

STANISŁAW UBERMANOWICZ
Uniwersytet im. Adama Mickiewicza
w Poznaniu

EWALUACJA NOWEJ MATURY

ABSTRACT. Ubermanowicz Stanisław, *Ewaluacja Nowej Matury* (Evaluation of New "Matura" Exam). „Neodidagmata” XXIV, Poznań 1999, Adam Mickiewicz University Press, pp. 73–91. ISBN 83-232-0956-1. ISSN 0077-653X.

In the frame of the Polish educational reform there are some attempts undertaken to improve the final exam of secondary school. The reformers' proposals undergo an outside analysis and critique. The author presents two models of democratic evaluation of the so-called in the "matura" exam. One model is a qualitative analysis of the criteria of grading essays in the subject of the Polish language. The other model is a quantitative study of qualitative differences between problem tasks within the exam in mathematics. Both naturalistic approaches are based on evaluating opinions of the subjects of the „matura”, i.e. students who take the exam and teachers who grade the output.

*Stanisław Ubermanowicz, Zakład Technologii Kształcenia Wydziału Studiów Edukacyjnych, Uniwersytet im. Adama Mickiewicza, ul. Słowackiego 20, 60-823 Poznań, Polska-Poland.
e-mail: uberman@main.amu.edu.pl*

W ramach reformy oświaty w Polsce prowadzone są systematyczne prace nad doskonaleniem egzaminu maturalnego. Propozycje reformatorów, skupionych w makroregionach wokół programu Nowa Matura, poddawane są ciąglej ewaluacji wewnętrznej i zewnętrznej. Najistotniejsza weryfikacja innowacji i dokonań zespołów przedmiotowych odbywa się co roku podczas eksperymentu zwanego „Próbną maturą”. W trakcie próbnej matury nauczyciele i uczniowie nie tylko zapoznają się z nowymi mechanizmami egzaminowania, lecz także wypełniają ankiety służące do oceny jakości i stopnia akceptacji proponowanych zmian. Znamienne dla tego eksperymentu jest bardzo szerokie zaangażowanie społeczne, zarówno reformatorów, jak i respondentów. Z jednej strony owocuje zdrowa rywalizacja między regionami, z drugiej napływa ze szkół wiele cennych wniosków optymalizacyjnych. Rolą autora, jako ewaluatora zewnętrznego, było przetworzenie części zebranego materiału empirycznego. Chociaż przygotowane na tej podstawie raporty przeznaczone były dla

specjalistów przedmiotowych, sędzę, że zainteresują one również czytelników „Neodidagmaty”.

W artykule przedstawiam dwa modele demokratycznej ewaluacji matury. Pierwszy model jest jakościową analizą kryteriów oceniania wypracowań z języka polskiego, opartą na studium przypadku. Drugi model jest ilościowym wyznaczaniem różnic w jakości czynników merytorycznych i organizacyjnych, wpływających na ogólną ocenę egzaminu w kolejnych trzech latach. Oba preferowane w środowisku poznańskim podejścia naturalistyczne oparte są na badaniu opinii głównych podmiotów matury, to jest nauczycieli oceniających prace i uczniów zdających „egzamin dojrzałości”.

OCENIANIE WYPRACOWAŃ Z JĘZYKA POLSKIEGO

Ważnym nurtem prac nad doskonaleniem matury jest przygotowanie rzetelnych kryteriów oceniania. W powszechnej opinii najtrudniejsze zadanie stoi przed polonistami, gdyż wypracowania maturalne z języka polskiego są niepowtarzalnymi (arcy)dzielami. Pomocne w optymalizacji kryteriów okazać się mogą badania sondażowe towarzyszące próbnej maturze. Zebrane w 1997 roku otwarte wypowiedzi nauczycieli współuczestniczących w maturze stanowią cenny materiał badawczy przydatny do opracowania jakościowego. Niniejsza praca stanowi syntezę opinii i autorską ocenę tych wskazanych przez respondentów zagadnień, które w przedstawionych przez reformatorów kryteriach trzeba jeszcze udoskonalić. To właśnie owe kwestie powinny bezwzględnie znaleźć się w przyszłorocznej ankiecie, ażeby nauczyciele mogli pomóc w rozstrzygnięciu wątpliwości i w podjęciu właściwych decyzji.

GLEBIA PROBLEMÓW

Podstawową metodą rozwiązywania skomplikowanych problemów jest rozłożenie danego zagadnienia na czynniki prostsze i zagłębienie się z osobna w każdy z komponentów. Taką drogę obrali twórcy „Arkusza oceny i recenzji”, proponując nauczycielom szczegółowe narzędzie obiektywizujące ocenianie pisemnych prac z języka polskiego. Narzędzie to poddane zostało weryfikacji i zaopiniowaniu. Z przeprowadzonego sondażu wynika, że ogólna filozofia kryteriów została zaaprobowana, jednakże duża ilość konstruktywnych uwag krytycznych wskazuje na potrzebę dalszej optymalizacji założeń. Okazało się przy tym, że respondenci łatwo dostrzegają wielość aspektów oceniania, lecz niekiedy ujmują poszczególne wątki zbyt płytko, tylko w jednym wymiarze. Jako pożywkę do dyskusji zacytuję w całości znamienne wypowiedzi:

„Zaproponowany «arkusz» mógłby stać się pomocny przy poprawianiu i ocenianiu prac maturalnych, gdyby dołączone do niego kryteria nie prowadziły do niebezpiecznego obniżenia wymagań intelektualnych. Jeśli zgodzimy się pochopnie na określenie mianem istoty dojrzałej człowieka, który nie potrafi samodzielnie sformułować argumentów, a zastępuje je sprawozdaniem, nie wybiera prawidłowo tekstów zgodnych z tematem, bezmyślnie powiela stereotypowe sądy, rozumie tylko dosłowny sens dzieła literackiego, posługuje się niewielką ilością słów i konstrukcji składniowych, w wyraźny sposób narusza reguły języka ojczystego, a stylu potocznego nie potrafi odróżnić od kulturalnej odmiany polszczyzny, to hasło «dojrzały» w słowniku wyrazów bliskoznacznych trzeba będzie wkrótce wzbogacić o kolejny synonim: «mierny». Natomiast wzorcem osobowym naszych przyszłorocznych maturzystów stanie się niewątpliwie Edek, postać występująca w dramacie «Tango» S. Mrożka”.

Powyższy tekst w bezpośrednim przesłaniu jest krytyką niskiego poziomu językowych kompetencji maturzystów. Co do takiego stanu jesteśmy zgodni. Jednak dodatkowo przewija się żal, iż matura oparta o łagodniejsze kryteria oceniania przestanie być egzaminem dojrzałości. Tych obaw ja nie podzielam. Jeśli bowiem zakłada się powszechność szkoły średniej, a matura z języka polskiego jest warunkiem podjęcia studiów, to egzamin winien być dopasowany do możliwości szerszej niż dotąd grupy uczniów. Nie wolno całej populacji serwować zbyt trudnego sprawdzianu z umiejętności bardziej literackich niż językowych. Ambiwalentne traktowanie uczniów o miernych zdolnościach pisarskich jest sprzeczne z humanistycznym duchem przedmiotu. Najbardziej jednak zdumiewa to, iż w cytacie jednym tchem wymienia się kryteria z poziomu bardzo dobrego jako konieczne atrybuty dojrzałości. Tymczasem żadne z kryteriów nie jest ani tak oczywiste, ani jednoznacznie spolaryzowane, jak mogłoby się wydawać. Spróbuję to uzasadnić.

