

**Filip Kulon**

Instytut Badań Edukacyjnych w Warszawie

**dr Henryk Szaleniec**

Polskie Towarzystwo Diagnostyki Edukacyjnej

## **Efekt egzaminatora w kontekście atrybutów pisma**

### **Wstęp**

Pomiar umiejętności tworzenia tekstów (pisania) poprzez ocenę ich jakości należy do bardzo złożonych zadań. Egzaminatorzy muszą ocenić to, co jest subiektywne, w tak dalece obiektywny sposób, jak to jest tylko możliwe. Dlatego też, jak pokazują wyniki badań (Breland i in., 1987; Knoch, 2011; Kulon i Żółtak, 2014; Szaleniec i in., 2015; Weigle, 2002), rzetelność i trafność oceny wypracowań z języka ojczystego i języka obcego często są kwestionowane, a efekt egzaminatora jest znaczącym źródłem wariacji wyników. Często oceniający wypracowania nie zgadzają się w ocenie jakości tej samej pracy niezależnie od przyjętej strategii oceniania (analityczne czy holistyczne), co ma istotny wpływ na rzetelność oceniania. Dyskusja dotycząca oceniania wypracowań to także dyskusja na temat trafności pomiaru z wykorzystaniem tekstu tworzonego przez zdających.

Tworzenie własnego tekstu to skomplikowana aktywność człowieka będąca złożeniem procesów poznawczych, kinestetycznych i percepcyjno-mechanicznych (Bonny, 1992; Reisman, 1993). Pisząc wypracowanie, uczeń wykonuje symultanicznie wiele zadań o charakterze poznawczym, obejmujących procesy myślowe dotyczące źródeł i kontekstu przedstawianej treści, planowania sprostania zawartym w poleceniu wymaganiom, kompozycji tekstu, tworzenia tekstu z zastosowaniem reguł gramatyki oraz zasad ortografii i interpunkcji. Ponadto piszący podczas egzaminu samodzielnie monitoruje i ocenia integrację ortograficzno-ruchową (Berninger i in., 1994; Hooper i in., 1993; Jones i Christensen, 1999). Umiejętność pisania rozwijana jest od najmłodszych lat szkoły podstawowej i w przypadku uczniów bez dysfunkcji w pisaniu sam proces pisania staje się z biegiem czasu automatyczny i nie koliduje z procesem twórczego myślenia (Scardamalia, Bereiter i Goelman, 1982). W badaniach prezentowanych w tym opracowaniu autorami prac są maturzyści – uczniowie w wieku 19 lat. Biorąc zatem pod uwagę ich wiek, można przyjąć, że teoretycznie powinni być oni zaawansowani w sztuce pisania, ponieważ faza doskonalenia umiejętności pisania przypada na okres od 16. do 25. roku życia. Umiejętność ta przejawia się w sprawnym ruchu pisarskim, charakteryzującym się ciągłością linii kreującej oraz ekonomicznym sposobem realizacji – przy najmniejszym zużyciu czasu i energii, co objawia się uproszczeniami i modyfikacjami modelunków liter oraz redukcją ich części konstrukcyjnych (Frąszczak, 2013).

W Polsce podczas egzaminów na progach edukacyjnych umiejętność pisania sprawdzana jest (poza przypadkami szczególnymi) bezpośrednio, poprzez ocenę jakości ręcznego zapisu własnego tekstu, tworzonych przez zdającego.

W 2015 roku umiejętności szóstoklasistów z języka polskiego sprawdzane były za pomocą 13 zadań, z czego dwa (12 i 13) sprawdzały bezpośrednio umiejętność pisania. Udzielając odpowiedzi na zadanie 12, uczeń musiał zapoznać się z załączonym w arkuszu egzaminacyjnym komiksem, a następnie miał wykonać polecenie: „Wyjaśnij, dlaczego Zeuksis uznał Parrazjosa za lepszego malarza od siebie”. W zadaniu 13 wymagano od ucznia napisania opowiadania – jego polecenie brzmiało: „Napisz opowiadanie zatytułowane «Tajemnicze pudełko»”.

W przypadku egzaminu gimnazjalnego z języka polskiego arkusz standardowy z 2015 roku zawierał 22 zadania, w tym 20 zadań zamkniętych i 2 zadania otwarte (21 i 22). Zadanie 21 było zadaniem krótkiej odpowiedzi i jego polecenie brzmiało: „Czy każdego celebrytę można nazwać człowiekiem sławnym? Uzasadnij swoją odpowiedź. W argumentacji odwołaj się do tekstów Leszka Kołakowskiego i Mirosława Pęczaka. Nie cytuj sformułowań z arkusza”. Natomiast zadanie rozszerzonej odpowiedzi (22), podobnie jak w przypadku sprawdzianu, polegało na napisaniu opowiadania wg polecenia: „Napisz opowiadanie o podróży, podczas której wydarzyło się coś, co sprawiło, że ktoś stał się sławny”.

Na maturze w nowej formule obowiązującej od 2015 roku dla LO jedno z zadań sprawdzających umiejętność pisania wymagało napisania streszczenia tekstu (Jerzego Bralczyka *Co się nosi w mówieniu?*), a drugie napisania wypracowania na jeden z wybranych tematów. Polecenie do tematu 1 brzmiało: „Wolna wola człowieka czy siły od niego niezależne – co przede wszystkim decyduje o ludzkim losie? Rozważ problem i uzasadnij swoje zdanie, odwołując się do podanego fragmentu *Lalki*, całego utworu Bolesława Prusa oraz innego tekstu kultury”. Temat 2 dotyczył zinterpretowania załączonego w arkuszu egzaminacyjnym wiersza *Ta jedna sztuka* Elizabeth Bishop w tłumaczeniu Stanisława Barańczyka („Zinterpretuj podany utwór. Postaw tezę interpretacyjną i ją uzasadnij”).

