

Ocenianie rozprawki egzaminacyjnej, czyli od klucza do skali

ADAM BROŻEK*, JOANNA DOBKOWSKA*

Autorzy artykułu krytycznie omawiają klucz punktowania rozprawki stosowany na egzaminie gimnazjalnym w latach 2002–2011. Przedstawiają podejmowane w Instytucie Badań Edukacyjnych próby opracowania holistycznej skali oceny tej formy wypowiedzi, które przyczyniły się do wprowadzenia zasadniczej zmiany w ocenianiu egzaminacyjnej rozprawki oraz stały się inspiracją dla zespołu pracującego w Centralnej Komisji Egzaminacyjnej nad modernizacją egzaminu maturalnego z języka polskiego.

Ilościowy klucz oceny treści egzaminacyjnej rozprawki

Od pierwszego egzaminu gimnazjalnego, który odbył się w 2002 roku, w systemie egzaminacyjnym wprowadzono ilościowy sposób oceny treści gimnazjalnej rozprawki, nie podając żadnego uzasadnienia tej decyzji. Jednoznacznym zwiastunem tego podejścia było sformułowanie tematu wypracowania w pierwszym teście gimnazjalnym: „Warto poznawać dorobek naszych przodków. Napisz rozprawkę, w której ustosunkujesz się do tej opinii. Odwołaj się do trzech przykładów osiągnięć cywilizacyjnych (jednego z arkusza i dwóch własnych)” (CKE, 2002, zad. 29.). W przytoczonym poleceniu wątpliwości budzi już sama jego teza, która ma postać pseudoproblemu: „Warto poznawać dorobek na-

szych przodków”. Do czego uczeń miał się właściwie ustosunkować? Czy możliwe jest racjonalne wystąpienie przeciwko tak postawionemu twierdzeniu? Bardzo wątpliwe.

Zwróćmy jednak uwagę na wskazówkę sformułowaną dalej: „Odwołaj się do trzech przykładów”. Dlaczego akurat do trzech, a nie do dwóch albo czterech? Oczywiście dlatego, żeby sprawdzający rozprawkę egzaminator mógł stwierdzić, czy ten ilościowy wymóg został spełniony. Potwierdza to wyjaśnienie ujęte w kluczu punktowania (CKE, 2002):

Rozwinięcie obejmuje:

- a) przywołanie przykładu z arkusza – 1 pkt.
- b) posłużenie się tym przykładem do uzasadnienia tezy – 1 pkt.
- c) podanie własnych przykładów – 1 pkt.

Artykuł powstał m.in. na podstawie badania *Diagnoza kompetencji gimnazjalistów 2011* przeprowadzonego w Instytucie Badań Edukacyjnych w ramach projektu „Badanie jakości i efektywności edukacji oraz instytucjonalizacja zaplecza badawczego” współfinansowanego ze środków Europejskiego Funduszu Społecznego (Program Operacyjny Kapitał Ludzki 2007–2013, priorytet III: Wysoka jakość systemu oświaty) oraz badania przeprowadzonego w Centralnej Komisji Egzaminacyjnej

w ramach projektu „Pilotaż nowych egzaminów maturalnych: modernizacja egzaminu maturalnego z języka polskiego” w komponencie IV.4 w obszarze języka polskiego „Modernizacja egzaminu maturalnego z języka polskiego” współfinansowanego ze środków Europejskiego Funduszu Społecznego.

* Pracownia Języka Polskiego, Instytut Badań Edukacyjnych. E-mail: j.dobkowska@ibe.edu.pl

- d) posłużenie się co najmniej jednym do uzasadnienia tezy – 1 pkt.

Po pierwsze, z tak sformułowanego klucza jednoznacznie wynika, jak uczeń powinien się ustosunkować do tezy – tylko pozytywnie. Wymienione „przykłady z arkusza” nie dają innej możliwości (są to trzy krótkie teksty o odkryciach: Pompejów, grobu Tutenchamona i starożytnej Troi). Po wtóre, w kluczu zupełnie pominięto kwestię jakości argumentacji. W tej sytuacji praca egzaminatora musi polegać na automatycznym wyszukiwaniu przykładów (jednego z arkusza i dwóch uczniowskich) w rozprawce i stwierdzeniu, czy owe przykłady zostały użyte jako argumenty potwierdzające słuszność tezy postawionej przez autora tematu. Można zatem stwierdzić, że rozprawka w tak ujętym kluczu punktowania traktowana jest raczej jako swoisty zbiór danych ilościowych, a nie struktura tekstowa podlegająca ocenie jakościowej.

Ilościowo sformułowany klucz ma rzekomą pomiarową zaletę – pozwala na w miarę dokładne punktowanie oczekiwanych cech rozprawki, w szczególności liczby przykładów, którymi posłużył się uczeń. Niestety, wymieniona zaleta nie równoważy bardzo istotnej wady dydaktycznej, którą jest konieczne „usztynwienie” tematu rozprawki poprzez podanie w nim wymaganej arbitralnie liczby koniecznych przykładów oraz – co istotniejsze – marginalizowanie kwestii jakościowych. Zauważmy bowiem, że wymienione w cytowanym kluczu oczekiwane cechy rozprawki nie obejmują najistotniejszej – oceny jakości rozumowania ucznia. A przecież rozprawka nie jest niczym innym, jak zapisem rozumowania. Zgodnie z cytowanym kluczem uczeń otrzyma 2 punkty w rozwinięciu tematu za samo wymienienie trzech przykładów osiągnięć cywilizacyjnych, bez względu na to, w jakim zostaną podane kontekście. Można zatem wyobrazić sobie, że uczeń zapisuje w swoim wypra-

cowaniu zdanie: „Carter odkrył grób Tutenchamona, dla ludzkości ważne również było wynalezienie żarówki i telefonu” – i zgodnie z kluczem za rozwinięcie tematu musi już otrzymać 2 punkty. Jeśli liczba błędów językowych, ortograficznych i interpunkcyjnych zmieści się w normie, uczeń otrzyma za wypracowanie większość punktów – nawet jeśli nie będzie ono rozprawką(!).