Dobór i funkcja utworów. Argumentacja czy sprawozdanie

Uczeń może posłużyć się konkretnym dziełem w funkcji argumentacyjnej tylko wtedy, jeśli ów utwór pozostaje w harmonii z jego osobistą hierarchią wartości. W przeciwnym wypadku argumentacja na rzecz własnych opcji zostanie uznana przez nauczyciela jako błędna interpretacja lub zły przykład. Właśnie z tego względu dobór tekstów może być oceniany wyłącznie w odniesieniu do koncepcji ucznia, a nie do wizji nauczyciela. Ponadto warto zwrócić uwagę, iż w praktycznej sferze języka sprawozdanie jest tak samo istotne jak uzasadnianie. Na równi z umiejętnością argumentacji potrzebne jest zatem kształtowanie kompetencji uczniów w zakresie komunikatywnej sprawozdawczości, a tę znajdziemy nie w *Tangu* Mrożka, lecz na przykład u Jacka Żakowskiego w książce *Co dalej panie Mrozek?* Problem tkwi w tym, że uczeń, który na maturze odwołałby się do tego drugiego utworu, mógłby być odruchowo i bezzasadnie słabiej oceniony.

Sądy wartościujące. Formułowanie czy odtwarzanie

Odtwarzanie norm wartościujących ma nie mniejszą rangę niż samodzielne formułowanie sądów. Przyjmuje się bowiem definicyjnie, iż standardy ewaluatywne, a nawet wartości uniwersalne, są właśnie uogólnieniem postaw i opinii większości. Uczeń wypowiadający się tak jak ogół nie powinien z tego powodu otrzymywać gorszej oceny. Co więcej – uznawanie takich czy innych wartości nie może mieć jakiegokolwiek wpływu na ocenę z języka polskiego. W kontekście wartości mówienie o stereotypie sądów jest w ogóle chybione, gdyż nadaje pejoratywny wydźwięk społecznej zgodzie w kwestiach zasadniczych. Ponadto ze swej perspektywy nauczyciel nie jest w stanie stwierdzić, czy dany pogląd był w procesie myślowym ucznia twórczy, czy odtwórczy. Wiemy wszak z psychologii, iż rzeczywiście jednostki często same dochodzą do czegoś, co inni już wcześniej odkryli.

Odczytywanie utworów. Sens dosłowny czy przenośny

Przed kilku laty głośnym echem w mass mediach odbiły się badania, które dowiodły, że większość Polaków nie potrafi czytać tekstów ze zrozumieniem. Nic dziwnego, skoro na „polskim” sięga się tylko po dzieła, a nie do pomyślenia jest np. analiza treści instrukcji obsługi urządzeń. Polonista nie widzi potrzeby, aby wziąć podręcznik z fizyki i na przykładzie zadań z treścią ćwiczyć ich percepcję. Wszystko wskazuje na paradoks, jakoby kształtowanie kompetencji w odczytywaniu dosłownym tekstów było czymś gorszym i leżało poza zadaniami szkoły. Na lekcjach króluje metafora, a jakoś dziwnie cicho na temat równie ekspresyjnego środka wyrazu, jakim jest izofora. Odczytywanie sensu przenośnego w wydaniu szkolnym ma wytyczone ramy „jedynie słusznej interpretacji”. Tymczasem rozkoszowanie się nektarem zawartym między wierszami dzieła ma swą wartość tylko wtedy, gdy zostanie ono samodzielnie doznane i zasmakowane przez czytelnika. Owe subiektywne odczucia i gusty nie powinny być podstawą oceny maturalnej. Zamiast przeciwstawiać sobie dosłowność i przenośność, lepiej stosować kryterium rozumienia w szerokim sensie, tj. z odczytywaniem także przesłania utworu.

Język wypowiedzi. Styl potoczny czy literacki

W obowiązujących standardach edukacyjnych znajduje się jednoznaczne sformułowanie o ogólnej, a nie wyłącznie o jakiejś kulturalnej odmianie polszczyzny. Ludzie z nielicznymi wyjątkami posługują się w życiu stylem potocznym. Czyżby tej formy nie trzeba było już szlifować? Oczywiście jest, że wypowiedzi ustne i pisemne rządzą się odmiennymi prawami, które uczeń powinien adekwatnie stosować. Jeśli jednak chodzi o język literacki i potoczny, to każdy z nich pełni doniosłą funkcję cywilizacyjną i społeczną, stąd trzeba je traktować równoważnie. Nie można przeznaczać 12 lat nauczania języka polskiego na kształtowanie tylko umiejętności wypowiedzi wzniosłej, zaniedbując komunikacyjną funkcję języka. Dziś owo zanie-

dbanie szkoły słycać wszędzie. Niestety – komunikatywności wypowiedzi nie da się doskonalić i wzorować na zalecanych obecnie lekturach, a nabyte kompetencje, li tylko literackie, z całą pewnością nie są żadną miarą dojrzałości społecznej.

STRUKTURA KRYTERIÓW

Do rzetelnej walidacji kryteriów oceniania niezbędne jest przyjrzenie się strukturze proponowanego systemu, a więc komponentom i ich wzajemnym powiązaniom. Postawiony w jednej z wypowiedzi zarzut „przerostu formy nad treścią” dotyczy zapewne nadmiarowości poszczególnych warstw narzędzia: arkusza oceny, recenzji i jeszcze uzasadnienia. Może ponadto odnosić się do wyraźnie naciąganego rozpisania kryteriów na oceny środkowe, zamiast tylko na skrajne. Skoncentrujmy się obecnie na propozycjach optymalizacyjnych, wyrażanych przez respondentów, a dotyczących struktury kryteriów.

Postać arkusza ocen i recenzji

W ankiecie pytano o „Arkusze ocen i recenzji” w całości, gdy tymczasem respondenci odmiennie opiniują każdy z elementów konstrukcji. Głównym wątkiem krytyki był zarzut zmuszania nauczycieli do trzykrotnego powtarzania podobnych sformułowań wobec każdej z poprawianych prac. Starano się wykazać zbędność recenzji i uzasadniania stopni. Gdyby jednak formy opisowe miały pozostać, to zdaniem wielu nie powinny one powielać haseł z karty ocen; wystarczyłoby puste miejsce na uwagi o wadach i zaletach pracy. Większość nauczycieli opowiedziała się za próbą ilościowego ujęcia w samym tylko arkuszu ocen. Znalazły się jednak również takie osoby, które zamiast wypełniać arkusz krzyżykami, wolą pisać werbalną recenzję opartą o kryteria. Sprawa nie jest więc rozstrzygnięta, choć z całą pewnością forma narzędzia wymaga zdecydowanej konsolidacji.

Szczegółowość a precyzja

Respondenci wyrażali troskę o to, ażeby drobiazgowa analiza według kryteriów nie brała prymatu nad oceną merytorycznego poziomu wypowiedzi. Zapewne chodziło im o relacje wielości aspektów formalno-językowych wobec ewentualnej wzniosłości, błyskotliwości lub oryginalności wypracowań. Kryteria wciąż jeszcze są niekompletne, a już teraz określono je jako nadmiernie szczegółowe i wymagające uproszczenia. Z drugiej strony zwrócono uwagę na enigmatyczne sformułowania: „w zasadzie”, „na ogół”, „właściwie”, „dopuszczalne”, podkreślając zupełny brak opisania tego, co jednoznacznie dyskwalifikuje pracę. Można streścić owe oczekiwania wobec kryteriów następująco: nauczyciele chcą narzędzia bardziej syntetycznego i precyzyjniejszego.



Operatywność narzędzia

Aż w kilkudziesięciu ankietach zaakcentowano ogrom czasu niezbędnego do rzetelnej oceny wypracowań według proponowanych kryteriów. Na całość procedur uruchamianych przy poprawianiu prac oraz wypełnianiu arkusza ocen i recenzji użyto wielokrotnie bardzo dosadnego określenia „biurokracja”. Nauczyciele twierdzą, że wnikliwa ocena wypracowań choćby tylko z jednej klasy, to olbrzymi nakład pracy. Tymczasem są przymiarki do wyprowadzenia tych czynności na zewnątrz. Wbrew potocznym mniemaniom uważam, że poprawianie prac poza macierzystą szkołą uczniów pogorszy rzetelność ocen. Ze względu na czasochłonność byłoby to bowiem mniej staranne, a bardziej rutynowe, w całym pejoratywnym kontekście tego słowa.