Przytoczone zadania sprawdzające umiejętność tworzenia własnego tekstu uczniów kończących odpowiednio drugi, trzeci i czwarty etap edukacyjny oceniane są przez egzaminatorów. Muszą oni jednak najpierw odczytać odręczny zapis przygotowany w odpowiedzi na zawarte w arkuszach egzaminacyjnych polecenia, aby móc go ocenić zgodnie z przyjętymi kryteriami. Wygląd pisma zapewne nie pozostaje bez wpływu na wynik otrzymany przez ucznia.

Peter Cooper (1984) donosi, iż w kilku badaniach wykazano, że wygląd pisma, „ładny zapis”, ma wpływ na ocenę eseju. Lynda Markham (1976, s. 277) podaje, że „analiza wariacji wykazała istotne statystycznie różnice w punktacji wyjaśniane charakterem pisma. Prace napisane ładniej uzyskiwały konsekwentnie wyższe wyniki niż prace cechujące się niską jakością pisma, bez względu na jakość treści”<sup>1</sup>. McColly (1970, s. 153–154) stwierdza:

Czytający mają tendencję zgodnego reagowania na wygląd pisma. Następująca sytuacja jest w jakiś sposób sprzeczna: dla prac napisanych odręcznie, wskaźnik

<sup>1</sup> Tłum. własne.

rzetelności szacowany jako zgodność oceniających będzie nieprawdziwie wysoki poprzez korelację czynnika, jakim jest wygląd pisma. Jednocześnie trafność oceniania jest obniżona, ponieważ umiejętność tworzenia własnego tekstu jest czymś innym niż umiejętność ręcznego pisania<sup>2</sup>.

Powyższe doniesienia były powodem tego, że autorzy artykułu podjęli próbę zbadania, w jakim stopniu atrybuty pisma zdającego (czytelność, wyrobienie i wielkość) będące wyznacznikiem jakości zapisu mogą być źródłem wariacji wyniku egzaminu maturalnego z języka polskiego.

## Metodologia

Próba prac wykorzystana w analizach pochodzi z „Badań porównywalności oceniania i efektu egzaminatora” przeprowadzonych przez Instytut Badań Edukacyjnych w latach 2013–2014. Prace do badania wylosowane zostały z baz Centralnej Komisji Egzaminacyjnej zawierających wyniki egzaminów maturalnych z Okręgowej Komisji Egzaminacyjnej w Jaworznie (2011 rok) i Krakowie (2012 rok). W wymienionych OKE po wybraniu prac przygotowane zostały kopie wypracowań (skany), z których usunięto identyfikatory szkół i uczniów oraz zapisy egzaminatorów oceniających poszczególne prace w trakcie sesji. Sumaryczny wynik dla analizowanych wypracowań może przyjmować wartości 0–54 punktów (uwzględniając kryterium szczególnych walorów pracy). Do badania wybrano prace równomiernie reprezentujące całą dostępną skalę ocen. Bardziej szczegółowy opis doboru próby prac znajdzie Czytelnik w monografii autorów wspomnianego badania (Szaleniec i in., 2015).

Do badania została wylosowana reprezentatywna próba egzaminatorów (z każdej OKE było 29 egzaminatorów i jeden przewodniczący zespołu egzaminatorów) z populacji oceniających prace maturalne z języka polskiego w sesji majowej w 2013 r. Łącznie w badaniu wzięło udział 232 egzaminatorów z języka polskiego i ośmiu przewodniczących zespołów. W podstawowej grupie 841 prac każda z nich była oceniana przez ośmiu egzaminatorów z różnych OKE, dzięki zastosowaniu niekompletnego zrównoważonego schematu blokowego (*Balanced Incomplete Block Design, BIBD*). Szczegółowy opis schematu przydzielania prac (uwzględniający dodatkowe prace i oceny wewnątrz poszczególnych OKE) znajdzie Czytelnik w przywoływanej już monografii (Szaleniec i in., 2015).

W badaniach dotyczących relacji pomiędzy wynikiem oceniania eseju a wyglądem pisma odręcznego do klasyfikacji jakości pisma najczęściej stosowano metodę sędziów kompetentnych do określenia skali czytelności (lub skali cechy definiowanej jako pismo ładne lub brzydkie). Po utworzeniu takiej skali pracom przypisywana była wartość skali lub też prace przepisywane były przez kopistów, których pismo charakteryzowało się daną cechą. Na przykład w często cytowanych badaniach, które prowadziła Lynda Markham (1973), wyselekcjonowane do eksperymentu prace o różnych poziomach jakości merytorycznej zostały przepisane odręcznie z zastosowaniem różnych poziomów jakości pisma (*handwriting styles*) definiowanych jako dobry, średni i słaby.

<sup>2</sup> Tłum. własne.