Taki model klucza punktowania z biegiem lat spotykał się z coraz większą krytyką – polonistycznych środowisk dydaktycznych i samych nauczycieli. Zarzucano mu przede wszystkim to, że koncentrując się na pomiarze aspektów drugorzędnych, wypacza sens rozprawki, a tym samym destrukcyjnie wpływa na nauczanie tej formy wypowiedzi w szkole. Te zarzuty system egzaminacyjny zawsze odpierał, akcentując konieczność zapewnienia rzetelności pomiaru osiągnięć uczniów. Co ciekawe – nigdy nie przedstawiono przekonujących dowodów na to, że stosując klucz, egzaminatorzy rzeczywiście punktuja wypracowania w sposób rzetelny (i porównywalny)¹. Tak więc dla iluzorycznej rzetelności, poświęcono trafność pomiaru.

Innym powodem krytyki klucza punktowania był sposób opracowania jego treści. Ostateczną wersję nierzadko redagowano po przeprowadzeniu egzaminu. Zajmowało się tym grono ekspertów z Centralnej i okręgowych komisji egzaminacyjnych. Bywało, że klucz był jeszcze korygowany w trakcie oceniania prac przez egzaminatorów. Ustaloną wersję komisje egzaminacyjne publikowały na swoich stronach internetowych nie wcześniej niż kilka tygo-

¹ Przeprowadzone przez Romana Dolatę badania efektu egzaminatora dowiodły, że stosując ilościowy klucz punktowania prac maturalnych, egzaminatorzy znacznie różnią się w ocenie tego samego wypracowania. Rozrzut w skrajnych wypadkach wynosił prawie 20 punktów, a rzetelność skali rozwinięcia tematu oszacowano na 54% (Dolata, Putkiewicz i Wiłkomirska, 2004).

dni po przeprowadzeniu egzaminu i po oceniu prac egzaminacyjnych. Nauczyciele, uczniowie i ich rodzice domagali się natomiast, aby szczegółowe zasady oceny wypracowania były znane, zanim gimnazjaliści do egzaminu przystąpią.

Mimo zarzutów i krytyki ilościowy system punktowania treści wypracowań był w egzaminach zewnętrznych stosowany przez dziesięć lat (2002–2011).

Skala oceny eseju w egzaminach SAT

Chcąc wspomóc system egzaminacyjny w reformowaniu nietrafnego sposobu oceniania wypracowania gimnazjalnego, eksperci Instytutu Badań Edukacyjnych (IBE) zaczęli w 2009 roku pracę nad skonstruowaniem szacunkowej skali oceny rozprawki. Szukając inspiracji w funkcjonujących na świecie systemach oceniania wypracowań egzaminacyjnych, skupili się na amerykańskiej skali stosowanej do oceny eseju w Scholastic Assessment Test (SAT)² (formę wypowiedzi pisemnej obowiązującej na egzaminie SAT w polskich warunkach nazwalibyśmy raczej rozprawką³).

Na tej skali wyodrębniono sześć hierarchicznie uporządkowanych poziomów jakości wypracowania (SAT College Board, bdw.):

1. Esej pokazuje znikomą biegłość w tworzeniu tekstu lub jej brak. Jest bardzo trudny w odbiorze.

2. Esej pokazuje niewielką biegłość w tworzeniu tekstu.
3. Esej pokazuje nierówną, choć widoczną biegłość w tworzeniu tekstu.
4. Esej pokazuje wystarczającą biegłość w tworzeniu tekstu, choć zawiera usterki jakościowe.
5. Esej pokazuje znaczną biegłość w tworzeniu tekstu, choć może zawierać sporadyczne usterki jakościowe.
6. Esej pokazuje oczywistą dla czytelnika i widoczną w całej pracy biegłość w tworzeniu tekstu, choć może zawierać kilka mniej znaczących błędów.

Każdemu z wymienionych poziomów przypisano zespół wyróżniających go cech. Na przykład wypracowanie na poziomie szóstym – najwyższym:

- efektywnie, dogłębnie i kreatywnie prezentuje poglądy autora na wybrany temat, dowodząc jego wybitnej zdolności do krytycznego myślenia; zawiera ponadto bezdyskusyjnie trafnie dobrane przykłady, argumenty oraz inne formy uzasadnienia wyrażonego stanowiska;
- ma dobrą kompozycję i jest skupiony wokół tematu, cechuje go oczywista dla czytelnika spójność oraz sprawne rozwijanie myśli;
- wskazuje na sprawność autora w posługiwaniu się językiem, zawiera zróżnicowane, precyzyjne i adekwatne do prezentowanych treści słownictwo;

² SAT to zestaw standardowych testów przeznaczony dla kandydatów aplikujących na studia na amerykańskich uczelniach. Jedną z jego sekcji jest poświęcona umiejętnościom formułowania wypowiedzi pisemnej i składa się z dwóch części: zamkniętej – pytania wyboru wielokrotnego (70% oceny cząstkowej) i krótkiego, pisanego w ciągu 25 minut wypracowania zwanego esejem (30% oceny cząstkowej).

