

dr Henryk Szaleniec

Polskie Towarzystwo Diagnostyki Edukacyjnej

dr Dorota Węziak-Białowolska

Szkoła Główna Handlowa w Warszawie

Czy e-ocenie może zapewnić większą rzetelność punktowania?

Wprowadzenie

Dobór egzaminatorów w egzaminach zewnętrznych tak, aby zapewnić jak najwyższą rzetelność oceniania, jest problemem niezwyklej wagi. Celem okręgowych komisji egzaminacyjnych, które zatrudniają ocenających, powinno być takie wybranie i przeszkolenie egzaminatorów, aby wyniki, jakie uzyskają uczniowie/absolwenci z egzaminu, w jak najmniejszym stopniu zależały od czynników związanych z doбором egzaminatorów. Z powyższym problemem borykają się nie tylko centralna i okręgowe komisje egzaminacyjne w Polsce, ale również ich odpowiedniki na całym świecie. Z tego też względu podejmowanych jest szereg badań, których celem jest rozpoznanie cech egzaminatorów determinujących jakość ich pracy w trakcie oceniania. W tego typu badaniach kładzie się szczególnie nacisk na wyznaczenie aspektów szkolenia, które mogą w znaczący sposób poprawić jakość pracy egzaminatorów. Kluczową sprawą dla rzetelności wyników egzaminu jest również sama organizacja procesu oceniania. Po pierwsze, powinien on zapewnić skuteczne monitorowanie i zapewnianie jakości oceniania. Po drugie, powinien być jak najmniej uciążliwy dla egzaminatora. Po trzecie, ocenający powinien mieć możliwość ciągłego kontaktu z przewodniczącym zespołu egzaminatorów w przypadku napotkania problemów trudnych do samodzielnego rozwiązania. Taki też cel przyświeca autorom projektu e-ocenie.

Jednym z najważniejszych czynników umożliwiających przeprowadzenie obiektywnego egzaminu jest zapewnienie takiego samego poziomu surowości w grupie egzaminatorów. Potwierdzają to wyniki badań prowadzonych obecnie lub w przeszłości między innymi w Wielkiej Brytanii, Stanach Zjednoczonych i Australii, w świetle których wybór egzaminatora nie pozostaje bez wpływu na jakość procesu oceniania (Meadows, Billington, 2007). Cechy, których wpływ na sposób, a zwłaszcza surowość oceniania, badano przede wszystkim to: poziom i kierunek wykształcenia, doświadczenie zawodowe (w pracy nauczyciela), doświadczenie egzaminacyjne, ale coraz częściej również cechy osobowości tzw. cechy behawioralne (mierzone za pomocą testów psychometrycznych), takie jak stosunek do rzetelnego oceniania, zaangażowanie, samodyscyplina, orientacja na osiągnięcia i inne (por. Meadows, Billington, 2007). Wyniki tych prac pokazują, że wpływ niemal każdej z nich można wyeliminować lub przynajmniej zminimalizować poprzez odpowiednie szkolenie egzaminatorów.

W Polsce (CKE) prowadzono również podobne badania. Wyniki tych badań¹ pokazały, że rezultaty egzaminu (bez względu na typ) obciążone były efektem egzaminatora. Efekt ten objawiał się w dwojaki sposób. Po pierwsze, wpływem przynależności egzaminatora do zespołu egzaminatorów na surowość/łagodność oceniania. Można przypuszczać, że był on wynikiem wpływu osobowości (w tym zwłaszcza poziomu surowości oceniania) przewodniczącego zespołu egzaminatorów na pracę podległych mu osób. Po drugie, za każdym razem² wystąpiły znaczące różnice w poziomach surowości oceniania między egzaminatorami z dużych i małych miejscowości.