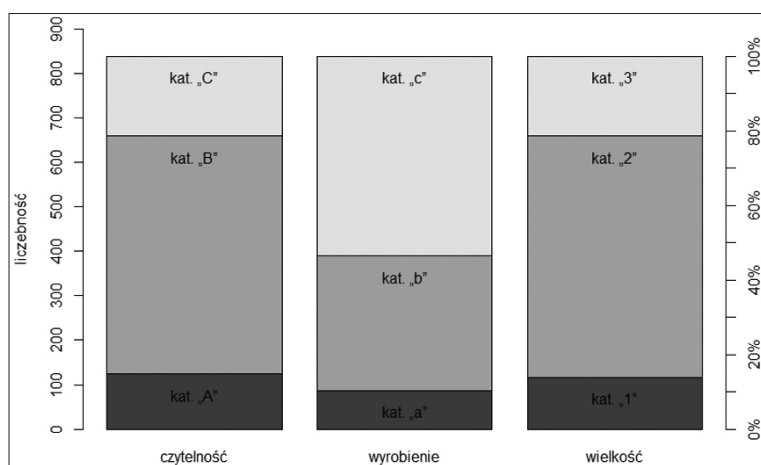
W prezentowanych w artykule badaniach przyjęto inną metodologię tworzenia skali opisującej jakość odręcznego pisma maturzystów. Próba prac wypracowań maturalnych wylosowanych do badań efektu egzaminatora została poddana analizie przez grafologa. W trakcie analizy cech syntetycznych pisma maturzystów ustalono, że najistotniejszymi parametrami, które wpływają na komunikatywność pisma odręcznego, są: (1) wielkość, (2) klasa pisma (wyrobiecie), (3) czytelność (Frąszczak, 2013). Podstawę oceny klasy pisma stanowią m.in. takie cechy jak: stopień płynności linii, rodzaj dominującego impulsu, liczba i rodzaj form powtarzalnych, równomierność nacisku, stopień stabilizacji wielkości kątowych oraz proporcji, rodzaj i standaryzacja uproszczeń. Czytelność pisma oznacza natomiast zrozumiałość, wyrazistość znaków graficznych dla czytającego tekst, zapisany pismem odręcznym przez inną osobę.

W tabeli 1 zamieszczono informacje o kategoriach wydzielonych dla każdego z aspektów pisma, wraz z liczebnością prac zaklasyfikowanych do każdej z nich (spośród 839 użytych w analizie<sup>3</sup>). W związku z bardzo małą liczebnością najbardziej skrajnych kategorii wielkości pisma zdecydowano się na połączenie ich z sąsiednimi kategoriami. Na rysunku 1 przedstawiono, w postaci wykresów słupkowych, proporcje prac w poszczególnych kategoriach aspektów pisma, ostatecznie użytych do analizy. Szczegółowy opis wymienionych kategorii wraz z przykładami próbek pisma z wypracowań maturalnych znajdzie Czytelnik w aneksie.

**Tabela 1. Opis kategorii atrybutów pisma wraz z przydzieloną do nich liczbą prac**

Atrybuty pisma	Kategorie		Liczba prac
wielkość	pismo bardzo duże: ponad 6 mm (I)	pismo duże: ponad 4 mm (1)	9
	pismo duże: 4,1–6 mm (II)		108
	pismo średnie: 2,6–4 mm (III)	pismo średnie: 2,6–4 mm (2)	543
	pismo małe: 1,5–2,5 mm (IV)	pismo małe: do 2,5 mm (3)	166
	pismo drobne: do 1,5 mm (V)		13
klasa (wyrobiecie)	wyrobite (a)		87
	średnio wyrobite (b)		303
	niwyrobite (c)		449
czytelność	czytelne (A)		124
	częściowo czytelne (B)		536
	częściowo nieczytelne (C)		179

<sup>3</sup> W przypadku dwóch prac niemożliwe było ich jednoznaczne przypisanie do odpowiednich kategorii.



Rysunek 1. Liczebność prac z próby w poszczególnych kategoriach aspektów pisma

## Umiejętność tworzenia tekstów i atrybuty pisma

Zastanawiając się nad związkiem atrybutów pisma z efektem egzaminatora, w pierwszej kolejności warto sprawdzić, czy mają one jakikolwiek związek z oceną wypracowań. Choć formalnie nie podlegają one ocenie, mogą wpływać na postrzeganie jakości pracy przez egzaminatora. W ten sposób, nawet nieświadomie, egzaminator może przydzielić uczniowi ocenę obciążoną błędem.

W związku z tym, że w opisywanych danych każda praca została oceniona ośmiokrotnie, ostateczna ocena może zostać obliczona w różny sposób (Szaleniec i Kulon, 2014). Jednym z nich jest oczywiście **średnia ocen wszystkich egzaminatorów**. Zastosowany w opisywanych badaniach hierarchiczny model efektu oceniającego oparty na teorii detekcji sygnałów (*hierarchical rater model with signal detection theory*, HRM-SDT) dostarcza jednak ocen „prawdziwych”, które można wykorzystać w dalszych analizach (Szaleniec i in., 2015). Korelacja tych dwóch miar<sup>4</sup> jest bardzo wysoka, współczynnik korelacji Pearsona wynosi 0,99 (SE = 0,005). Oceny „prawdziwe” otrzymane w wyniku użycia HRM-SDT powinny być jednak lepszym oszacowaniem od średniej pojedynczych ocen, gdyż model ten uwzględnia istnienie różnych efektów oceniającego. Oceny te powinny być zatem pozbawione wpływu efektów egzaminatora, w przeciwieństwie do średniej faktycznych ocen przyznanych przez egzaminatorów (Kulon, 2014).

Do analizy związku atrybutów pisma i oceny wypracowania wykorzystając metodę regresji liniowej, która w dość prosty sposób zilustruje te zależności. Dysponując dwoma miarami ostatecznej oceny wypracowania, warto użyć obydwu tych miar jako zmiennych zależnych i porównać wyniki. Jako zmiennych niezależnych w obydwu przypadkach użyto wskaźników poszczególnych atrybutów pisma kodowanych zerojedynkowo. Jako kategorie odniesienia we wszystkich przypadkach wybrano kategorie środkowe,

<sup>4</sup> Oceny „prawdziwe” zostały obliczone jako średnia z łańcucha Markowa (por. Szaleniec i Kulon, 2014).

czyli odpowiednio: dla czytelności kategoria „B”, dla wyrobienia kategoria „b”, a dla wielkości kategoria „2”. Taki wybór jest podyktowany tym, że te kategorie można uznać za przeciętne, a w przypadku czytelności i wielkości pisma są to również kategorie najbardziej liczne (por. rys. 1).