Hierarchia kryteriów

Kontrowersje w wypowiedziach dotyczą kwestii, czy wszystkie kryteria są jednakowo ważne i czy konieczne jest ocenianie bez wyjątku wszystkich cech. Grupa nauczycieli po doświadczeniach z zajęć metodycznych zaproponowała elastyczność w podejściu do wypełniania rubryk, jeżeli forma analizowanego utworu nie zawiera którejś z cech wyszczególnionych w kryteriach. Część respondentów zastanawiała się, czy kolejność na liście kryteriów ma jakiś związek z hierarchią ważkości. Wielu domagało się określenia relacji między poszczególnymi pozycjami kryteriów, z ustaleniem rankingu aspektów merytorycznych wypracowania względem poprawności językowej. Proponowano wprowadzenie punktacji o odmiennych wagach, nadanych różnym kryteriom. Wszystkie te koncepcje zmierzają do rozwikłania zagadki, jak z wielu wątków oceniania pracy przejść do pojedynczego stopnia szkolnego. Kryteria powinny ten problem rozstrzygać jednoznacznie.

Granice i progi wymagań

Respondenci krytykują brak ujęcia w kryteriach oceny niedostatecznej i celującej. Domagają się zwłaszcza uściślenia wymagań koniecznych. Zdarza się bowiem tak, że jedno z kryteriów jest na poziomie niedostatecznym, a pozostałe na dostatecznym. Problemem jest wówczas to, czy liczyć średnią, czy zdyskwalifikować pracę. Każde uśrednianie stopni częściowych prowadzi do tendencji skupiającej, w której mniej jest ocen skrajnych, najlepszych i najgorszych. Kilkoro nauczycieli zmartwiło się tym, że przy proponowanych kryteriach trudniej będzie postawić je-dynkę. Znaleźli się i tacy, którzy twierdzą, że za mierną według kryteriów pracę należy się niedostateczny. Cóż, mają pełne prawo do wyrażania w ankiecie dowolnych opinii. Jeśli już jednak nowe kryteria obejmą maturę regulaminową, to wszystkim wypadnie trzymać się przyjętych, jednolitych zasad oceniania.

MERITUM KRYTERIÓW

Nie bez powodu respondenci tak obszernie i dość krytycznie wypowiedzieli się o strukturze kryteriów. Dopiero bowiem w przypadku prawidłowej formy narzędzia, można będzie owe ramy wypełnić doskonalszą, bardziej precyzyjną treścią. Przyjrzyjmy się teraz równie konstruktywnym opiniom nauczycieli w kwestii warstwy merytorycznej kryteriów.

Zgodność prac z tematem

Zdaniem respondentów autorzy kryteriów nie uwzględnili sytuacji, kiedy to dobór tekstów jest wprawdzie zgodny z koncepcją ucznia, lecz koncepcja ta jest nie-spójna z tematem. Brak ujęcia w kryteriach dość powszechnego błędu pisania wypracowań nie na temat, był najczęściej stawianym zarzutem. W opinii nauczycieli ścisłość realizacji zadania, zawartego w temacie pracy, powinna być zupełnie odrębnym i wręcz podstawowym kryterium.

Ujęcie problemu

Okazało się, że ankietowani nie podzielili moich wątpliwości, czy argumentacja i sprawozdanie mogą być umieszczane na przeciwległych krańcach skali wartości. Pojawily się za to kwestie czysto techniczne. Respondenci wskazali na błąd w sformułowaniu „problematyczne ujęcie” zamiast problemowe. Sugerowali, aby oceniać funkcjonalność ewentualnych cytatów, a nie oczekiwać ich w każdym wypracowaniu. Ponadto prosili o redagowanie tematów raczej w formie pytań oraz o sięganie po realne problemy bez zmuszania uczniów do „rozważań abstrakcyjnych i filozofujących”.

Konteksty kulturowe

Twórcy kryteriów jednoznacznie usytuowali brak kontekstu kulturowego w ocenie miernej. Wzbudziło to sprzeciw tej części nauczycieli, którzy za brak kontekstu bądź już za sam nietrafny dobór dzieł automatycznie chcą wystawiać ocenę niedostateczną. W kilku ankietach domagano się, ażeby matura sprawdzała przede wszystkim znajomość lektur. Wypada więc przypomnieć, że zawężenie kontekstu kulturowego tylko do literatury wyzwała największy i zupełnie uzasadniony opór uczniów.

Wskazywanie wartości

W wypowiedziach pojawiły się wątpliwości, czy słuszne jest rozbijanie na dwa odrębne punkty kryteriów związanych z wartościowaniem. Respondentom chodziło zapewne o zrównoważenie ocen za dostrzeganie i wskazywanie wartości z rangą

formułowania własnych sądów. Do tej kwestii odniosłem się już wcześniej, dodam więc jedynie, że w jednej z ankiet znalazłem zbliżony do mojego pogląd o potrzebie doceniania choćby samego postrzegania wartości.

Odczytywanie dzieł

Respondenci sygnalizowali, że nie wszystkie utwory i formy literackie kwalifikują się do odczytywania warstwy metaforycznej. Jako przykład wymieniano esej. Padły propozycje, ażeby zrównoważyć wartość percepcji sensu dosłownego i przenośnego. Innym sugerowanym rozwiązaniem byłoby pomijanie kryterium nie mającego zastosowania wobec danego dzieła. Z drugiej jednak strony, w kilku ankietach opowiedziano się za wystawianiem oceny niedostatecznej w przypadkach wyłącznie dosłownego odczytywania dzieł, mimo że kryteria dopuszczają stopień dostateczny.

Kształtowanie wypowiedzi

Uwagi nauczycieli świadczą o niejednoznacznym odczytywaniu kryteriów konstrukcyjnych, cytując: „przejrzystość kompozycyjna wiąże się z poprawnym graficznie opracowaniem tekstu, dlatego należałoby te kryteria połączyć” (?). W celu uniknięcia nieporozumień konieczne jest wyraźniejsze zaakcentowanie, że kompozycja odnosi się do struktury narracyjnej i warstwy semantycznej wypowiedzi, a nie do układu typograficznego tekstu. Należy oceniać trafność śródtytułów oraz prawidłowość podziału wątków na akapity, lecz na stopień nie powinno wpływać to, czy akapity wyróżniono wcięciem, czy może pustym wierszem albo, czy marginesy i światła są proporcjonalne do wielkości liter.

Błędy językowe

Nauczyciele mają wątpliwości co robić, jeśli w pracy jest dużo błędów interpunkcyjnych, podczas gdy leksyka, składnia, fleksja i frazeologia są prawidłowe. Proponują, ażeby kryterium interpunkcji wyodrębnić. Domagają się uściślenia wagi błędów ortograficznych w odniesieniu do uchybień stylistycznych i merytorycznych. Chcą też ujęcia ilościowego: za ile i za jakie błędy ma być wystawiana ocena niedostateczna, bez względu na poziom wypowiedzi.

Obiektywność ocen

Trzy osoby orzekły, że oceny za wypracowania zawsze pozostaną subiektywne. Jeśli jest tak rzeczywiście, to może – zamiast organizować maturę z języka polskiego – lepiej rzucać kostką i wpisywać na świadectwie liczbę uzyskanych oczek? Żartując z metody statystycznie sprawiedliwszej przyjmijmy, że owej trójce chodziło raczej o zaakcentowanie trudności w osiągnięciu pełnego obiektywizmu. Zdecydowana większość respondentów dostrzega pozytywną rolę kryteriów, argumentując to tym, że:

zmuszają one nauczyciela do wnikliwości przy poprawianiu prac, rozstrzygają w sytuacjach wątpliwych, służą uwiarygodnieniu i obiektywizacji oceny, a uczniom pomagają poznać wymagania. Są jednak również nauczyciele uważający, że proponowane, zbyt łagodne kryteria prowadzą do sztucznego zawyżania wyników i „poszerzania kręgu pseudointeligentów”. Nie mogą oprzeć się pokusie przewrotnego odczytania tej ostatniej frazy jako zamknięcia się właśnie w... owym kręgu.