³ Tematy proponowane na egzaminie SAT dotyczą zwykle zagadnień społecznych lub filozoficznych. Przykładowy temat podany w materiale przygotowującym do tego egzaminu brzmi: „Wiele osób wierzy, że – żeby wspiąć się po drabinie sukcesu i osiągnąć – należy zapomnieć

o przeszłości, tłumić ją lub wyrzec się jej. Inni mają odmienny punkt widzenia – wspomnienia traktują jako szansę rozliczenia się z przeszłością oraz integracji przeszłości i teraźniejszości” (na podstawie: Sara Lawrence-Lightfoot, *I’ve Known Rivers: Lives of Loss and Liberation*). Czy wspomnienia przeszkadzają, czy pomagają ludziom w ich wysiłkach, aby uczyć się na podstawie przeszłości i odnosić sukcesy w teraźniejszości? Napisz esej, w którym przedstawiś swój punkt widzenia na tę sprawę. Wesprzyj swoją tezę argumentacją i przykładami zaczerpniętymi z lektury, studiów oraz własnych doświadczeń i obserwacji [tłum. J. Dobkowska]. Źródło: SAT College Board, bdw.

- wskazuje na sprawność autora w świadomym różnicowaniu struktur składniowych;
- jest zasadniczo wolny od błędów gramatycznych, stylistycznych oraz usterek w zapisie.

Natomiast esej na poziomie najniższym (1) wyróżnia się jedną lub więcej niż jedną z następujących cech:

- brak czytelnych poglądów na omawiany temat lub poglądy bardzo słabo poparte jakimikolwiek formami uzasadnienia;
- zasadnicze usterki kompozycyjne, brak skupienia na temacie, co prowadzi do chaotycznej i niespójnej prezentacji myśli;
- poważne błędy w budowie zdań, bardzo liczne błędy gramatyczne, stylistyczne oraz usterki graficzne, które utrudniają zrozumienie treści.

Choć na skali SAT osobno opisano poszczególne cechy eseju, ma ona jednak charakter holistyczny, gdyż występowanie tych cech ocenia się całościowo. Innymi słowy – ocena za esej jest funkcją współwystępowania w nim wyszczególnionych cech, a nie sumą ocen cząstkowych wystawianych osobno za poszczególne cechy. Skłania to oceniających do pracy dwuetapowej – przeprowadzenia analizy występowania poszczególnych cech, a następnie dokonania oceny ich funkcjonowania w strukturze wypracowania. Zwróćmy też uwagę na pragmatyczne podejście do oceny języka. W SAT używa się wprawdzie w pewnym sensie ilościowego określenia „bardzo liczne błędy”, ale odnosi się je następnie do ich wpływu na czytelność tekstu (zrozumienie jego treści).

Aby ułatwić osobom zamierzającym przystąpić do egzaminu SAT przygotowanie się do napisania eseju, skala została wyposażona w obszerny komentarz, w którym syntetycznie wyjaśniono:

- cechy eseju podlegające ocenie (zdolność do krytycznego myślenia, umiejętność

prezentacji tematu, kompozycja tekstu, sposób wykorzystania języka);

- co jest w eseju najważniejsze, a co drugorzędne (obecność czytelnej myśli przewodniej, odpowiedni dobór argumentów, właściwe rozwinięcie tematu, skuteczna kompozycja oraz precyzyjne wykorzystanie języka – zostaną wysoko ocenione bez względu na styl czy technikę pisania);
- jaka jest waga usterek, błędów językowych i błędów w zapisie w stosunku do jakości treści (uczeń może uzyskać najwyższy wynik nawet, gdy popełnił drobne błędy).

Na koniec dodajmy, że w celu zapewnienia obiektywizmu ocen eseje w egzaminach SAT standardowo oceniają dwie osoby. Jeśli ich oceny różnią się ponad dopuszczalną normę, w roli arbitra występuje trzeci oceniający.

Tworzenie nowej skali oceny rozprawki

Zespół ekspertów IBE w 2009 roku opracował wzorowany na SAT projekt holistycznej skali oceny rozprawki, w którym wyszczególniono pięć hierarchicznie ujętych poziomów oceny wypracowania. Zasadniczymi czynnikami wyróżniającymi kolejne poziomy były: sposób prezentacji poglądów przez autora, jakość argumentacji oraz trafność przykładów ją wzmacniających. Na poziomie najwyższym opis wymaganych cech odnoszących się do treści rozprawki wyglądał następująco: „precyzyjnie i twórczo zaprezentowane poglądy autora, pogłębiona argumentacja i trafne przykłady ją wzmacniające”. A o stopień niżej: „czytelnie zaprezentowane poglądy autora, trafna i wyczerpująca argumentacja i przykłady ją wzmacniające” (IBE, 2009).

Niestety, badanie pilotażowe pokazało, że nauczyciele mają bardzo duże kłopoty z rozróżnieniem tak opisanych poziomów. Wydaje się to zrozumiałe, bo o ile w wypadku sposobu prezentacji poglądów przez

autora poziom najwyższy został jakościowo oddzielony od poziomu niższego przy słówkiem „twórczo”, to już trudno się zgodzić z tym, że zapisany na poziomie najwyższym przymiotnik „pogłębiona” odnoszący się do argumentacji oznacza jakość lepszą niż określenie „trafna i wyczerpująca”. Podobnie jest z próbą opisaną przykładów wzmacniających argumentację. Określenie ich na poziomie najwyższym przymiotnikiem „trafne” niczego jakościowo lepszego w stosunku do poziomu niższego nie wnosi, gdyż przykłady, które rzeczywiście wzmacniają argumentację, siłą rzeczy muszą być trafne. Przykłady nietrafne mogą argumentację jedynie osłabić.

Nieefektywne okazały się również ambitne próby pragmatycznego podporządkowania formalnych cech rozprawki jej treści i skonstruowania na tej podstawie skali w pełni holistycznej, zmuszającej egzaminatora do punktowania wszystkich cech rozprawki. W pilotażu okazało się, że nauczyciele napotykają w tym zakresie na co najmniej dwie poważne przeszkody: nie rozróżniają sąsiadujących ze sobą poziomów opisów cech języka i zapisu (IBE, 2009), np.:

- tekst jest w przeważającej mierze wolny od błędów (poziom wyższy),
- tekst zawiera niewielką liczbę błędów, które nie utrudniają rozumienia tekstu (poziom niższy).

