011 i 2009 roku, gdyż w obydwu zestawach egzaminacyjnych sprawdzano poziom opanowania umiejętności uczniów w zakresie pisania niemal identycznymi zadaniami. Sprawdzały one umiejętność tworzenia tych samych form wypowiedzi – podania i rozprawki. Także podobnie były sformułowane tematy obydwu zadań, a w kartotekach wskazano te same umiejętności.

Stąd właśnie za kontekst analizy wyników z 2011 roku przyjęto wyniki z arkusza zastosowanego w 2009 roku.

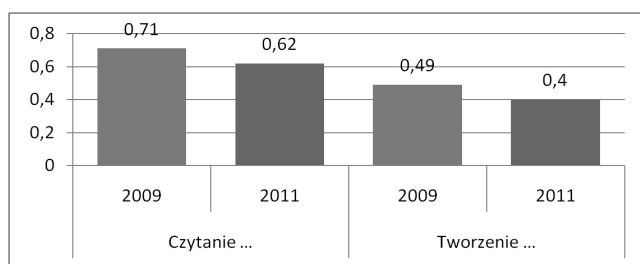
Już najprostsze porównania pokazały, że mamy do czynienia z nową sytuacją. Dotychczas rozkłady wyników z części humanistycznej egzaminu gimnazjalnego były przesunięte w kierunku wyników wysokich. Rozkład z 2011 roku ma wyraźnie inny kształt, jest niemal zbliżony do rozkładu normalnego.



**Wykres 1. Rozkłady wyników części humanistycznej egzaminu gimnazjalnego w województwie śląskim w 2009 i 2011 roku**

Analiza kształtu rozkładów sugeruje, że w 2011 roku uczniowie mieli o wiele większe problemy z rozwiązaniem zadań niż uczniowie w 2009 roku.

O obniżeniu poziomu umiejętności uczniów można by wnioskować również, porównując wskaźniki na przykład łatwości standardów wymagań egzaminacyjnych.



**Wykres 2. Wyniki uczniów województwa śląskiego w obszarach standardów wymagań egzaminacyjnych w latach 2009 i 2011**

Zestawienie pokazuje, że w 2011 roku wskaźniki łatwości były wyraźnie niższe w obu obszarach standardów. Czy to oznacza, że uczniowie umieją mniej?

Podobne pytanie pojawia się także po analizie porównawczej wskaźników łatwości w obszarze standardu *tworzenie własnego tekstu* w poszczególnych przedziałach staninowych. Skala staninowa dostarcza przecież rzetelnej informacji o osiągnięciach uczniów w kolejnych sesjach egzaminacyjnych. Pozwala zaobserwować, jaki jest wskaźnik łatwości badanego standardu wśród uczniów, którzy uzyskali wyniki w poszczególnych przedziałach. Pozwala również określić odsetek tych, którzy osiągnęli poziom zadowalający (od 70% opanowania umiejętności).

**Tabela 2. Wskaźnik łatwości zestawu egzaminacyjnego i obszaru standardu *tworzenie własnego tekstu* w poszczególnych przedziałach staninowych w województwie śląskim**

2009									
Opis wyniku	Najniższy	Bardzo niski	Niski	Niżej średniego	Średni	Wyżej średniego	Wysoki	Bardzo wysoki	Najwyższy
Skala staninowa	1	2	3	4	5	6	7	8	9
Zestaw	0,24	0,34	0,45	0,56	0,66	0,74	0,80	0,86	0,93
Tworzenie ... (II)	0,07	0,17	0,31	0,47	0,58	0,66	0,73	0,80	0,90
2011									
Zestaw	0,15	0,23	0,31	0,40	0,50	0,60	0,70	0,79	0,90
Tworzenie ... (II)	0,04	0,10	0,18	0,28	0,39	0,49	0,60	0,71	0,87

Z tabeli wynika, że uczniowie z wynikiem od 1. do 6. stanina, tj. 77%, osiągnęli niski poziom umiejętności z zakresu *tworzenie własnego tekstu*. Poziom zadowalający osiągnęli ci gimnazjaliści, których wyniki znalazły się w przedziałach od 7. do 9. stanina w 2009 roku, natomiast w 2011 roku dopiero uczniowie z wynikami w 8. i 9. staninie. To różnica 12% piszących.