**Tabela 2. Współczynniki modeli regresji liniowej atrybutów pisma na ostateczną ocenę pracy**

		Ostateczna ocena szacowana jako							
		średnia ocen indywidualnych				oceny „prawdziwe”			
		b	SE	t	p	b	SE	t	p
wyraz wolny		<b>27,232</b>	0,742	36,721	0,000	<b>27,824</b>	0,845	32,911	0,000
czytelność	kat. „A”	0,288	1,083	0,266	0,790	0,511	1,234	0,414	0,679
	kat. „C”	<b>-3,212</b>	0,948	-3,389	0,001	<b>-3,583</b>	1,081	-3,316	0,001
wyrobienie	kat. „a”	<b>3,747</b>	1,312	2,855	0,004	<b>4,404</b>	1,496	2,944	0,003
	kat. „c”	<b>-3,978</b>	0,820	-4,850	0,000	<b>-4,653</b>	0,935	-4,977	0,000
wielkość	kat. „1”	<b>-3,176</b>	1,098	-2,894	0,004	<b>-3,388</b>	1,251	-2,708	0,007
	kat. „3”	1,073	0,938	1,144	0,253	1,347	1,069	1,260	0,208

Pogrubiono wartości współczynników istotnych statystycznie przy  $p < 0,05$ .

Wszystkie atrybuty pisma posiadają istotny statystycznie związek z ostateczną oceną prac z próby badawczej, niezależnie od tego, czy za ostateczną ocenę przyjmujemy ocenę „prawdziwą” z HRM-SDT, czy średnią pojedynczych ocen egzaminatorów. Dla ocen „prawdziwych” można zaobserwować nieco wyższe efekty niż dla średniej z ocen przyznanych przez egzaminatorów. Prace zaklasyfikowane pod względem czytelności do kategorii „A” nie mają istotnie różnych wyników niż te z kategorii „B”, lecz dla prac z kategorii „C” wyniki są wyraźnie niższe. W przypadku wyrobienia prace zaklasyfikowane zarówno do kategorii „a”, jak i kategorii „c”, uzyskały wyniki wyraźnie różne w stosunku do kategorii „b”. Prace o piśmie mniej wyrobionym otrzymują niższe oceny, natomiast o piśmie bardziej wyrobionym – wyższe. W porównaniu z pracami charakteryzującymi się pismem średnim, te z większym pismem są oceniane istotnie statystycznie niżej. Prace o piśmie małym nie różnią się statystycznie istotnie wynikami od prac z pismem średnim. Współczynniki omawianych modeli regresji (wraz z wielkością błędu i statystyki testowej) zawiera tabela 2.

Trudno jednoznacznie przypisać zaobserwowane efekty jedynie charakterowi pisma. Być może uczniowie o wyższym poziomie umiejętności tworzenia tekstów charakteryzują się pismem wyrobionym i czytelnym. Prawdopodobnie rozwijaniu tych umiejętności (pod warunkiem ręcznego pisania) towarzyszy również poprawa poszczególnych atrybutów pisma. Im więcej tekstów komponują uczniowie, tym ich pismo staje się bardziej wyrobione, a z czasem poprawie może ulec także jego czytelność. Wielkość pisma nie powinna jednak ulegać zmianom w miarę ćwiczenia w pisaniu, zatem powyższe rozumowanie zawodzi w przypadku tego atrybutu.



## Atrybuty pisma i efekt egzaminatora

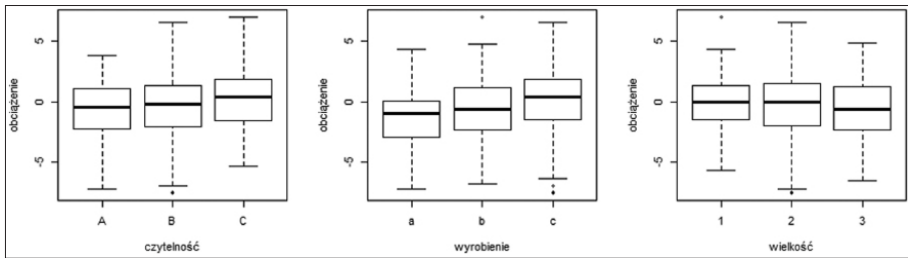
Obliczając różnicę średniej ocen przydzielonych przez egzaminatorów i oceny „prawdziwej” z HRM-SDT, możemy oszacować obciążenie (*bias*) ocen przyznanych każdej pracy. Dodatkowo wartości obciążenia oznaczają, że **średnio egzaminatorzy przyznają wyższe oceny niż ocena „prawdziwa”, a ujemne – niższe oceny egzaminatorów w stosunku do ocen „prawdziwych” z modelu.** Związek poszczególnych atrybutów pisma i błędów popełnianych przez egzaminatorów został oszacowany w modelu regresji liniowej, gdzie obciążenie było zmienną zależną. Poszczególne atrybuty pisma zostały wprowadzone do modelu jako zmienne niezależne w sposób analogiczny jak w poprzednich modelach. Współczynniki takiego modelu prezentuje tabela 3.