Krytyka tak rozróżnionych poziomów wydaje się uzasadniona. Praca w „przeważającej mierze wolna od błędów” to w istocie rzeczy to samo co praca „zawierająca niewielką liczbę błędów”.

Inny problem wynikał z tego, że niekiedy w uczniowskich rozprawkach poziom języka i zapisu nie pokrywał się z przypisanym mu teoretycznie na skali poziomem treści. Innymi słowy – zdarzały się rozprawki, w których poziom treści znacznie przewyższał przypisa-

ny mu w skali poziom języka. Bywało również odwrotnie – w rozprawkach formalnie nienagannych pod względem językowym egzaminatorzy stwierdzali pustosłowie.

W kolejnych latach zdecydowano się rozdzielić konstruowaną skalę na trzy podskale oceniające odrębne aspekty rozprawki: (a) treść, (b) warstwę językowo-stylistyczną (w tym interpunkcję), (c) poprawność ortograficzną.

W skali oceny treści wyróżniono cztery poziomy rozwinięcia tematu oznaczone jako A, B, C, D. Na najwyższym poziomie (A) wskazano sześć cech, którymi powinna się charakteryzować bardzo dobra rozprawka (Bańkowska, Dobkowska i Mikołajczuk, 2010):

- odnosi się do problemu sformułowanego w temacie,
- jest poprawna merytorycznie,
- przedstawia stanowisko autora,
- zawiera trafną argumentację,
- zachowuje logikę wyводу podkreśloną segmentacją tekstu,
- zawiera podsumowanie przeprowadzonego rozumowania.

Na niższych poziomach niektóre cechy zostały zredukowane (np. na poziomie C praca tylko „w przeważającej części” odnosi się do problemu sformułowanego w temacie) lub usunięte (np. w pracy, która ze względu na zasadnicze dla rozprawki elementy: związek z problemem sformulowanym w temacie, poprawność merytoryczną, przedstawienie stanowiska autora i trafność argumentacji została sklasyfikowana na poziomie C, nie ocenia się już logiki wyvodu i ewentualnego podsumowania rozważań, ponieważ cechy te – choć ważne – są w takim przypadku drugorzędne). Taka struktura skali podkreśla więc hierarchię poszczególnych elementów.

Interesujące kryteria zastosowano w skali oceny warstwy językowo-stylistycznej – jako wy-

nik kompromisu pomiędzy sugerowanym przez ekspertów podejściem holistycznym a narzuconym przez CKE tradycyjnym podejściem analitycznym. Na przykład na poziomie najwyższym opis wyglądał następująco (Bańkowska, Dobkowska i Mikołajczuk, 2010):

- Umiejętne stosowanie składni argumentacyjnej, oddającej tok logicznego wyводу i nie więcej niż dwa błędy składniowe. Składnia podkreślona poprawną interpunkcją i nie więcej niż trzy błędy interpunkcyjne.
- Umiejętne zastosowanie słów i związków frazeologicznych, które umożliwiają precyzyjne wyrażanie myśli i nie więcej niż dwa błędy słownikowo-frazeologiczne.
- Poprawna fleksja.

Synkretyczne (pragmatyczno-ilościowe) zestawienie składni i frazeologii na tej skali sprawia, że przewagę ma nadal podejście ilościowe – nawet biegłość w stosowaniu składni argumentacyjnej, dobrze podkreślającej logikę wyводу, nie ma bowiem żadnego znaczenia po przekroczeniu limitu dwóch błędów składniowych, które powodują obniżenie oceny. Analogicznie rzecz ma się w przypadku związków frazeologicznych.

Takie samo podejście, skutkujące deprecjacją kryterium jakościowego, zastosowano w skali oceny poprawności ortograficznej. Na przykład na poziomie C wskazano:

- Zapis powoduje znaczne spowolnienie odbioru tekstu – nie więcej niż cztery błędy ortograficzne rażące lub ich równoważność z udziałem błędów drugorzędnych.

Omówione wyżej skale zostały w sierpniu 2010 roku przedstawione Centralnej Komisji Egzaminacyjnej, która nie zdecydowała się jednak na ich zamieszczenie w informatorze gimnazjalnym. Znalazła się tam natomiast czteropunktowa skala oceny treści rozprawki oparta na kategorii rozumowania stosowanej w logice (Konarzewski, 2010, s. 13–14.):

- poziom 4.: rozumowanie pełne (teza, argumentacja, wniosek), argumentacja wyczerpuje wymagania tematu, wszystkie argumenty trafne;
- poziom 3.: rozumowanie pełne (teza, argumentacja, wniosek), ale argumentacja nie wyczerpuje wymagań tematu, lub argumentacja wyczerpuje wymagania tematu, ale niektóre argumenty są nietrafne;
- poziom 2.: rozumowanie niepełne (nie jest jasne, za czym lub przeciw czemu zdający argumentuje), większość argumentów trafnych;
- poziom 1.: próba rozumowania (nie da się stwierdzić, za czym lub przeciw czemu zdający argumentuje).

Łatwo zauważyć, że kryteria zamieszczone na skali nie zawsze są wobec siebie konsekwentne. Na przykład jeśli na poziomie 2. zakłada się, że „nie jest jasne, za czym lub przeciw czemu zdający argumentuje”, to jak stwierdzić, czy argumenty są trafne? Trafność jest przecież w wypadku argumentacji kategorią obligatoryjnie odnoszoną do stanowiska, *na rzecz* którego się argumentuje.