Matura sprawdzająca czy różnicująca

W ankietach przewinęła się obawa o zagubienie indywidualności ucznia wskutek zastosowania szablonowych kryteriów. We wskazówkach metodycy wręcz zachęcają do rangowania wypracowań przed wystawieniem oceny. Wszystko to świadczy o braku wyraźnych rozstrzygnięć w sprawie funkcji matury: czy ma sprawdzać wiedzę, czy ma różnicować uczniów. Trudno jest te opcje połączyć, a każda z osobna jest w pewnym sensie niesprawiedliwa: Czyż wolno obniżyć ocenę tylko dlatego, że uczeń chodzi do elitarniej szkoły, w której wypadł słabiej na tle innych wybitnie uzdolnionych? Czyż można wszystkim maturzystom dać ten sam zestaw tematów, skoro mieli różny program nauczania? Ja opowiadam się za jednolitą dla wszystkich (także dla pracujących) maturą sprawdzającą, opartą na realnych, łagodniejszych kryteriach.

Nowe kryteria interpretacji

Spora grupa nauczycieli nie mogła wypowiedzieć się w kwestii kryteriów interpretacji wierszy, gdyż żaden z ich uczniów nie pisał tematu czwartego. Niektórzy ocenili je pozytywnie, mimo że krytykowali kryteria z arkusza ocen. Świadczy to jednak bardziej o pewnym nowatorstwie ujęcia niż o jakości. Tradycyjne kryteria są znane i stosowane, stąd mogą wydawać się zbędne, gdyż naruszają dotychczasową rutynę oceniania. Z jednej strony uznano za wartościową ogólną antynomię: wady i zalety wypracowania, a z drugiej – spora grupa respondentów krytykowała kryteria interpretacyjne, określając je jako ogólnikowe, niejednoznaczne i zbyt literackie. Niejasna jest wypowiedź o zbyt małym wyeksponowaniu poetyki i o braku kryterium łączącego analizę porównawczą z językiem teorii literatury. Czyżby należało to rozumieć jako chęć oceniania licealistów za umiejętność posługiwania się hermetycznym językiem literaturoznawców, krytyków i recenzentów?

Sprawą smutną jest ingerencja nauczycielska, cytuję: „Dobrze byłoby, gdyby nauczyciel mógł krótko zinterpretować tematy”. W innej, dosadniejszej wypowiedzi nauczyciel uznał, iż ma pełne prawo do interpretacji, gdyż czuje się podmiotem matury. Zapomniał, że w ten sposób ucznia pozbawia podmiotowości. Wbrew zakazowi tematy są nadal komentowane. Potwierdzili to uczniowie, tłumacząc w ankiecie, że nie wybrali tematu czwartego, ponieważ nauczyciele określali to zadanie jako najtrudniejsze (!).

WALIDACJA KRYTERIÓW

Podczas formułowania wniosków co do jakości „Arkusza ocen i recenzji” trzeba uwzględnić fakt, że większość nauczycieli po prostu zaakceptowała kryteria w ich aktualnej postaci. Świadczy to o właściwym kierunku prac Zespołu zajmującego się obiektywizacją oceniania. Tym niemniej wiele z nauczycielskich uwag, dotyczących formy i treści kryteriów, trzeba uwzględnić przy doskonaleniu narzędzia. Podczas wprowadzania zmian konieczne jest jednak podejście metodologiczne, a nie intuicyjne. W przeciwnym wypadku nasila się, sporadycznie na razie, opinie, że lepsze były kryteria ubiegłoroczne, a nawet te z 1983 r. W przyszłym roku nieodzowne jest przeprowadzenie pełnej walidacji kryteriów oceniania, aby za dwa lata można było maturę pisemną z języka polskiego poddać ewaluacji.

Wyjaśnię syntetycznie różnice między dwoma mało znanymi pojęciami. Otóż *walidacja* oznacza wartościowanie i zatwierdzanie eksperckie, natomiast *ewaluacja* jest wartościowaniem społecznym. Domeną pierwszej są m.in. uniwersum kulturowe i standardy edukacyjne, a drugiej – wartości uniwersalne i standardy ewaluacyjne. Do zoptymalizowania matury najpierw potrzebne jest zbadanie i stwierdzenie jej wartości w aspekcie cywilizacyjnym, a potem uzyskanie społecznej akceptacji. Podczas walidacji nauczyciele są o tyle ekspertami, o ile potrafią wystąpić w roli obiektywnych sędziów kompetentnych, tłumiących osobiste nastawienia. W ewaluacji natomiast nauczyciele i uczniowie jako podmioty matury ujawniają właśnie postawy i swój punkt widzenia, nie tylko oceniając, lecz także wyrażając doznawane przez siebie emocje i motywacje.

Poloniści zapewne nie zdają sobie sprawy, iż to właśnie oni są najbardziej kompetentni do przygotowania kwestionariusza walidacji kryteriów. Cała bowiem trudność polega na ułożeniu precyzyjnych zdań, ujmujących sedno każdego problemu w parę komplementarnych stwierdzeń. Przedstawiam przykład prawidłowej konstrukcji profesjonalnej skali pomiarowej. Treść owych stwierdzeń wynika z syntezy otwartych wypowiedzi nauczycieli.

Matura winna różnicować, stąd ocena musi uwzględniać porównanie z poziomem innych prac.
Matura winna sprawdzać, stąd ocena pracy musi wynikać wyłącznie ze spełnienia kryteriów.
Kryteria powinny być szczegółowsze, ażeby ujmowały z osobna każdą z pożądanych cech.
Kryteria powinny być bardziej uogólnione, z podobnie dużym zasięgiem cech jak standardy.
Kryterium wyjątkowo nie pasujące do formy omawianego utworu może zostać pominięte.
W ustalaniu oceny musi być uwzględnione odniesienie do wszystkich bez wyjątku kryteriów.
Trzeba wskazać te kryteria, których nie spełnienie z góry wyklucza możliwość zaliczenia.
Trzeba wskazać kryteria, które mogą być niedostateczne, a mimo to pracę wolno zaliczyć.
Warto kryteriom przypisać wagi punktowe, aby ustalić ważkość danej cechy wypracowania.
Zamiast zróżnicowanej punktacji, lepiej wszystkie kryteria oceniać w skali stopni szkolnych.
Kryteria winny wyznaczać tylko dwa skrajne progi, tj. brak zaliczenia oraz ocenę celującą.
Warto rozpisać kryteria na każdy ze stopni szkolnych, od niedostatecznego do celującego.
Zamiast przeciwstawiać sobie odmienne cechy, lepiej oceniać natężenie jakości każdej z nich.
Dosłowne i metaforyczne odczytywanie to prawidłowy układ skrajnych jakości na skali ocen.

Podane przykłady dotyczą głównie struktury kryteriów, trzeba by więc je poszerzyć o wszystkie wątki merytoryczne, co łącznie dałoby ok. 40 pozycji. Ta dość pokaźna liczba wynika z przyjętej metody dublowania, w której sąsiednie stwierdzenia tworzą specyficzną parę kontrolną, niezbędną do głębszej interpretacji wypowiedzi. Dotykają tej samej kwestii z odmienną perspektywą. Z tego powodu nie wolno odczytywać powyższych zdań jako zbioru słusznych tez. O prawdziwości/falszywości stwierdzeń rozstrzygną respondenci, wybierając jedną z czterech opcji: *nie*, *chyba nie*, *chyba tak*, *tak*. Zbędna jest neutralna opcja *brak zdania* lub *nie wiem*, skoro wypowiadać się mają sędziowie kompetentni. Zachęcam czytelników do zastanowienia się, jak w tej skali ustosunkowaliby się do poruszonych kwestii.