Podsumowując, wyniki dotychczas przeprowadzonych badań (zarówno w Polsce, jak i na świecie) wskazują na konieczność uwzględnienia problematyki efektu egzaminatora w szkoleniu egzaminatorów, jak również ciągłego monitorowania procesu oceniania w trakcie jego trwania. Oba kierunki działań powinny koncentrować się nie tylko na sposobie posługiwania się kryteriami oceniania, ale również powinny zwracać uwagę na zachowanie stabilności sposobu oceniania i to nie tylko w czasie, ale i w przestrzeni. Innymi słowy, w czasie szkolenia należy zwracać uwagę egzaminatorom na konieczność zachowania takich samych kryteriów i takiego samego poziomu surowości oceniania w całym procesie oceniania oraz w całym kraju. Ponadto powinno się zadbać o to, żeby do wszystkich grup egzaminatorów docierały takie same informacje, co powinno pozwolić wyeliminować różnice w podejściu do oceniania między egzaminatorami z małych i dużych miejscowości.

Taki proces monitorowania (wraz z bieżącą korektą sposobu oceniania) przeprowadzony został przez Centralną Komisję Egzaminacyjną (CKE) w 2007 roku na terenie całego kraju (w 2008 roku proces monitorowania został znacznie zawężony w stosunku do roku 2007). Objął on wszystkie typy egzaminów zewnętrznych, tj. sprawdzian, egzamin gimnazjalny i maturę. Jego celem była ocena sposobu punktowania prac egzaminacyjnych, która miała prowadzić do zwiększenia porównywalności oceniania w skali kraju. Proces ten przeprowadzony został z wykorzystaniem narzędzi elektronicznych w następujący sposób. Dwie wybrane prace kalibrujące punktowane były przez wszystkich egzaminatorów w różnych momentach oceniania. Pierwsza z nich z reguły na początku oceniania, a druga – w połowie. Po ocenieniu pierwszej pracy i wprowadzeniu wyników on-line zespoły i poszczególni egzaminatorzy otrzymali informację zwrotną o poprawności i zgodności oceniania. Zaraz potem przeprowadzone zostały również krótkie szkolenia dla egzaminatorów zorientowane na omówienie popełnianych błędów oraz kwestii problematycznych (por. Czarnotta-Mączyńska, Lipska, 2007; Mytkowski, 2007). Przeprowadzone później analizy efektywności tej metody (por. Węziak, 2008) pokazały, że dzięki prowadzonej na bieżąco (on-line) możliwości weryfikacji sposobu oceniania różnice w poziomie

¹ Badania przeprowadzono w latach 2005 i 2006 na poziomie sprawdzianu (por. Dubiecka, Szaleniec, Węziak, 2006, s. 98) oraz na poziomie sprawdzianu, obu części egzaminu gimnazjalnego oraz matury z matematyki w roku 2007 (por. Węziak, 2008).

² Wyjątkiem był tu egzamin maturalny z matematyki.

surowości oceniania egzaminatorów w przypadku każdego egzaminu zmniejszyły się. Zmniejszeniu uległ również efekt wpływu przynależności egzaminatora do zespołu egzaminatorów. Te optymistyczne rezultaty pozostające w zgodzie z wynikami prac prowadzonych w innych krajach (por. Meadows, Billington, 2007) stanowiły silną przesłankę do rozpoczęcia prac związanych z wdrożeniem e-oceniania do systemu oceny egzaminów zewnętrznych w Polsce.

Od oceniania tradycyjnego do e-oceniania

E-ocenianie, czyli ocenianie z wykorzystaniem sieci (electronic marking), jest nowym rozwiązaniem zarówno pod względem technologicznym, jak i organizacyjnym. Główna różnica między dotychczasowymi rozwiązaniami a e-ocenianiem polega przede wszystkim na korzystaniu z innego nośnika, na którym dostarczana jest egzaminatorowi praca zdającego. Zamiast papierowego oryginału egzaminator otrzymuje obraz rozwiązania, który może przeglądać i punktować na ekranie komputera. Kolejna różnica dotyczy sposobu komunikacji. Dzięki zastosowaniu Internetu zapewniona jest interakcja egzaminatora z przewodniczącym zespołu egzaminatorów niezależnie od tego, czy ocenianie odbywa się w zespole, czy też w domu egzaminatora. Zmienia się też sposób monitorowania pracy egzaminatora – przeprowadzane jest ono na bieżąco, a egzaminator oceniający nierzetelnie może zostać odsunięty od dalszego oceniania, przy czym jego już ocenione prace łatwo mogą być skierowane do powtórnej oceny przez innego egzaminatora. Ponadto e-ocenianie umożliwia przejście od punktowania przez egzaminatorów całych prac obejmujących od kilku do kilkudziesięciu zadań do specjalizacji w ocenianiu poszczególnych zadań w danej sesji.