Do oceny języka skonstruowano w cytowanym informatorze kryteria oparte na jednoznacznych normach ilościowych (CKE, 2010):

Język:

- poziom 2: najwyżej 3 błędy składniowe lub leksykalne, lub frazeologiczne, lub fleksyjne
- poziom 1: 4 błędy składniowe lub leksykalne, lub frazeologiczne, lub fleksyjne
- poziom 0: ponad 4 błędy składniowe lub leksykalne, lub frazeologiczne, lub fleksyjne

Ortografia:

- poziom 1: najwyżej 2 błędy
- poziom 0: ponad 2 błędy

Interpunkcja:

- poziom 1: najwyżej 3 błędy
- poziom 0: ponad 3 błędy

Korzystając z zebranych doświadczeń, w IBE opracowano czterostopniową skalę oceny treści wypowiedzi argumentacyjnej, która posłużyła do oceny wypracowań w badaniu *Szkoła samodzielnego myślenia* (Tabela 1). We wrześniu 2011 roku skalę tę przejęła od IBE CKE i uczyniła z niej oficjalne narzędzie oceniania egzaminacyjnej rozprawki gimnazjalnej, przy czym przekodowała wcześniej poziomy A, B, C, D na punkty – odpowiednio: 4, 3, 2, 1.

Metodologiczne i pomiarowe własności skali

Skala oceny treści rozprawki zakłada możliwość przyporządkowania jej cech do jednego z pięciu poziomów, przy czym cztery z nich dotyczą prac, które można określić jako rozprawki, piąty poziom obejmuje prace ocenione na 0 punktów, których ze względu na brak argumentacji (będącej dla tej formy wypowiedzi elementem konstytutywnym) do rozprawek zaliczyć nie można lub które nie odnoszą się do problemu sformułowanego w temacie.

Przyjęcie zmodyfikowanej skali pociąga za sobą konieczność starannego formułowania tematu, który musi zawierać rzeczywisty problem, możliwy do rozważenia w różny sposób i interesujący dla ucznia. Dobrym tego przykładem jest temat jednej z rozprawek z cytowanego informatora o egzaminie gimnazjalnym (CKE, 2010): „*Zemsta Aleksandra Fredry kończy się deklaracją zgody. Rozważ, jak trwała będzie ta zgoda. Wykorzystaj podany fragment, znajomość całego dramatu i własne doświadczenie*” (uzupełnienie tematu stanowi końcowy fragment *Zemsty*). Przykładu negatywnego dostarcza tegoroczny arkusz egzaminacyjny, w którym temat rozprawki („*Literatura pozwala lepiej poznać i zrozumieć minionie wieki. Rozważ słuszność tego stwierdzenia w rozprawce. zilustruj swoje argumenty przykładami literackimi.*”) zawiera – tak jak przykład cytowany na wstępie niniejszego artykułu – tezę właściwie bezdyskusyjną.

Cztery cechy uwzględnione na skali odnoszą się do podstawowych elementów wy-

Tabela 1
Skala punktowania treści rozprawki

A	B	C	D*	0 p.	N
1. Praca odnosi się do problemu sformułowanego w temacie.	1. Praca odnosi się do problemu sformułowanego w temacie.	1. Praca w przeważającej części odnosi się do problemu sformułowanego w temacie.	1. Praca jest luźno związana z problemem sformułowanym w temacie.	Praca nie odnosi się do problemu sformułowanego w temacie. lub: Praca nie zawiera argumentacji.	Brak pracy lub tekst nieczytelny.
2. Przedstawia stanowisko autora.	2. Przedstawia stanowisko autora.	2. Przedstawia stanowisko autora.	2. Przedstawia stanowisko autora.		
3. Zawiera trafną i wnikliwą argumentację.	3. Zawiera trafną argumentację.	3. Zawiera częściowo trafną argumentację.	3. Zawiera próbę argumentacji.		

*Ocena 1 przypisywana jest pracy, która nie spełnia kryteriów tekstu argumentacyjnego, ale świadczy o podjęciu przez ucznia próby argumentacji.

maganych w rozprawce z założeniem, że ze względu na czas przeznaczony na pisanie rozprawki egzaminacyjnej (ok. 30–40 minut) praca ucznia nie musi zawierać skomplikowanych przykładów i charakteryzować się wyszukaną kompozycją; powinna jednak być precyzyjna, komunikatywna i zgodna z założeniami gatunku (Mikołajczuk i Puzynina, 2004, s. 49). Cechy te to:

- zgodność z tematem – odniesienie do problemu sformułowanego w temacie,
- stanowisko autora wobec problemu (sąd wyrażony bezpośrednio lub pośrednio – postawiona teza lub hipoteza),
- trafność i wnikliwość argumentacji,
- logika wyводу (błąd rzeczowy zakłócający logikę wyводу powoduje obniżenie oceny o jeden poziom).

Rozprawka oceniona najwyżej powinna: (1) odnosić się do problemu sformułowanego w temacie (w tradycyjnym ujęciu polonistycznym określano to mniej precyzyjnie jako „praca na temat”); (2) przedstawiać stanowisko autora (bezpośrednio – w postaci postawionej tezy lub hipotezy albo też pośrednio, tak aby jasne było na rzecz jakiego stanowiska autor rozprawki przedstawia argumentację); (3) zawierać trafną i wnikliwą argumentację (liczba argumentów jest nieistotna, całość argumentacji powinna przekonująco świadczyć o słuszności przyjętego stanowiska) i (4) zachowywać logikę wyводу podkreślona segmentacją tekstu.

Na niższych poziomach niektóre cechy zostały zredukowane lub usunięte, co sugeruje hierarchię poszczególnych elementów.