* * *

Na obecnym etapie tworzenia kryteriów oceniania w pewnych obszarach konieczne jest podejście elastyczne, w innych bardziej szczegółowe. Swoista nadmiarowość musi współbrzmieć z klarownością i słusznością zasad opartych na kulturowym uniwersum. Podstawę tworzyć powinny harmonijnie dobrane tylko te standardy edukacyjne, które na danym etapie rozwoju społeczeństwa uznawane są przez ogół za wartościowe i konieczne. Podkreślam: przez ogół, a więc także i przez uczniów. Jest jeszcze czas na dyskusję i optymalizację. Po starannym dopracowaniu i uzyskaniu aprobaty, kryteria maturalne, zespolone z nowym programem edukacji, kształtować będą standard reguł, którym wypadnie się podporządkować. Martwi mnie jednak to, że część polonistów wybiera trajektorię Ikara, wskutek czego wkrótce dojść może do wyrugowania języka polskiego z egzaminu dojrzałości! Stanie się tak, gdy polskie suwerenne uczelnie zamiast rodzimej matury zaczną uznawać bardziej pragmatyczną maturę międzynarodową.

Gniezno, 12 X 1997 r.

EWALUACJA PRÓBNEJ MATURY Z MATEMATYKI

Od trzech lat próbnej maturze z matematyki towarzyszy *różnicowy pomiar wartościujący*, oparty na poznańskim modelu ewaluacji demokratycznej. W metodzie tej propozycje kreatorów Nowej Matury poddawane są pod ocenę **podmiotom egzaminu**, tj. zdającym uczniom i ich nauczycielom. Reakcje respondentów na bodźce – stwierdzenia ankiety – są wskaźnikami postaw wobec konkretnych cech matury. Uogólnione postawy danej zbiorowości tworzą względnie trwałe obraz standardów ewaluacyjnych. Dostrzegalne na tym tle zmiany opinii świadczą o istnieniu i oddziaływaniu czynników modyfikujących, a więc pośrednio o jakości działań reformatorskich, a także o trendach społecznych przewartościowań, zachodzących zwrótnie jako efekt upowszechniania dorobku Nowej Matury.

Głębia pomiaru różnicowego ujawnia się w badaniu długofalowym, dlatego bieżące wyniki zestawiono z uzyskanymi poprzednio. W analizie tabel szczególną uwagę warto zwrócić na *tendencje do zmian w rozbieżności zdań* wobec poszczególnych kwestii oraz na *kierunki i nasilenie trendów wypowiedzi* na wszystkich trzech poziomach HML. Pożądanym zjawiskiem niekoniecznie musi być wzrost średniej wartości danej cechy, wszak akceptacja konkretnego zagadnienia mogła już osiągnąć poziom normy standardu ewaluatywnego. Owe normy dla każdej z cech różnią się, gdyż dużo łatwiej jest na przykład zaakceptować zawarte w ankiecie stwierdzenia *o jakości zadań* niż *o słuszności stopni*. Z tego względu nie można upraszczać interpretacji do stwierdzeń, iż tę pierwszą cechę oceniono „na piątkę z plusem”, a drugą tylko „na trójkę”. W metodzie ważonych ocen istotna jest analiza przesuwania się wyników w stronę zgodności wypowiedzi, o czym świadczy miara rozziewu [-b] opinii respondentów.

OBJAŚNIENIE SPOSOBU INTERPRETACJI PARAMETRÓW STATYSTYCZNYCH

Użyta do badań skala postaw ma specyficzną biegunowość, dzięki czemu po nadaniu wskaźnikom wag z zakresu $< -3; +3 >$ można wyznaczać dwa wymiary, tj.: jakość oraz intensywność wypowiedzi. Jakość jako parametr jest średnią arytmetyczną z wag, natomiast intensywność to średnia z modułów wag. Pierwszy parametr odzwierciedla *wartościowanie przez respondentów danej cechy*, drugi zaś oznacza *stopień zdecydowania respondentów co do słuszności swych racji*. Z różnicy między jakością a intensywnością wypowiedzi obliczany jest rozziew [-b]. Parametr ten jest miarą wzajemnej zgodności respondentów. Sytuacja idealna ma miejsce wtedy, gdy rozziew jest zerowy, czyli żadna z osób nie neguje pozytywnej opinii, wyrażanej przez ogół badanych. Sprzeczność opinii jest cechą niekorzystną, dlatego parametr ten przyjmuje wyłącznie wartości niedodatnie. W praktyce pomiarowej niewielki rozziew w zakresie $< 0; -0,5 >$ interpretuje się jako jednomysłność respondentów, natomiast znaczny rozziew (od -2) wskazuje na społeczne zakwestionowanie jakości danej cechy matury.

Aby ułatwić nauczycielom intuicyjne odczytywanie wyników, przeniesiono wskaźniki pozycyjne na 6-stopniową skalę stopni szkolnych. W ten sposób jakość przeszła na poziom średni wartości [M]. Trzeba jednak brać pod uwagę fakt, że przyjęto łagodne kryterium konwersji z progiem o poziomie 3, stąd wyniki średnie poniżej tego progu oznaczają przewagę wypowiedzi krytycznych. O jednoznacznej aprobacie danej cechy matury można mówić dopiero od wartości 4 wzwyż, natomiast średnie poniżej 2 są wyrazem dezaprobaty. Z kolei te przypadki, w których poziom średni bliski jest wartości 3 oznaczają, że wobec danej kwestii nie ukształtował się jeszcze w społeczeństwie standard ewaluacyjny i, z powodu równowagi przeciwstawnych opinii, nie ma podstaw do orzekania o bezwzględnej jakości owej cechy. Właśnie w tych przypadkach ważne jest obserwowanie wieloletnich trendów

i różnic między typami szkół. Z wyników wieloletnich wyłania się średni poziom oczekiwany, zwany normą.

Obok parametrów pozycyjnych do analizy niezbędne są też wskaźniki rozproszenia wypowiedzi. Zamiast klasycznych odchyłeń standardowych wykorzystano identycznie interpretowane, lecz bliższe rozkładowi rzeczywistym, odchylenia realne. Liczy się je z wariancji cząstkowych, analizując osobno ujemne i dodatnie odchylenia od średniej, co jest konieczne przy typowej dla skali postaw lewostronnej asymetrii rozkładów. Po dodaniu odchyłeń realnych do wartości średniej powstają odpowiednio: poziom niski [L] oraz poziom wysoki [H]. Wyznaczają one punkty przebieg wykresu estymowanej funkcji gęstości. W dokładnej analizie zmian mierzonej cechy warto zawsze brać pod uwagę łączną oraz wzajemną fluktuację całej trójki poziomów HML w kolejnych latach.

W tabelach z wynikami przyjęto następującą konwencję: literowe kody parametrów HML oraz - b ze znakiem prim [''] oznaczają wyniki z roku 1996, znak bis ["] wskazuje na rok 1997, natomiast symbole bez indeksów dotyczą próbnej matury z roku 1998. Do opracowania poniższych statystyk wykorzystano łącznie ok. 500 ankiet nauczycielskich oraz ponad 2000 uczniowskich. Niepokoi jednak fakt, że na ostatniej próbnej maturze przeprowadzono egzamin tylko według formuły problemowej, toteż prezentowane tutaj statystyki dotyczą jedynie tradycyjnej formy matury.