## **Dlaczego gimnazjaliści w 2011 roku osiągnęli niższe wyniki?**

### **Czy rzeczywiście umieli mniej?**

Aby odpowiedzieć na te pytania, autorki artykułu przeprowadziły badanie porównawcze poziomu osiągnięć uczniów w zakresie umiejętności tworzenia własnego tekstu na przykładzie zadań rozszerzonej odpowiedzi – podania i rozprawki.

Celem badania było sprawdzenie:

- jakie byłyby wyniki uczniów z roku 2009, jeśli ich prace sprawdzano by zgodnie z kluczem zastosowanym w roku 2011?
- Jakie wyniki uzyskaliby uczniowie zdający egzamin w roku 2011, gdyby ich prace oceniano zgodnie z kryteriami obowiązującymi w roku 2009?

W konsekwencji analiza miała posłużyć ustaleniu, czy i jak na ocenę osiągnięć uczniów, czyli na ich wynik, wpływają klucze punktowania budowane według różnych strategii oceniania – analitycznej i holistycznej).

### **Obszary wymagań poddawane analizie porównawczej**

Badaniem porównawczym objęto tylko te obszary, które oceniane były inaczej na egzaminach w 2009 i 2011 roku.

W zadaniu 28. – podaniu – powtórnie sprawdzono i dokonano analizy osiągnięć w zakresie realizacji tematu, dostosowania stylu do formy wypowiedzi i przestrzegania poprawności językowej. Z kolei w zadaniu 29. – rozprawce – realizacji tematu, przestrzegania poprawności językowej, dostosowania stylu do formy wypowiedzi (przestrzegania poprawności stylistycznej).

Nie porównywano wyników uczniów z 2009 i 2011 roku w zakresie poprawności ortograficznej i interpunkcyjnej, ponieważ zasady oceniania w obydwu latach były tożsame.

### **Opis próby badawczej**

Badaniem odbywało się na próbie losowej liczącej 260 prac. Poddano obserwacji 130 prac egzaminacyjnych z 2009 roku i 130 prac z roku 2011.

Do badania wykorzystano arkusze z 2009 roku zarchiwizowane w Okręgowej Komisji Egzaminacyjnej w Jaworznie zgodnie z *Instrukcją archiwizacji prac egzaminacyjnych przeznaczonych do badań jakościowych*, opracowaną i zatwierdzoną przez CKE w 2007 roku. Było to 130 prac uczniów, którzy uzyskali na egzaminie co najmniej 20 punktów za cały arkusz (od 3. stanina).

Dla zapewnienia porównywalności materiału badawczego zgodnie z tą instrukcją dokonano wyboru prac z 2011 roku – wybrano 130 prac z wynikiem egzaminacyjnym od 14 punktów (od 3. stanina). Doboru próby losowej dokonano z wykorzystaniem programu SPSS.

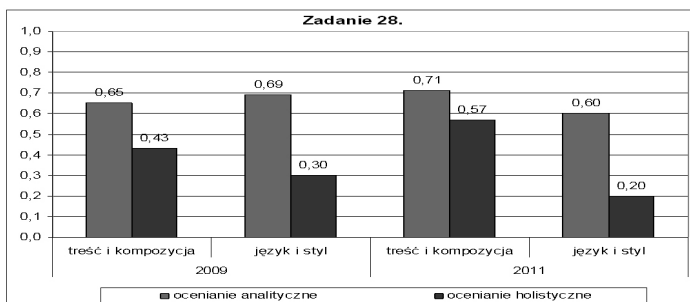
### **Opis narzędzi badawczych i przebiegu badania**

W badaniu zastosowano 4 klucze punktowania: dwa opracowane przez CKE (analityczny z 2009 r. i holistyczny z 2011 r.), wykorzystywane podczas sesji egzaminacyjnych przez egzaminatorów oraz dwa kolejne, opracowane przez autorki (holistyczny dla prac z 2009 r. i analityczny dla prac z 2011 r.) ściśle według zasad obowiązujących w kluczach CKE.