Przygotowana skala czterostopniowa została zastosowana podczas kodowania rozprawek pozyskanych w badaniu *Diagnoza kompetencji gimnazjalistów 2011*. Kodowanie odbyło się w dwóch etapach. Najpierw zespół 30 egzaminatorów zakodował 1000 losowo wybranych prac (wylosowanych z pró-

by liczącej ok. 8000. prac). Celem drugiego etapu kodowania było zbadanie, jaką zgodność osiągają egzaminatorzy w ustalaniu ocen za treść rozprawki oraz w ustalaniu liczby błędów językowych, ortograficznych i interpunkcyjnych. Aby to stwierdzić, spośród 1000 zakodowanych wcześniej rozprawek zostało wybranych 200 prac, każda z nich została powielona 6 razy. Dobrano następnie (nielosowo) 40 egzaminatorów i podzielono ich (losowo) na dwa dwudziestoosobowe zespoły. Każdemu egzaminatorowi przydzielono do zakodowania 30 rozprawek. Przy tak założonym planie punktowania każda praca była w obu zespołach punktowana trzykrotnie przez różnych egzaminatorów.

Korelacja średnich wyników punktowania treści rozprawek, mierzona współczynnikiem *r*-Pearsona, w obu zespołach okazała się wysoka – wyniosła 0,85. Nieoczekiwanie bardzo małą zgodność egzaminatorzy osiągnęli w ustalaniu w pracach liczby błędów językowych, interpunkcyjnych i ortograficznych. W wypadku tych kryteriów trójki egzaminatorów zgodnie ustaliły liczbę błędów w niespełna 10% wypracowań (!), z ortografią poszło im nieco lepiej, ale i tak zgodność nie przekroczyła 50%. Taki stan rzeczy stawia pod znakiem zapytania sensowność posługiwania się przez egzaminatorów kluczem punktowania, w którym ściśle określono limit dopuszczalnych błędów.

Implikacje dla oceniania rozprawki na zmodernizowanej maturze

Doświadczenia efektów prac nad skalą oceny rozprawki gimnazjalnej zostały wykorzystane przez zespół powołany w 2010 roku przy CKE do opracowania projektu zmodernizowanej matury z języka polskiego na rok 2015. Celem projektowanych zmian było przede wszystkim dostosowanie egzaminu maturalnego do nowej podstawy programo-

wej, a w zakresie części pisemnej – zwiększenie liczby wariantów wypracowania na egzaminie. Zespół przyjął, że na poziomie podstawowym uczniowie będą pisać rozprawkę opartą na problemie zarysowanym w tekście epickim lub dramatycznym, bądź interpretację pojedynczego tekstu lirycznego. Na poziomie rozszerzonym maturzyści będą mieli do wyboru rozprawkę lub szkic na podstawie tekstu teoretycznego, bądź interpretację porównawczą. W związku z tym, że stosowany od 2005 roku klucz analityczny oparty na wyszukiwaniu w pracy określonych treści oraz cech kompozycyjno-językowych, który skłaniał zdających do nienaturalnego zachowania się w sytuacji testowania polegającego na pisaniu „pod klucz”, oceniany był jako nietrafny i zniekształcający edukację polonistyczną w szkołach, istotnym celem prac modernizacyjnych stało się opracowanie i empiryczne sprawdzenie nowych wariantów kryteriów oceniania, w tym przede wszystkim skali holistycznej.

Założeniem zespołu pracującego nad modernizacją egzaminu maturalnego było stworzenie – tam, gdzie to możliwe – spójnego systemu oceniania prac uczniów na III i IV etapie edukacji, który wspierałby kształcenie umiejętności pisania egzaminacyjnych form wypowiedzi, a zarazem umożliwił uchwycenie postępów uczniów kończących kolejne etapy edukacji. Formą wypowiedzi, która dawała możliwość stworzenia takiego narzędzia oceny, jest rozprawka obecna na egzaminie gimnazjalnym i na maturze.

Przygotowana w projekcie maturalnym skala oceny rozprawki łączy w sobie kryteria oceny treści oraz warstwy językowej wypracowania (skale oceniania pozostałych form wypowiedzi skonstruowane są w analogiczny sposób). Rozprawkę ocenia się na podstawie ośmiu niżej wymienionych kryteriów, z których każde ma przypisaną odpowiednią wagę.

Sformułowanie stanowiska (tezy lub hipotezy, waga 2). Uczeń powinien zrozumieć problem postawiony w poleceniu i sformułować stanowisko będące propozycją jego rozwiązania (nie musi ono stanowić wyodrębnionej części rozprawki). Stanowisko jest oceniane pod względem tego, czy jest udaną próbą rozwiązania problemu oraz czy uwzględnia więcej niż jeden punkt widzenia. Stanowisko będące nieudaną próbą rozwiązania problemu jest skutkiem albo jego niezrozumienia spowodowanego niewłaściwą interpretacją słów lub fraz polecenia, albo niepełnego zrozumienia przekazu polecenia. Stanowisko uwzględnia więcej niż jeden punkt widzenia, gdy zawiera przynajmniej dwa sposoby rozumienia lub rozwiązania problemu. Stanowisko uwzględnia tylko jeden punkt widzenia, gdy bezpośrednio zakłada jedno rozumienie problemu i zawiera tylko jedno rozwiązanie. Rozprawka, w której brak stanowiska, może zawierać streszczenie tekstu, biografię autora itp. lub rozwijać myśli niezwiązane z poleceniem.