ANALIZA WYPOWIEDZI NAUCZYCIELI O CECHACH JAKOŚCI MERYTORYCZNEJ I ORGANIZACYJNEJ

Tabela 1

Wartość próbnej matury problemowej w opinii nauczycieli z obu typów szkół

Zmienna Parametr	Kod	Zgodność zadań z programem		Jakość doboru zestawu		Przekrojowość zagadnień		Jasność sformułowań		Trafność trudności zadań		Zbieżność ocen ze szkolnymi		Spójność ocen i oczekiwań		Osąd formuły problem.		Akceptacja dwu etapów	
		Ogół	Zaw	Ogół	Zaw	Ogół	Zaw	Ogół	Zaw	Ogół	Zaw	Ogół	Zaw	Ogół	Zaw	Ogół	Zaw	Ogół	Zaw
1996	H'	6,0	6,0	6,0	6,0	4,8	4,9	6,0	6,0	6,0	5,1	4,4	4,4	4,9	4,9	5,0	5,1	5,4	5,4
Wartość	M'	5,7	5,1	5,1	4,7	3,7	3,4	5,6	5,4	4,5	3,8	3,1	3,1	3,2	2,8	2,5	2,4	3,7	3,6
	L'	3,8	3,3	3,1	2,8	1,8	1,3	2,0	3,3	3,2	1,3	0,8	1,2	0,9	0,7	1,0	0,9	0,7	0,9
Rozziew	-b'	0,0	-0,2	-0,2	-0,4	-0,6	-1,1	-0,2	-0,1	-0,3	-0,9	-1,3	-1,3	-1,6	-2,0	-2,2	-2,4	-1,5	-1,5
1997	H''	6,0	6,0	6,0	6,0	4,7	4,5	5,0	6,0	4,9	4,8	4,2	4,3	4,6	4,6	4,7	4,8	6,0	5,4
Wartość	M''	5,5	5,3	4,1	4,1	2,9	3,2	3,9	4,2	3,9	3,4	3,2	3,2	3,0	2,9	2,4	2,7	4,9	4,0
	L''	3,9	3,3	2,5	2,6	1,3	1,5	1,1	2,8	1,6	1,3	1,2	1,6	0,9	1,1	0,9	1,0	3,1	1,0
Rozziew	-b''	0,0	-0,1	-0,7	-0,6	-1,6	-1,2	-0,7	-0,5	-0,6	-1,1	-1,0	-1,0	-1,6	-1,6	-2,2	-2,0	-0,3	-1,0
1998	H	6,0	6,0	6,0	6,0	4,7	4,2	6,0	6,0	5,2	5,0	4,6	4,3	4,7	4,6	4,8	4,0	-	-
Wartość	M	5,5	5,5	4,5	5,0	3,3	3,4	5,5	5,4	4,0	3,8	2,9	2,5	3,1	2,7	2,4	2,4	-	-
	L	3,6	4,0	2,9	3,4	1,5	1,7	3,8	4,0	2,6	1,7	1,1	1,2	1,1	1,2	0,9	1,2	-	-
Rozziew	-b	-0,1	0,0	-0,4	-0,2	-1,1	-0,7	0,0	0,0	-0,6	-0,7	-1,6	-1,9	-1,5	-1,7	-2,3	-1,9	-	-

Przyjrzyjmy się zestawionym w tabeli 1 parametrom statystycznym zmiennych cząstkowych, wyznaczającym jakość wybranych cech matury w oparciu o poziomy i trendy opinii nauczycieli:

Zgodność zadań z programem – w tej kwestii charakterystycznie wysoka wartość średnia oraz znikomy współczynnik rozrzewu wskazują na jednomyślnie wysoką aprobatę, przy czym wzrostowa tendencja w szkołach o profilu zawodowym osiągnęła już poziom oczekiwanej normy dla tegoż wskaźnika.

Jakość doboru zestawu – po ubiegłorocznym dużym spadku wartości nastąpiła poprawa, choć w szkołach ogólnokształcących wyraźnie niższa. Uzyskany tam przed dwoma laty wynik świadczy o tym, że realne jest ustalenie tej cechy na poziomie 5, chociaż ułożenie doskonałych zadań wcale nie jest łatwe.

Przekrojowość zagadnień – ta cecha w formule problemowej nie ma szans na wysoką ocenę, lecz mimo to, po lepszym wyniku w liceach z 1996 roku, nie ma usprawiedliwienia dla tak niskiego poziomu wartości średnich. Normą mógłby być tu poziom bliższy 4, z mniejszym rozrzewem.

Jasność sformułowań – po ubiegłorocznej zapaści tegoż wskaźnika, tym razem uzyskane wyniki są wzorcowe, wyznaczając niezbędną normę na poziomie 5,5 z zerowym rozrzewem. Warto tu odnotować wysokie wartości na dolnym poziomie L, co oznacza aprobatę nawet osób krytycznych.

Trafność trudności zadań – to wskaźnik, który różnicuje oba typy szkół. Możliwa do osiągnięcia norma w szkołach zawodowych jest niższa niż w liceach ogólnokształcących. Podobnie jak z innymi cechami, także i tu, po spadku w roku ubiegłym, wyniki w liceach ogólnokształcących nie powróciły na poziom sprzed dwóch lat.

Zbieżność ocen ze szkolnymi – w tej kwestii w bieżącym roku nastąpiło zdecydowane załamanie, zwłaszcza w szkołach zawodowych. Nie można tego tłumaczyć jedynie zawyżaniem ocen szkolnych, lecz trzeba uznać na podstawie wskaźników, że wymagania maturalne są po prostu nadmierne.

Spójność ocen i oczekiwań – wprawdzie tylko niewielki spadek w szkołach zawodowych, lecz trzeba podkreślić, iż są to opinie nauczycieli. Jeśli oni, doskonale znając możliwości swych uczniów, liczyli na lepsze wyniki z egzaminu, to brak adekwatności ocen jest z pewnością słabą stroną matury.

Osąd formuły problemowej – nauczyciele nadal surowo oceniają egzamin w tradycyjnej postaci. To właśnie wskutek tak ostrej krytyki konieczne są bardziej radykalne zmiany i prace nad formułą testową.

Akceptacja dwu etapów – w poprzednim roku nastąpiła wyraźna poprawa, lecz w roku 1998 nie włączono tej kwestii do ankiety, stąd brak danych o aktualnym porównaniu matury dwuetapowej przez nauczycieli.

ANALIZA WYPOWIEDZI UCZNIÓW O CECHACH JAKOŚCI MERYTORYCZNEJ
I ORGANIZACYJNEJ

Tabela 2

Wartość próbnej matury problemowej w opinii uczniów z obu typów szkół

Zmienna Parametr	Kod	Zgodność zadań i nastawień		Atrak- cyjność zestawu		Jasność sformu- łowań		Percepcja treści zadań		Trafność trudności zadań		Przystęp- ność formuły		Spójność ocen i oczekiwań		Samo- dzielność uczniów		Osąd formuły problem.		Akcepta- cja dwu etapów	
		Ogół	Zaw	Ogół	Zaw	Ogół	Zaw	Ogół	Zaw	Ogół	Zaw	Ogół	Zaw	Ogół	Zaw	Ogół	Zaw	Ogół	Zaw	Ogół	Zaw
1996	H'	5,3	5,0	5,1	4,4	6,0	6,0	6,0	6,0	5,2	4,7	4,8	4,8	4,7	4,8	5,6	5,4	5,5	5,4	5,2	4,2
Wartość	M'	2,7	2,8	3,7	2,9	5,2	4,5	5,3	4,6	3,4	2,3	3,6	3,4	2,7	2,7	3,9	3,6	3,4	2,9	2,3	1,7
	L'	0,9	0,9	1,2	1,1	3,6	3,2	3,9	3,3	0,8	0,6	1,1	1,3	0,8	0,9	0,8	0,8	0,7	0,6	0,4	0,0
Rozziew	-b'	-2,2	-2,1	-0,9	-1,6	-0,1	-0,3	0,0	-0,2	-1,5	-2,6	-0,9	-1,1	-2,0	-2,1	-1,3	-1,5	-1,8	-2,3	-2,9	-3,7
1997	H''	4,5	4,4	4,6	4,5	5,0	6,0	6,0	6,0	4,6	4,5	4,7	4,7	4,5	4,8	5,4	5,3	4,9	4,1	5,5	5,3
Wartość	M''	2,9	2,9	2,6	2,8	3,9	4,0	4,6	4,4	2,1	2,4	3,0	3,2	2,3	2,6	3,8	3,6	2,0	1,9	3,8	3,2
	L''	1,2	1,2	1,3	1,3	1,2	2,6	3,4	3,1	0,7	1,0	1,4	1,3	0,6	0,7	0,9	0,9	0,5	0,0	0,6	0,7
Rozziew	-b''	-1,5	-1,5	-1,8	-1,6	-0,7	-0,6	-0,2	-0,3	-2,7	-2,1	-1,4	-1,3	-2,5	-2,2	-1,2	-1,4	-3,0	-3,2	-1,5	-1,9
1998	H	4,8	4,4	4,6	4,5	6,0	6,0	6,0	6,0	4,8	4,4	4,6	4,5	4,6	3,6	5,1	5,2	3,9	3,6	5,5	5,5
Wartość	M	3,4	2,6	2,9	2,2	4,7	4,4	4,9	4,5	2,8	2,0	3,2	2,6	2,3	1,5	3,6	3,2	1,9	1,2	3,8	3,8
	L	1,3	1,1	1,3	0,9	3,4	3,1	3,6	3,2	0,9	0,6	1,5	1,0	0,5	0,0	1,1	0,8	0,0	0,0	0,8	0,7
Rozziew	-b	-1,1	-1,9	-1,5	-2,4	-0,2	-0,4	-0,1	-0,3	-1,8	-2,8	-1,2	-2,0	-2,6	-3,6	-1,2	-1,8	-3,1	-4,1	-1,3	-1,4