**Tabela 3. Współczynniki modelu regresji liniowej atrybutów pisma na obciążenie pojedynczych ocen pracy**

		b	SE	t	p
wyraz wolny		-0,592	0,165	-3,594	0,000
czytelność	kat. „A”	-0,223	0,241	-0,926	0,355
	kat. „C”	0,371	0,211	1,759	0,079
wyrobienie	kat. „a”	<b>-0,657</b>	0,292	-2,255	0,024
	kat. „c”	<b>0,675</b>	0,182	3,704	0,000
wielkość	kat. „1”	0,212	0,244	0,869	0,385
	kat. „3”	-0,274	0,208	-1,317	0,188

Pogrubiono wartości współczynników istotnych statystycznie przy  $p < 0,05$ ; pochyłono przy  $p < 0,1$ .

W przypadku czytelności i wyrobienia pisma widoczny jest ich związek z obciążeniem ocen przyznanych przez egzaminatorów. Obciążenie prac o czytelności zaklasyfikowanej do kategorii „A” nie różni się statystycznie istotnie od tego dla prac z kategorii „B”, analogicznie jak we wcześniejszych modelach regresji (na ostateczne oceny). Prace z kategorii „C” są jednak oceniane przez egzaminatorów o średnio 0,37 punktu wyżej niż ich ocena „prawdziwa” (przy 90% poziomie ufności). W porównaniu z kategorią „b” na skali wyrobienia, prace zaklasyfikowane do kategorii „a” są przez egzaminatorów oceniane o 0,66 punktu niżej niż ich ocena „prawdziwa” z HRM-SDT. Prace zaklasyfikowane do kategorii „c” są natomiast oceniane średnio o 0,68 punktu wyżej niż ocena „prawdziwa” w porównaniu z kategorią „b”. Co ciekawe, wielkość pisma zdaje się nie mieć wpływu na oceny egzaminatorów. Na rysunku 2 przedstawiono wykresy skrzynkowe obciążenia w podziale na kategorie poszczególnych atrybutów pisma.



**Rysunek 2. Wykresy skrzynkowe obciążenia w podziale na kategorie poszczególnych atrybutów pisma**

Obciążenie ocen przyznanych przez egzaminatorów można powiązać głównie z efektem ich łagodności lub surowości. Obrazuje ono błędy systematyczne polegające na przydzielaniu ocen wyższych lub niższych niż oceny „prawdziwe”. W celu zbadania, czy atrybuty pisma mają również związek z tak istotnym efektem oceniającego jak precyzja oceniania, konieczne jest zastosowanie innej miary. Precyzja oceniania jest związana z błędami losowym i objawia się zmianami w rzetelności przyznawania ocen przez egzaminatorów. Miara precyzji w omawianym przypadku może być odchylenie standardowe pojedynczych ocen przyznanych danej pracy. Niskie odchylenie standardowe rzeczywistych ocen egzaminatorów jest objawem dużej precyzji, a wysokie odchylenie – małej precyzji. Trudno ustalić normy dla takiej miary precyzji, gdyż mogą one zależeć od specyfiki mierzonej cechy. W omawianym przypadku nie jest to jednak konieczne, ponieważ interesujące jest to, czy wielkość odchylenia różni się dla prac z poszczególnych kategorii analizowanych atrybutów pisma. W tym celu znów można posłużyć się modelem regresji liniowej, tym razem z odchyleniem standardowym poszczególnych ocen jako zmienną zależną i tym samym zestawem zmiennych niezależnych, jak do tej pory. W tak wyspecyfikowanym modelu regresji żaden ze współczynników dla poszczególnych kategorii atrybutów pisma nie jest istotny statystycznie. Można zatem przyjąć, że charakterystyka pisma analizowanych prac nie ma związku z precyzją oceniania. Średnie odchylenie standardowe ocen prac wynosi 4,87 (SE = 0,055) i nie różni się istotnie statystycznie dla poszczególnych kategorii opisywanych atrybutów pisma.

W celu potwierdzenia powyższych wniosków dotyczących zależności precyzji oceniania i atrybutów pisma użyto innej metody do analizy tych samych danych niż analiza regresji. Dla grup prac wyznaczonych przez kategorie wszystkich opisywanych atrybutów pisma obliczono współczynniki korelacji wewnątrzklasowej (*intraclass correlation coefficient*, ICC; Shrout i Fleiss, 1979). Jest to oparty na analizie wariancji współczynnik często używany w celu obliczenia zgodności lub spójności wielokrotnych ocen. Jego wartość równa 0 oznacza całkowity brak zgodności pomiędzy ocenami, a wartość 1 – pełną zgodność. Analizowane dane dotyczą sytuacji, kiedy każda praca jest oceniana przez inny zestaw ośmiu egzaminatorów wybranych losowo z populacji wszystkich egzaminatorów. Ponadto, zainteresowanie skupia się na zgodności ocen pojedynczych egzaminatorów, zatem odpowiednim współczynnikiem jest ICC(1, 1) (Shrout i Fleiss, 1979). Granice 95% przedziałów ufności dla współczynników z wszystkich kategorii poszczególnych atrybutów pisma przedstawiono w tabeli 4. Żaden z przedziałów



ufności ICC nie jest rozłączny z innym, zatem wyznaczone charakterystyki pisma nie prowadzą do obniżenia precyzji oceniania. Analiza z użyciem ICC potwierdza wnioski płynące z wcześniejszej analizy z użyciem regresji liniowej.