Uzasadnienie stanowiska (waga 4). Uczeń powinien uargumentować wszystkie elementy swojego stanowiska. Uzasadnienie, które nie musi stanowić wyodrębnionej kompozycyjnie części rozprawki, jest oceniane pod względem trafności i szerokości ujęcia. Uzasadnienie trafne zawiera zgodne z prawdą lub prawdopodobne oraz logicznie poprawne argumenty za przyjętym rozwiązaniem problemu i przeciw rozwiązaniu, które uczeń odrzucił. Uzasadnienie jest nietrafne, gdy operuje argumentami nieprawdziwymi, zawiera rozumowania *non sequitur* lub sądy wzajemnie sprzeczne. Szerokość uzasadnienia ocenia się na podstawie liczby źródeł, z których pochodzą argumenty użyte przez ucznia. Uzasadnienie jest szerokie, jeśli uczeń trafnie wykorzystał wszystkie źródła wymienione w poleceniu (np. podany fragment *Lalki*, cały utwór Bolesława Prusa oraz inny tekst kultury). Uzasadnienie wąskie ogranicza się

do trafnych argumentów tylko z niektórych poleconych źródeł. Rozprawka, w której brak uzasadnienia, dostarcza argumentów jedynie w sprawie drugorzędnych elementów przyjętego rozwiązania lub rozwija myśli niezwiązane ze stanowiskiem.

Poprawność rzeczową (waga 1) ocenia się na podstawie liczby i wagi błędów rzeczowych. Za błąd istotny uważa się błąd świadczący o nieznanomości lektur obowiązkowych lub kontekstu historycznoliterackiego. Pomyłki (np. w nazwach własnych lub datach) uważa się za błędy nieistotne.

Zamyśl kompozycyjny (waga 2) ocenia się pod względem funkcjonalności struktury i uporządkowania tekstu stosownie do wybranego przez ucznia gatunku wypowiedzi. Należy wziąć pod uwagę, czy w tekście zostały wyodrębnione – językowo i graficznie – części pracy i akapity niezbędne dla jasnego sformułowania stanowiska i uzasadniających go argumentów oraz czy wyodrębnione części i akapity są uporządkowane konsekwentnie (bez luk i zbędnych powtórzeń). Wagę zaburzenia funkcjonalności ocenia egzaminator na podstawie całości pracy (np. brak zakończenia w pracy, która jasno rozwija przyjęte stanowisko, uznaje się za niewielkie zaburzenie, natomiast podważenie w zakończeniu wcześniejszych wywodów – za znaczne zaburzenie).

Spójność lokalną (waga 2) w rozprawce ocenia się na podstawie zgodności logicznej i gramatycznej między zdaniem w akapitach.

Poprawność stylistyczną (waga 2) ocenia się pod względem stosowności i jednolitości stylu rozprawki. Styl uznaje się za stosowny, jeśli uczeń dostosował środki językowe do wybranego przez siebie gatunku wypowiedzi i używał konstrukcji składniowych i jednostek leksykalnych w sposób celowy. Styl

jest niestosowny, jeśli np. w tekście typowej rozprawki problemowej występują fragmenty przesadnie zmetaforyzowane lub potoczne, wtręty ze stylu urzędowego itp. Styl uznaje się za jednolity, jeśli autor posługuje się swoistymi dla niego środkami językowymi w sposób konsekwentny.

Poprawność językową (waga 2) ocenia się na podstawie liczby (szacowanej odpowiednio do objętości tekstu) i wagi błędów składniowych, leksykalnych (w tym słowotwórczych), frazeologicznych, fleksyjnych i stylistycznych.

Poprawność zapisu (waga 2) ocenia się pod względem liczby błędów ortograficznych i interpunkcyjnych (szacowanej odpowiednio do objętości tekstu) oraz ich wagi (błędy rażące i nierażące).

Proponowana skala ma charakter synkretyczny, łączy w sobie cechy narzędzia holistycznego – dla każdego z kryteriów określone są poziomy jakości wykonania⁴ oraz analitycznego – za każde kryterium uczeń osobno otrzymuje punkty, które sumują się, dając ostateczny wynik. Decydująca jest punktacja uzyskana za treść pracy – jeśli maturzysta uzyska tu minimalny wynik, egzaminator nie przyznaje punktów w pozostałych kategoriach. Takie połączenie wydaje się być rozwiązaniem optymalnym: skala oparta wyłącznie na ocenianiu holistycznym nie jest optymalnym rozwiązaniem ze względu na złożoność zadania, które wykonuje uczeń

⁴ Na przykład w czterostopniowym kryterium „Sformułowanie stanowiska” rozprawka oceniona na poziomie najwyższym (6 punktów) zawiera stanowisko, które jest udaną próbą rozwiązania problemu z uwzględnieniem więcej niż jednego punktu widzenia. W przypadku rozprawki o poziom słabszej (4 punkty) stanowisko jest udaną próbą rozwiązania problemu z uwzględnieniem tylko jednego punktu widzenia. W rozprawce ocenionej na 2 punkty stanowisko jest nieudaną próbą rozwiązania problemu, zaś 0 punktów otrzymuje uczeń, który nie sformułował stanowiska.

(napisanie maturalnej rozprawki egzaminacyjnej muszą poprzedzać takie czynności, jak: analiza polecenia, interpretacja dołączonego tekstu kultury pod kątem wskazanego w poleceniu problemu, wyszukanie odpowiednich kontekstów – tekstów kultury i problemów, które łączą się z tematem wypowiedzi); specyfikę języka polskiego (w przeciwieństwie do języka angielskiego trudno zaakceptować w polonistycznej pracy maturalnej „rażące błędy w budowie zdań, bardzo liczne błędy gramatyczne, stylistyczne oraz usterki graficzne, które utrudniają zrozumienie treści” (SAT College Board, *bdw.*); tradycję polonistyczną (trudno wyobrazić sobie nauczyciela-polonistę, który pomija w ocenie aspekt poprawności językowej i poprawności zapisu) i – wreszcie – skłonność egzaminatorów do oceniania intuicyjnego, które może skutkować przyjmowaniem indywidualnych kryteriów oceny lub wybiórczym ocenianiem określonych cech.