Przyjrzyjmy się zebranych w tabeli 2 parametrom zmiennych cząstkowych, wyznaczającym jakość wybranych cech matury w oparciu o poziomy i trendy opinii uczniów zdających egzamin:

Zgodność zadań i nastawień – ciekawe jest to, że w szkołach ogólnokształcących uczniowie w zasadzie spodziewali się tego typu zadań na maturze. Wszystkie jednak pozostałe wskaźniki świadczą o tym, że pod pozorem typowości kryły się zagadnienia o trudności wyższej od zadań liczonych w szkole. Na dwa miesiące przed maturą regulaminową nie powinno być już zaskoczenia rodzajem zadań matury próbnej.

Atrakcyjność zestawu – ów wskaźnik z punktu widzenia ucznia dotyczy zapewne w mniejszym stopniu tego, aby rozwiązywanie zadań sprawiało przyjemność. W większym natomiast dotyczy takiej cechy, kiedy to uczeń ma do wyboru zestaw, z którego może wybrać zadania, jakie umie rozwiązać. W tej sferze jest niestety znacznie gorzej niż przed dwoma laty.

Jasność sformułowań – nastąpiła wyraźna poprawa w porównaniu z ubiegłym rokiem, lecz w szkołach ogólnokształcących średnia jest gorsza niż w roku 1996. Warto zwrócić uwagę na zgodną opinię nauczycieli i uczniów co do jakości tej cechy w konkretnym roku, a odmienną dla różniących się zestawów w obu typach szkół.

Percepcja treści zadań – także i z tą cechą jest lepiej niż przed rokiem, chociaż rozrzew $-0,3$ w szkołach o profilu zawodowym wskazuje, iż znaleźli się uczniowie, którzy czegoś w zadaniach nie zrozumieli. Żelazną regułą powinien tu być rozrzew zerowy. Znamienne, iż wskaźnik ten zawsze wypada na wyższym poziomie niż ocena jasności sformułowań, co może wynikać stąd, że wybierane są te zadania, których treść uczniowie lepiej rozumieją bądź wbrew zakazowi nauczyciele „omawiają” zestawy.

Trafność trudności zadań – w analizie kolejnych cech warto pomijać wynik z poprzedniego roku, kiedy to przeprowadzono wyraźnie nieudany eksperyment wprowadzenia zadania „z treścią”. Z odniesienia bieżących wyników do tych sprzed dwu lat dostrzec można silną korelację opinii uczniów i nauczycieli z tych samych szkół i z tych samych lat w sprawie poziomu trudności zadań. Nadal próg trudności jest zbyt wysoki, a już szczególnie w szkołach o profilu zawodowym.

Przystępność formuły – wartości tegoż wskaźnika są potwierdzeniem, iż obecny poziom wymagań maturalnych nie jest dopasowany do możliwości uczniów. Szczególnie nadmierna jest tegoroczna różnica w ocenie przystępności formuły problemowej między opiniami pochodzącymi ze szkół ogólnokształcących a tymi ze szkół zawodowych.

Spójność ocen i oczekiwań – już i tak złe w poprzednich latach wskaźniki tej cechy pogorszyły się jeszcze w szkołach zawodowych do poziomu zupełnie niedopuszczalnego. O tym, że oceny coraz bardziej mijają się z oczekiwaniami, twierdzą zgodnie i uczniowie, i nauczyciele. Eksperyment próbnej matury nie powinien mieć na celu stymulowania uczniów do nauki poprzez stawianie im słabych ocen.

Samodzielność uczniów – przez dwa poprzednie lata zmiany tego wskaźnika były statystycznie nieistotne. W roku 1998 nastąpiło pogorszenie, wbrew podstawowemu założeniu Nowej Matury, dotyczącemu obiektywności egzaminu. Na nic się zda kodowanie i poprawianie prac poza szkołą, jeśli jedynym sposobem zdania matematyki będzie uzyskanie podpowiedzi od innego ucznia lub nauczyciela.

Osąd formuły problemowej – w takiej postaci, jakiej doświadczają uczniowie, formuła problemowa jest z roku na rok coraz bardziej odrzucana. Czyżby chodziło o to, aby zniechęcając do niej łatwiej można było wprowadzić nowatorski egzamin testowy?

Akceptacja dwu etapów – wskaźnik ten w bieżącym roku jest mniej miarodajny, gdyż w odróżnieniu od poprzedników ci uczniowie nie pisali testu i nie doświadczali trudów zdawania matury w obu postaciach.

ANALIZA WYPOWIEDZI NAUCZYCIELI O CECHACH JAKOŚCIOWYCH
KRYTERIÓW OCENIANIA

Tabela 3

Kryteria oceniania porównawczo w opinii nauczycieli z obu typów szkół

Zmienna ----- Parametr	Kod	Akceptacja kryteriów		Rzetelność kryteriów		Przydatność kryteriów		Operatywność kryteriów	
		Ogóln.	Zawod.	Ogóln.	Zawod.	Ogóln.	Zawod.	Ogóln.	Zawod.
Typ szkół									
1996	H'	6,0	6,0	5,1	5,1	6,0	6,0	5,5	5,5
Wartość	M'	4,8	4,9	3,9	3,9	4,8	4,8	4,0	4,0
	L'	3,3	3,1	1,5	1,5	2,6	2,8	0,6	1,7
Rozziew	-b'	-0,2	-0,3	-0,7	-0,7	-0,5	-0,4	-1,2	-1,2
1997	H''	6,0	6,0	6,0	5,0	6,0	6,0	6,0	5,5
Wartość	M''	4,8	4,7	4,2	3,9	4,5	4,6	4,0	3,5
	L''	3,3	3,3	3,1	1,2	3,0	2,9	2,1	0,8
Rozziew	-b''	-0,2	-0,2	-0,3	-0,6	-0,4	-0,4	-0,9	-1,6
1998	H	6,0	6,0	5,1	4,8	6,0	6,0	6,0	6,0
Wartość	M	4,6	4,3	4,0	3,7	4,6	4,8	4,6	4,6
	L	3,1	2,9	1,7	1,4	2,9	3,8	2,6	2,8
Rozziew	-b	-0,3	-0,5	-0,6	-0,8	-0,4	-0,1	-0,5	-0,5

Akceptacja kryteriów – w poprzednich latach miała miejsce wysoka zgodność wypowiedzi nauczycieli z obu typów szkół. W bieżącym roku, zwłaszcza w szkołach o profilu zawodowym, uwydatnił się spadek wartości średniej i zwiększył rozrzew opinii. Zmiany są większe od ewentualnego błędu pomiaru.

Rzetelność kryteriów – także i ta cecha w roku 1998 wypadła gorzej, przy czym zarówno średnie, jak i parametry rozrzewu są poniżej wartości oczekiwanych. Warto tu podkreślić, że struktura „Kryteriów” jest ustalona, lecz co roku zmienia się ich szczegółowa treść, zależna od meritum zadań.