**Tabela 4. Współczynniki korelacji wewnątrzklasowej dla pojedynczych ocen egzaminatorów**

	Czytelność			Wyrobieenie			Wielkość		
	kat. „A”	kat. „B”	kat. „C”	kat. „a”	kat. „b”	kat. „c”	kat. „1”	kat. „2”	kat. „3”
dolna granica	0,79	0,81	0,75	0,70	0,80	0,79	0,77	0,80	0,79
górnna granica	0,87	0,85	0,82	0,82	0,85	0,84	0,86	0,84	0,86

## Podsumowanie

Chociaż coraz częściej w codziennym życiu, gdy chcemy coś napisać, sięgamy po klawiaturę, umiejętność pisania odręcznego jest nadal bardzo istotna. Dlatego też w wielu systemach edukacyjnych umiejętność tworzenia tekstów sprawdzana jest na egzaminie bezpośrednio poprzez ocenę tekstu napisanego ręcznie przez zdającego. Dzieje się tak pomimo większych kosztów takiego sprawdzania w porównaniu z ocenianiem pośrednim za pomocą zadań zamkniętych i pomimo obciążenia wyników efektem egzaminatora. Efekt ocenającego jest nierozzerwalnie związany z procesem oceniania, w którym wartościowania jakości pracy dokonuje człowiek. Nie da się go całkowicie wyeliminować. Konieczne natomiast jest poznanie jego natury i dążenie do jego minimalizacji w celu zwiększenia rzetelności procesu oceniania. Poznanie, jak cechy charakterystyczne pisma odręcznego wpływają na ten efekt, jest jednym z działań prowadzących do tego celu.

Zastanawiać może uzyskany w prezentowanych badaniach wynik, różny od rezultatów otrzymanych we wcześniejszych badaniach nad atrybutami pisma (Bull i Stevens, 1979; Chase, 1986; Huck i Bounds, 1972; Markham, 1976; Marshall i Powers, 1969). Warto jednak podkreślić inny sposób wyboru atrybutów pisma, które stanowiły zmienne niezależne w prezentowanych badaniach. Wielkość pisma, wyrobienie i czytelność są czym innym niż wygląd pisma oceniany holistycznie na skali: bardzo ładne, średnie i nieładne pismo (por. Markham, 1976). Otrzymane wyniki są jednak zgodne z badaniami, które wykazują, że ocenający mają tendencję oceniać niżej prace o takiej samej zawartości prezentowane jako wydruk komputerowy, a więc bardzo czytelne (Graham, Harris i Hebert, 2011; Powers i in., 1994; Russell i Haney, 1997; 2000; Szaleniec i Kulon, 2014). Steve Graham i współpracownicy (2011) sugerują, że prawdopodobna przyczyna tego zjawiska wiąże się z łatwiejszym dostrzeganiem błędów w piśmie czytelnym niż w mało czytelnym. Ponadto w polskich egzaminach (jak zapewne w innych krajach) w przypadkach wątpliwości egzaminator ma obowiązek podjąć decyzję na korzyść zdającego.

Choć przeprowadzone analizy wskazują na powiązanie atrybutów pisma z efektem łagodności/surowości ocenających, nie ujawniono takiego związku z precyzją oceniania. Jest to istotna informacja, gdyż oznacza, że ocenający

spójnie przyznają oceny wyższe lub niższe (w zależności od atrybutów pisma) niż odpowiednio do poziomu umiejętności uczniów, lecz rzetelność pomiaru nie ulega obniżeniu. Jest to potwierdzeniem otrzymywanych wcześniej wyników (McColly, 1970) i stanowi dobrą wiadomość, choć jednocześnie rodzi pytanie o trafność dokonywanej oceny.

## Bibliografia

1. Berninger, V. W., Cartwright, A. C., Yates, C. M., Swanson, H. L. i Abbott, R. D. (1994). *Developmental skills related to writing and reading acquisition in the intermediate grades*. Reading and Writing: An Interdisciplinary Journal, 6(2), 161–196.
2. Bonny, A. M. (1992). *Understanding and assessing handwriting difficulties: Perspective from the literature*. Australian Occupational Therapy Journal, 39(3), 7–15.
3. Bull, R. i Stevens, J. (1979). *The effects of attractiveness of writer and penmanship on essay grades*. Journal of Occupational Psychology, 52(1), 53–59.
4. Breland, H., Camp, R., Jones, R. J., Morris, M. M. i Rock, D. A. (1987). *Assessing writing skill*. New York, NY: College Entrance Examination Board.
5. Chase, C. I. (1986). *Essay test scoring: Interaction of relevant variables*. Journal of Educational Measurement, 23(1), 33–41.
6. Cooper, P. L. (1984). *The assessment of writing ability: A review of research*. GRE Board Research Report GREB No. 82-15R, ETS Research Report 84-12. Princeton, NJ: Educational Testing Service.
7. Frąszczak, J. (2013). *Czytelność odręcznego tekstu wypracowań z egzaminu maturalnego z języka polskiego*. Analiza na zamówienie IBE. Maszynopis w archiwum Zespołu Analiz Osiągnięć Uczniów.
8. Graham, S., Harris, K. i Hebert, M. A. (2011). *Informing writing: The benefits of formative assessment*. A Carnegie Corporation Time to Act report. Washington, DC: Alliance for Excellent Education.
9. Hooper, S. R., Swartz, C., Montgomery, J., Reed, M. S., Brown, T., Wasileski, T. i Levine, M. D. (1993). *Prevalence of writing problems across three middle schools samples*. School Psychology Review, 22(4), 610–622.
10. Huck, S. W. i Bounds, W. G. (1972). *Essay Grades: An Interaction between Graders' Handwriting Clarity and the Neatness of Examination Papers*. American Educational Research Journal, 9(2), 279–283.
11. Jones, D. i Christensen, C. (1999). *The relationship between automaticity in handwriting and students' ability to generate written text*. Journal of Educational Psychology, 91(1), 44–49.
12. Knoch, U. (2011). *Rating scales for diagnostic assessment of writing: what should they look like and where should the criteria come from?* Assessing Writing, 16(2), 81–96.
13. Kulon, F. (2014). *Modele analizy efektu oceniającego w pomiarze edukacyjnym*. Edukacja, 128(3), 95–111.
14. Kulon, F. i Żółtak, M. (2014). *Zróżnicowanie łagodności egzaminatorów między okręgowymi komisjami egzaminacyjnymi*. [w:] B. Niemiernko i M.K. Szmigel (red.). *Diagnozy edukacyjne. Dorobek i nowe zadania* (s. 142–150), Kraków: gRUPA TOMAMI.
15. Markham, L. R. (1976). *Influences of handwriting quality on teacher evaluation of written work*. American Educational Research Journal, 13(4), 277–283.