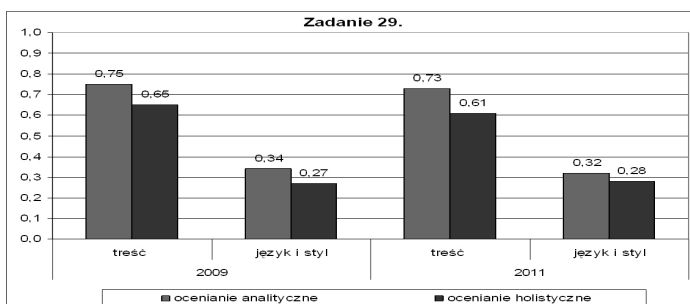
Zadania 28. i 29. znajdujące się w arkuszach egzaminacyjnych z 2009 i 2011 roku zostały powtórnie sprawdzone według kluczy punktowania opracowanych przez autorki. Zadania z arkusza 2009 roku na egzaminie sprawdzane kluczem analitycznym zostały sprawdzone kluczem holistycznym. Natomiast holistycznie oceniane podczas egzaminu prace z 2011 roku powtórnie sprawdzono z wykorzystaniem klucza analitycznego.

### Wyniki badania

Wyniki oceniania zilustrowano na wykresach nr 3 i 4, zestawiając wskaźniki łatwości dla poszczególnych kryteriów po zastosowaniu kryteriów analitycznych i holistycznych.



Wykres 3. Zadanie 28. (podanie) – zestawienie wskaźników łatwości za poszczególne kryteria sprawdzane kluczem analitycznym i holistycznym



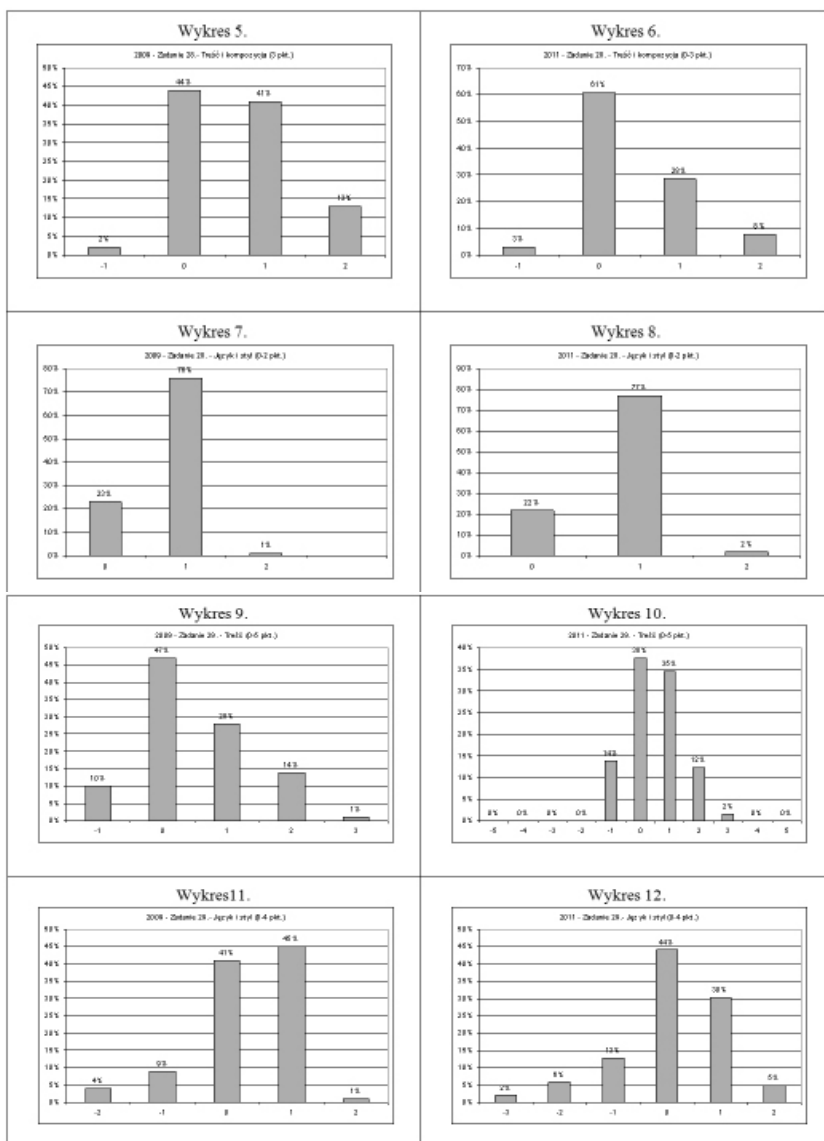
Wykres 4. Zadanie 29. (rozprawka) – zestawienie wskaźników łatwości za poszczególne kryteria sprawdzane kluczem analitycznym i holistycznym