Przydatność kryteriów – w przypadku tego wskaźnika sytuacja jest odwrotna, okazało się bowiem, że w szkołach zawodowych uznano kryteria za bardziej przydatne niż w ogólniakach. Można to tłumaczyć różnicą kompetencji nauczycielskich dotyczących umiejętności oceniania w próbach z obu typów szkół.

Operatywność kryteriów – wyraźny trend wzrostowy w ostatnim roku wskazuje na to, że nauczyciele opanowali już umiejętność sprawnego posługiwania się zaproponowanymi im kryteriami oceniania.

WNIOSKI Z EWALUACJI

Wprawdzie badania prowadzone były na próbnej maturze, lecz z całą pewnością wnioski rozciągnąć można również na maturę regulaminową. Początkowo można było sądzić, iż krytykowana forma problemowa egzaminu z matematyki zostanie

wyparta przez formę testową. Ostatnie jednak kierunki wskazują, że to raczej egzamin ustny byłby zastąpiony przez test. W oparciu o wyniki ewaluacji spróbujmy zastanowić się na tym, czy i w jakim stopniu forma problemowa matury została poprawiona. Gdyby odnieść wyniki wyłącznie do uzyskanych w roku 1997, to okazałoby się, że w bieżącym roku jest zdecydowanie lepiej. Trzeba jednak wyjaśnić, że słabsze wyniki ubiegłego roku były skutkiem próbnego wprowadzenia zadania z treścią. Z tego powodu konieczne jest odniesienie do ewaluacji z roku 1996. Niestety niewiele wskaźników z roku 1998 jest lepszych od tych sprzed dwóch lat. W takiej sytuacji najbardziej ogólny wniosek z przedstawionych tu wyników pomiaru jest oczywisty – matura w formie problemowej wymaga jeszcze **daleko idącej optymalizacji**, w przeciwnym wypadku trzeba będzie z tej formy zrezygnować.

Z wielu wątków dyskusji nad dalszą optymalizacją matury wybierzmy jedynie następujące:

- problem dopasowania skali trudności egzaminu maturalnego do poziomu uniwersum kulturowego i do realnych standardów osiągnięć, w kontekście zamiaru upowszechnienia wykształcenia średniego oraz uprzyśpieszenia możliwości studiowania;

- zagadnienie braku możliwości rzetelnego sprawdzenia osiągnięć szkolnych w formule problemowej za pomocą tylko trzech choćby najbardziej złożonych zadań oraz zastrzeżenia co do obiektywizmu oceniania prac pisemnych i nadmiernej losowości na egzaminie ustnym;

- problem odmiennych zestawów zadań na maturze ze względu na inne programy nauczania matematyki w różnych szkołach, a sprawa porównywalności ocen maturalnych przy stosowaniu tzw. konkursów świadectw w niektórych uczelniach.

Każda z powyższych kwestii dotyczy pewnych niedoskonałości w zakresie realizowania odpowiednio innej funkcji: cywilizacyjnej, sprawdzającej oraz różnicującej.

Funkcja cywilizacyjna osadzona jest w podstawach programowych, nad którymi wciąż trwają prace. Skutki zmian będą widoczne dopiero po wielu latach. Na razie w kompetencjach króluje odtwarzanie szablonów teoretycznych (np. analiza funkcji), przy zauważalnym braku umiejętności rozwiązywania prostych problemów praktycznych (np. analiza opłacalności kredytu). Rozwijanie abstrakcyjnego myślenia jest oczywiście celem, lecz musi ono być werbalizowane także przez realne, doskonałe zredagowane zadania z treścią. Niestety szkoła nie potrafi wykształcić kompetencji do rozwiązywania zadań z treścią, dlatego nie wolno „serwować” ich na maturze. Wyniki próbnej matury z 1997 roku są dowodem fiaska pomysłu wprowadzenia „zadań praktycznych”. Obecny dobór zadań, będących przekrojem przez program nauczania, jest wyraźnym dowodem poruszania się głównie w kręgu tradycyjnych kanonów. Najgorszą sprawą jest jednak rosnąca z roku na rok przepaść między zawyżonymi standardami wymagań a realnymi standardami osiągnięć. Zarówno abstrakcyjność, jak i poziomy trudności zadań są nadmierne, a przecież zapowiedziano, że egzamin z matematyki ma być obowiązkowy dla wszystkich.

Funkcje sprawdzająca i różnicująca wykluczają się wzajemnie. W tym tkwi źródło intuicyjnie wyczuwanej niesprawiedliwości, którą wyrażają przeciwstawne

pytania: – *Dlaczego w szkole dziennej zadania są trudniejsze niż w szkole dla pracujących?* bądź – *Dlaczego zestaw zadań jest identyczny, skoro uczniowie mieli programowo w danej szkole mniej godzin zajęć?* Wyjaśnić tu trzeba, iż matura była i jest nadal przede wszystkim sprawdzianem osiągnięć na końcu etapu kształcenia średniego. Wskutek dominacji funkcji sprawdzającej nie można żądać, ażeby była ona równocześnie podstawą weryfikacji na uczelnie. Tymczasem, wbrew opinii uczonych, projekt reformy oświaty nakłada na maturę absurdalne zadanie selekcji na studia. Przecież ignorowanie specyfiki ocen powoduje, że w konkursach świadectw przegrywają uczniowie z elitarnych szkół, mający ambitnych nauczycieli, którzy w ogóle nie wystawiają stopni celujących. Pojawiła się presja na większą obiektywizację i standaryzację egzaminu maturalnego. Wprowadzenie obu form pisemnych umożliwiłoby rozwiązanie powyższych sprzeczności. Konieczne jednak byłoby konsekwentne przestrzeganie zasady następującej – niech formuła problemowa sprawdza poziom wiedzy i umiejętności, a formuła testowa różnicuje i klasyfikuje uczniów według ich kompetencji. Z porównania wyników pomiarów przeprowadzonych trzykrotnie wynika, że problemowa matura próbna (miejmy nadzieję, że nie regulaminowa) różnicowała uczniów nie tylko ze względu na rodzaj szkoły, co jest zrozumiałe, lecz niestety także w poszczególnych latach, co już jest niedopuszczalne.

W funkcji sprawdzającej pełne uzasadnienie mają odmienne zestawy zadań problemowych w różnych typach szkół, przy czym dopasowanie do programów nauczania oraz ustalenie trafnych poziomów jest niezwykle trudnym zagadnieniem. Ponieważ formuła problemowa w zasadzie dubluje rolę szkoły, nadal aktualne pozostaje pytanie: *Czy od niej po prostu nie odstąpić?* Proponowana nowa formuła testowa, ażeby była klarownie różnicująca, musi opierać się na identycznym zbiorze zadań we wszystkich typach szkół, na realnym poziomie trudności, dopasowanym do populacji wszystkich maturzystów. Uzyskany z testu stopień stałby się swoistym języczkiem u wagi bądź przeciwwagą dla ocen rocznych na świadectwie. Trzeba jednak zaniechać wszelkich prób uproblemowienia testu, co już miało miejsce w 1997 r. i co wyniknęło z mylnej autosugestii autorów, jakoby za trzy punkty zadania testowe musiały być trudniejsze.

* * *

W pomiarze ewaluacyjnym nauczyciele występują w roli sędziów kompetentnych, stąd wyrażane przez nich opinie powinny być wiążące dla zespołu optymalizującego maturę z matematyki. Udział w próbnej maturze jest cennym doświadczeniem, które z pewnością ułatwi nauczycielom przygotowanie zarówno siebie, jak i uczniów z kolejnych roczników do planowanych zmian w regulaminie egzaminu. Uczniowie zdający są podmiotami tego społecznie doniosłego procesu. Z uwagi na szczególną rolę matury w życiu pokoleń wkraczających w dojrzałość, trzeba dołożyć wszelkich starań, ażeby uzyskała ona wyższą niż dotąd akceptację.