Przygotowaną w projekcie skalę oceniania wraz z przykładowymi zadaniami egzaminacyjnymi poddano badaniu pilotażowemu, którego celem było porównanie rozkładów wyników uzyskanych w poszczególnych zadaniach egzaminacyjnych oraz sprawdzenie zgodności punktowania wypracowań. Badanie dostarczyło wielu interesujących i ważnych wniosków, które posłużyły do modyfikacji skal oceniania (Konarzewski, 2011). Na uwagę zasługują wyniki dotyczące zgodności punktowania, które badano dwiema metodami. Pierwsza polegała na oszacowaniu podobieństwa dwóch wyników sumarycznych tego samego wypracowania. Wartość współczynnika korelacji dla wszystkich rozprawek wyniosła 0,67, przy czym należy zauważyć, że na obniżenie tej wartości w sposób znaczny wpłynęły najmniej trafne i rzetelne skale poprawności merytorycznej i językowej (wyniki dotyczące stosowania tych dwóch skal wskazują ze względu na – wydawałoby się – arbitralny charakter oceny błędów rzeczowych

i językowych). Druga metoda badania zgodności punktowania polegała na obliczeniu procentu zgodnych decyzji w poszczególnych skalach częściowych. Najbardziej zgodnie punktowano wypracowania w skalach treści wypracowań: tezy (1,95), uzasadnienia (2,0) oraz kompozycji (1,9), natomiast najmniej zgodnie w skalach poprawności merytorycznej (1,4), języka (1,3) i zapisu (1,3). Problemy z oceną dwóch ostatnich aspektów rozprawki, które nie dotyczą bezpośrednio treści wypracowania, zostały już częściowo omówione w innej publikacji (Dobkowska i Hącia, 2012), warto jednak podkreślić, że część skali dotycząca oceny treści sprawdziła się w pilotażu, a po udoskonaleniu na podstawie wniosków z badania może być rekomendowana do użytku szkolnego i egzaminacyjnego.

Podsumowanie

Przedstawiony w artykule proces ewolucji sposobu oceniania rozprawki dotyczy nie tylko procedury egzaminacyjnej, ale także – a może przede wszystkim – praktyki dydaktycznej. Zastosowanie spójnej metody oceniania dla egzaminu gimnazjalnego i maturalnego może powodować, że uczniowie będą przygotowywać się do obu egzaminów w podobny sposób, a więc – w przypadku matury – utrwaląc umiejętności zdobyte w gimnazjum, rozszerzając je i uzupełniając. Jawność, dostępność i niezmiennność kryteriów oceniania ustalonych na kilka lat przed egzaminem sprawi, że zasady zdawania egzaminu – tak jak w przypadku SAT – będą jasne, a przygotowanie do egzaminu stanie się prostą ścieżką, na której liczyć się będą przede wszystkim pracowitość i sumienność w ćwiczeniu, a nie – jak dotychczas – umiejętność odgadywania intencji egzaminatorów. Uczeń i nauczyciel dokładnie będą wiedzieć, jakie cechy powinna

⁵ Wartości podane w nawiasach stanowią ilorzad faktycznej zgodności punktowania przez zgodność oczekiwaną przy założeniu czysto losowego punktowania.

mieć rozprawka egzaminacyjna, dzięki czemu będą mogli precyzyjnie zaplanować działania dydaktyczne rozwijające umiejętność tworzenia tej formy wypowiedzi pisemnej. Wszystko to daje nadzieję, że zmiany w sposobie oceniania prac egzaminacyjnych nie tylko zapobiegną uczeniu się „pod klucz”, ale będą służyć rozwijaniu niezbędnej (również w rzeczywistości pozaszkolnej) umiejętności argumentowania.

Literatura

- Centralna Komisja Egzaminacyjna (2002). Arkusze z egzaminu gimnazjalnego w zakresie przedmiotów humanistycznych. Pobrano z: http://www.cke.edu.pl/images/stories/Arkusze/gimnazjum_2002/gh_a1_2002_arkusz.pdf
- Centralna Komisja Egzaminacyjna (2010). *Informator o egzaminie gimnazjalnym od roku szkolnego 2011/2012*. Warszawa: Autor.
- Bańkowska, E., Dobkowska, J. i Mikołajczuk, A. (2010). *Skala oceny rozprawki* [Niepublikowany maszynopis]. Warszawa: Instytut Badań Edukacyjnych.
- Dobkowska J. i Hącia A. (2012). Ocena poprawności językowej prac egzaminacyjnych uczniów III klasy gimnazjum: wewnętrznojęzykowe przyczyny trudności w ocenie, wyniki zastosowania skali egzaminacyjnej, zalecenia dla systemu egzaminacyjnego. *Edukacja*, 2(118), 93–117.
- Dolata, R., Putkiewicz, E. i Wilkomirska, A. (2004). *Reforma egzaminu maturalnego: oceny i rekomendacje*. Warszawa: Instytut Spraw Publicznych.
- Instytut Badań Edukacyjnych (2009). *Holistyczna skala oceny rozprawki* [Niepublikowany maszynopis]. Warszawa: Autor.
- Konarzewski, K. (2010). Skala oceny rozprawki. W: *Informator o egzaminie gimnazjalnym od roku szkolnego 2011/2012*. Warszawa: Centralna Komisja Egzaminacyjna.
- Konarzewski K. (2011). *Nowe formy wypracowania maturalnego z języka polskiego. Badanie pilotażowe*. Warszawa: Centralna Komisja Egzaminacyjna.
- Mikołajczuk, A. i Puzynina, J. (red.). (2004). *Wiedza o języku w zreformowanej szkole*. Warszawa: Nowa Era.
- SAT College Board (bdw.) *SAT preparation booklet 2007–08*. Pobrano z: <http://sat.collegeboard.org/practice/writing-sat-essay>