Z zestawienia danych na wykresach wynika, że te same uczniowskie rozwiązania uzyskiwały różną liczbę punktów, najczęściej wyższą przy zastosowaniu kryteriów analitycznych.

Wobec takich obserwacji nasuwają się kolejne pytania:

- Jaki procent uczniów uzyskuje inny wynik przy zmianie zasad oceniania?
- O ile punktów wynik oceniania analitycznego jest wyższy od wyniku ustalonego według kryteriów holistycznych?

Różnice w punktacji wynikające z zastosowania kryteriów analitycznych i holistycznych ilustrują wykresy od 5. do 12.

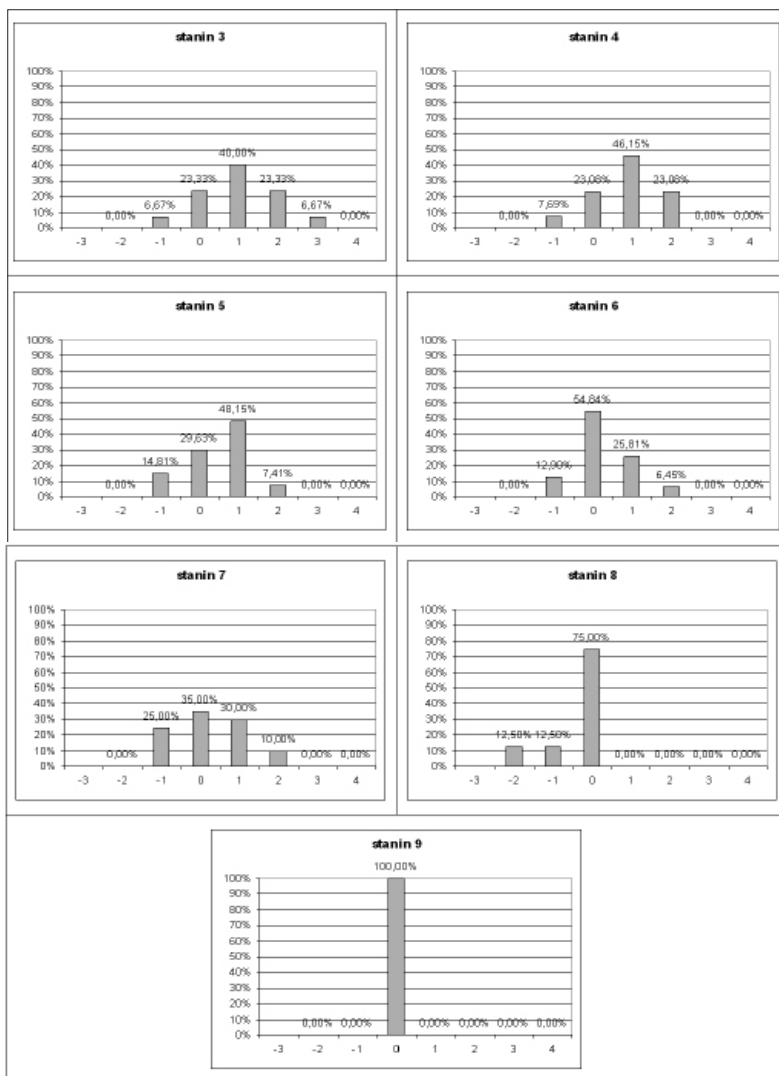


Wykresy 5.-12. Różnice w punktacji wynikające z zastosowania kryteriów analitycznych i holistycznych

Z analizy wynika, że uczniowie uzyskaliby wyższy wynik, gdyby ich prace były sprawdzane kluczem analitycznym – w podaniu najczęściej o dwa punkty: jeden za treść, drugi za język. Podobnie w rozprawce prawie połowa badanych otrzymałaby o 1 punkt więcej za ocenę języka i stylu. Natomiast w obszarze treści około 30% zdających – o jeden punkt więcej, a 14% o dwa punkty; niewielki procent uczniów (1 lub 2) nawet o 3 punkty więcej. Można więc stwierdzić, że