16. Marshall, J. C. i Powers, J. C. (1969). *Writing neatness, composition errors, and essay grades*. Journal of Educational Measurement, 6(2), 97–101.
17. McColly, W. (1970). *What does educational research say about the judging of writing ability?* The Journal of Educational Research, 64(4), 148–156.
18. Powers, D., Fowles, M., Farnum, M. i Ramsey, P. (1994). *Will they think less of my handwritten essay if others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays*. Journal of Educational Measurement, 31(3), 220–233.
19. Reisman, J. E. (1993). *Development and reliability of the research version of the Minnesota Handwriting Test*. Physical and Occupational Therapy in Pediatrics, 13(2), 41–55.
20. Russell, M. i Haney, W. (1997). *Testing Writing on Computers*. Education Policy Analysis Archives, 5(3). doi:<http://dx.doi.org/10.14507/epaa.v5n3.1997>
21. Russell, M. i Haney, W. (2000). *Bridging the Gap between Testing and Technology in Schools*. Education Policy Analysis Archives, 8(19). doi:<http://dx.doi.org/10.14507/epaa.v8n19.2000>
22. Scardamalia, M., Bereiter, C. i Goelman, H. (1982). *The role of production factors in writing ability*. W M. Nystrand (red.), *What writers know: The language, process, and structure of written discourse* (s. 173–210). New York: Academic Press.
23. Shrout, P. E. i Fleiss, J. L. (1979). *Intraclass correlations: uses in assessing rater reliability*. Psychological bulletin, 86(2), 420–428.
24. Szaleniec, H., Kondratek, B., Kulon, F., Pokropek, A., Skórska, P., Świsł, K., Wołodźko, T. i Żóltak, M. (2015). *Efekt egzaminatora w ocenianiu prac maturalnych z języka polskiego i matematyki*. Warszawa: Instytut Badań Edukacyjnych.
25. Szaleniec, H. i Kulon, F. (2014). *Wpływ rodzaju pisma (ręczne vs komputerowe) na ocenę wypracowania maturalnego z języka polskiego*. [w:] B. Niemierko i M.K. Szmigiel (red.), *Diagnozy edukacyjne. Dorobek i nowe zadania* (s. 75–88), Kraków: gRUPA TOMAMI.
26. Weigle, S. C. (2002). *Assessing writing*. Cambridge, England: Cambridge University Press.

## Aneks

### Opis kategorii jakości pisma odręcznego zastosowanych w badaniach

#### Czytelność:

- A. Pismo czytelne w aspekcie wszystkich desygnatów brzmieniowych i znaczeniowych. Konstrukcje znaków graficznych zbliżone do wzorców elementarnych i drukowych, z nielicznymi modyfikacjami modelunków znaków graficznych – wynikającymi z ekonomiki kreślenia.
- B. Pismo czytelne w aspekcie desygnatów brzmieniowych i znaczeniowych, ale zawierające uproszczenia i modyfikacje konstrukcji graficznych, które jednakże nie utrudniają ich prawidłowej interpretacji.
- C. Pismo częściowo nieczytelne – zawierające uproszczenia, modyfikacje i redukcje poszczególnych konstrukcji liter, części i całych wyrazów oraz ciągów wyrazowych, powodujące trudności w ich jednoznacznej interpretacji brzmieniowej i znaczeniowej.

Wyrobienie<sup>5</sup>:

- Pismo wyrobione, charakteryzujące się między innymi dużą płynnością linii graficznej, subtelnym cieniowaniem pisma, swoistym rozwiązaniem budowy znaków oraz znaczną różnorodnością elementów powtarzalnych.
- Pismo średnio wyrobione, charakteryzujące się między innymi umiarkowaną płynnością linii graficznej, zauważalnym cieniowaniem pisma, pojawiającymi się niekiedy swoistymi rozwiązaniami budowy znaków, średnim bogactwem elementów powtarzalnych i średnim tempem pisania.
- Pismo niewyrobione, charakteryzujące mało zróżnicowanym cieniowaniem, mało płynnym przebiegiem linii i wolnym tempem kreślenia.

Wielkość:

- pismo bardzo duże – ponad 6 mm
- pismo duże – od 4,1 mm do 6 mm
- pismo średnie – od 2,6 mm do 4 mm
- pismo małe – od 1,5 mm do 2,5 mm
- pismo drobne – do 1,5 mm

Wybrane przykłady fragmentów pisma odpowiadającego poszczególnym kategoriom pisma (na podstawie: Frąszczak, 2013)

WYPRAWOWANIE

na temat nr ...2.....

Miły to miła, prosta dziewczyna  
& bolebały, lecz w oczach nieletomych  
ludzi mił jest takim amótkiem, jakby  
się mogło wyłowić. Ma tylko miłkę,  
Inaczej jako siurpca w mieszanostwie  
dokumenty, widać się w momencie z  
kierownikiem. Jego stosunek do rodziny  
co chwila się zmienia, ponieważ tego,  
je one cały czas jest taka sama.  
Dzisiaj jest mił, co mił obco, widać  
siurpca, która spełnia mił dla  
mił, mił znaczą, a jej obecność

Przykład wypracowania charakteryzującego się pismem w pełni czytelnym (A), wyrobionym (a) i dużym. Reprezentuje ono w przewadze modelunki elementarne i nie stwarza żadnych trudności w odczytaniu podczas oceniania prac.