około połowa badanych mogłaby uzyskać za zadania rozszerzonej odpowiedzi średnio 4 punkty więcej.

Istotne byłoby wobec tego stwierdzenie, dla których uczniów – słabszych czy zdolniejszych – zastosowanie kryteriów holistycznych jest bardziej niekorzystne? Aby to zaobserwować, obliczono różnice w punktacji dla uczniów, których wynik znalazł się w staninach od 3. do 9. Wyniki obliczeń ilustrują poniższe wykresy od 13. do 17.



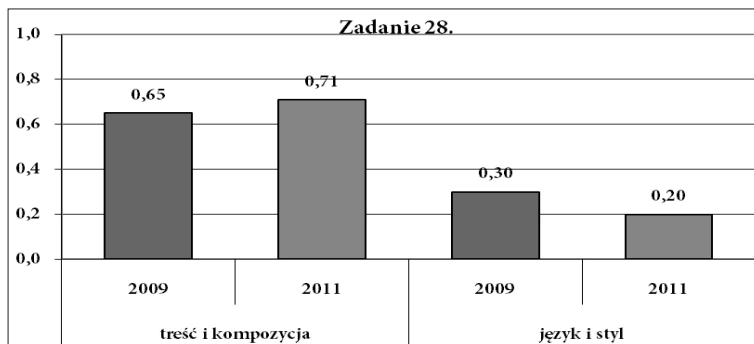
Wykresy 13.–17. Różnice w punktacji wynikające z zastosowania kryteriów analitycznych w pracach z 2011 roku ocenianych holistycznie w przedziałach staninowych od 3. do 9.

Można zauważyć, że zmiana klucza holistycznego na analityczny nie wpłynęłaby na punktację w pracach uczniów z najwyższymi wynikami (stanin 9.). Część uczniów z wynikami z 8. stanina straciłby punkty, natomiast dość wyraźnie zyskaliby uczniowie z wynikami z przedziałów 4., 5. i 6. stanina.

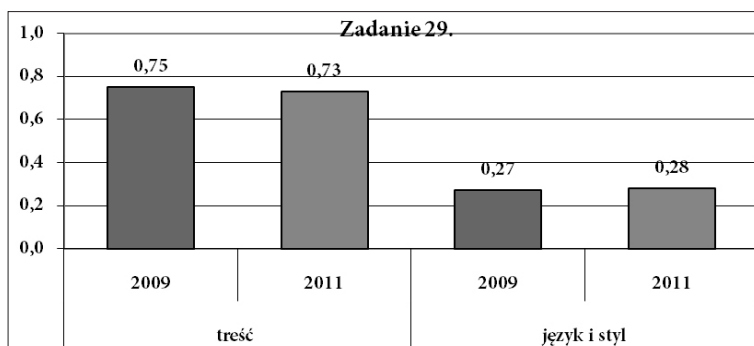
Wydaje się więc, że po zastosowaniu klucza holistycznego na egzaminie gimnazjalnym w 2011 roku wyniki szkół, w których przeważają uczniowie zdolni, niewiele się zmieniły w stosunku do lat ubiegłych, ale w szkołach, w których jest duża liczba uczniów słabszych, wyraźnie się pogorszyły.

Zestawiając wskaźniki łatwości za umiejętności sprawdzane zadaniami 28. i 29. można stwierdzić, że uczniowie piszący egzamin w 2011 roku w rzeczywistości

- opanowali umiejętność redagowania podania w obszarze treści i kompozycji lepiej niż uczniowie piszący egzamin w 2009 roku.
- potrafią tworzyć tekst argumentacyjny na tym samym poziomie co absolwenci gimnazjów z 2009 roku, podobnie radzą sobie z językiem i stylem.
- prezentują niższe umiejętności w obszarze języka w podaniu niż uczniowie w 2009 roku.



Wykres 18. Porównanie współczynników łatwości treści i kompozycji oraz języka i stylu w zadaniu 28. – podaniu ocenianym kluczem analitycznym w latach 2009 i 2011



Wykres 19. Porównanie współczynników łatwości treści oraz języka i stylu w zadaniu 29. – rozprawce ocenianej kluczem analitycznym w latach 2009 i 2011



## Podsumowanie

Przeprowadzone badanie porównawcze i analiza zgromadzonego materiału pozwalają na wyciągnięcie następujących wniosków dotyczących ewaluacji komunikowania wyników oraz prowadzenia badań nad przyrostem wiedzy i umiejętności sprawdzanych na egzaminach zewnętrznych:

- Porównanie wyników egzaminów z różnych lat bez uwzględniania kontekstu wymagań może prowadzić do nieuprawnionych wniosków i nie powinno stanowić podstawy do wydawania sądów ani o poziomie nauczania, ani o przyroście lub spadku poziomu umiejętności gimnazjalistów.
- Dokonując analizy i interpretacji wyników uczniów, a także porównując ich osiągnięcia, należy uwzględniać jeszcze jeden kontekst – narzędzie oceniania.
- Przy porównywaniu wyników egzaminów – nawet przy niewielkiej zmianie zasad oceniania – możliwy jest błąd w komunikowaniu wyników, a tym samym błąd w ocenie osiągnięć uczniów.
- Konieczne jest więc informowanie odbiorców o tym, jak zmieniają się zasady oceniania i że ta zmiana może mieć wpływ na wynik egzaminu.
- Ocenianie holistyczne bardziej różnicuje populację uczniów przystępujących do egzaminu niż ocenianie analityczne.
- Przy zastosowaniu kryteriów holistycznych informacja zwrotna o osiągnięciach uczniów jest na pewnym poziomie ogólności. Należy rozważyć, czy komunikując wyniki uczniów szkołom, nie należałoby oprócz wyniku punktowego kodować informacji o cechach rozwiązania.

*Autorki składają podziękowanie Panu Krzysztofowi Słomczyńskiemu, kierownikowi Wydziału Badań i Analiz w OKE w Jaworznie, za pomoc w opracowywaniu danych statystycznych badanej populacji zdających.*