<sup>5</sup> Na podstawie: Słownik Terminów Pismoznawczych, <http://prawouam-stp.home.amu.edu.pl/>.

## WYPRACOWANIE

na temat nr 1

Adam Mickiewicz stworzył III cz. „Dziadła” w okresie romantyzmu. Był to dramat. Samo słowo dramat powstało już w Antyku, w VI w. p.n.e. Gdy Grecy po raz pierwszy zaczęli występować dramaty na scenie.

W prefacyjnym fragmencie III cz. „Dziadła” mamy przedstawioną, lewą stronę (do której zaliczamy urzędników, radców, registratorów, domy, mezczyzn) i prawą stronę (do której zaliczamy mezczyzn, senatora, gubernatora, starostę, gubernatorów). Lewa strona jest

Przykład wypracowania charakteryzującego się pismem czytelnym (A) o małym wyrobieniu (c). Pismo ma elementarzowe konstrukcje liter i mimo braku wyrobienia może być jednoznacznie interpretowane w aspekcie zawartej treści.

Egzamin maturalny z języka polskiego  
Poziom podrzędny

nie wystraszony w miły strachu się powiem  
była potrzebna nie jest to to samo co w teatrze  
niektóre napisanie. Niekiedy widziałem że para się  
rozważają światem istniejącymi i światem zmagającym  
z którego pochodzi nasz codzienny świat. To podziw  
iżnawo przeproszono postawę miłego drzewka, młotów, domy,  
obrony tego że świat, świat iżnawo u mnie w teatrze  
światu jednakże przez światła napisanie światu w  
niektórych napisanie światu gdy same światła napisanie  
wielko tego napisanie. Na końcu z „miłego”  
światu była napisanie. Niekiedy przez światła że  
niektórych napisanie napisanie, ale napisanie to  
niektórych napisanie napisanie światu iżnawo

Pismo małe (IV), częściowo czytelne (B) i średnio wyrobione (b) z lekką deformacją konstrukcji, charakteryzujące się bazowaniem linii podstawowej pisma nad liniamentem. Jest ono jednakże czytelne w ciągach wyrazowych, w kontekście treści danego zdania.



Egzamin maturalny z języka polskiego  
Poziom podstawowy

Isabelle Leica w „Lalce” Bolesława Prusa  
jest panną zbijaną bardzo bezbożną,  
nie dbającą o nic, jest piękna,  
kochała, nie było dla niej rzeczy niemożliwych.  
Kiedy ją ustraszono na kardynal karku,  
karmu ustraszono się, nie interakowała.  
Zawsze była straszona i od czasu do czasu  
nie była straszona, kiedy ktoś się zabawał  
panną, która była wystraszona i  
parcia podziwiała, przywołała im panienki  
choćby była bardzo kochała, i była  
bezbożnością to szanowała ludzi co

Pismo częściowo czytelne (B), niewyrobane (c) i bardzo duże (I), zawierające uproszczone konstrukcje znaków graficznych, które jednakże są czytelne w graficznej kompozycji wyrazowej i w wierszach tekstu.

**WYPRACOWANIE**

na temat nr 7

Przeanalizuj fragment powieści z III części: „Przebieg  
Adama Mickiewicza”, która została napisana w czasie  
romantyzmu, gdy Polacy byli pod zaborami, a sam  
Mickiewicz przeżywał parę zyczeń.  
Początek dzieła się w powieści o autorze  
Nowickim, na romantycznym przesłaniu, który  
sam gospodarz będący w centrum uwagi, siłami  
nie bawi, trafia w rytm myśli. Po obu jego  
stronach dotykamy motywu góry, które, które do-  
-szereżają Nowickim.  
Po lewej stronie autoru skupili się  
wiedzący i wiedzący, którzy cały czas wyobrażają

Przykład pracy, w którym pismo jest częściowo nieczytelne (C) i średnio wyrobione (b). Jest wprawdzie trudne w odczytaniu, ale w konsekwencji czytelne. Zwraca uwagę nierówny dukt pisma, oraz ostrość grammowa<sup>6</sup> spowodowana gwałtownymi ruchami kreującymi.

<sup>6</sup> Gramma to konstrukcyjnie wyodrębniona część znaku graficznego. Wśród liter pisanych zgodnie ze wzorcem elementarnym wyróżnia się litery jednogrammowe (np. c, o, s), dwugrammowe (np. a, n, p, u) i trójgrammowe (m) (Słownik Terminów Pismoznawczych, <http://prawouam-stp.home.amu.edu.pl/>).



WYPACOWANIE

na temat nr 9

Obawę Justyna i jego przyjaciół z Zespołu  
wzrostu intelektualnego, oraz humanistycznego i politycznego  
z widocznymi spójnikami oraz innymi znakami

Justyna jest postacią, która bardzo lubi  
nie tylko swoją historię, ale także swoje  
przekazywanie. Oni także nie chcą przekazywać  
swoich historii innym. Jedni z nich z zespołu z  
Zespołu

Justyna i jego przyjaciele bardzo lubią  
Justyna i jego przyjaciele z zespołu, ogólnie  
mają taki sam poziom de faktu

Pismo częściowo nieczytelne (C), niewyrobane (c), wskazuje na brak wyrobionego ruchu kreującego oraz na zaburzenia motoryczne piszącego. W dużej części wyrazów występuje całkowita destrukcja modelunków poszczególnych liter.