System e-oceniania został już na dużą skalę wprowadzony między innymi w Wielkiej Brytanii czy Stanach Zjednoczonych. Doświadczenia, jakie zdobyły w tym obszarze cztery duże komisje egzaminacyjne w tych krajach (AQA, OCR i EDEXCEL – Wielka Brytania; ETS – Stany Zjednoczone), są nie do przecenienia. Na podstawie dotychczas przeprowadzonych w tych krajach badań można stwierdzić, że przejście od oceniania tradycyjnego do e-oceniania wiąże się nie tylko ze zmianą organizacji procesu przygotowania prac do oceniania, ale również poprawia jego jakość. Badanie wpływu zmiany procesu oceniania na jego jakość pokazało, że jeśli tylko zmiana ta nie pociąga za sobą zmiany w obszarze sposobu przygotowania egzaminu, to pozwala zachować jego pierwotną trafność i podnieść jego rzetelność (por. Fowles, Adams 2005). Przykładowo Williams i van Lent (za Fowles, Adams 2005, s. 15) twierdzą, że wprowadzenie oceniania zadaniowego w systemie e-oceniania pozwala zwiększyć jego jakość oceniania dzięki:

- wprowadzeniu całkowitej anonimowości ocenianych zadań,
- uniknięciu efektu halo – sposób rozwiązania zadań wcześniejszych przez danego ucznia nie rzutuje na sposób oceny ocenianego zadania,
- równomiernemu rozłożeniu potencjalnych błędów w ocenianiu – zadania do oceny są przydzielane egzaminatorom w sposób losowy.

Williams i van Lent dodają, że zwłaszcza ten trzeci czynnik powinien wpłynąć na wzrost rzetelności oceniania. Podkreślają jednak również, że spodziewane efekty mogą różnić się w zależności od rodzaju egzaminu. Przykładowo Raikes (za Fowles, Adams 2005, s. 15) pisze, że dla egzaminatorów oceniających prace z matematyki ocenianie zadaniami jest nużące, co z pewnością obniża jakość ich pracy.

Biorąc pod uwagę powyższe argumenty, aby odpowiednio przygotować wprowadzenie do praktyki polskiej e-oceniania w latach 2007 – 2009, pod patronatem CKE realizowany jest projekt *Monitorowanie efektu egzaminatora oraz przygotowanie wprowadzenia oceniania z wykorzystaniem narzędzi elektronicznych (Electronic Marking)*³.

Głównym problemem pomyślnego wdrożenia e-oceniania w każdym z krajów, gdzie jest ono powszechnie stosowane, było uporanie się z uwarunkowanymi kulturowo przyzwyczajeniami egzaminatorów (Hudson, 2005). Egzaminatorzy oceniający prace od wielu lat mieli bezpośredni kontakt z pracą egzaminowanego na nośniku, na jakim została ona przez autora (ucznia, absolwenta) zapisana. Dlatego też we wstępnej fazie projektu przeprowadzono wiele spotkań z przewodniczącymi zespołów egzaminatorów, egzaminatorami i przedstawicielami wszystkich OKE. Łącznie we wszystkich spotkaniach uczestniczyło prawie 2000 osób.

W 2008 roku prowadzony jest pilotaż e-oceniania, w którym uczestniczyło 100 egzaminatorów z czterech egzaminów:

- sprawdzianu,
- egzaminu gimnazjalnego w części humanistycznej,
- egzaminu gimnazjalnego w części matematyczno-przyrodniczej,
- egzaminu maturalnego z matematyki.

Natomiast w 2009 roku przeprowadzony zostanie pilotaż z pięciu egzaminów na poziomie maturalnym.

W ramach projektu planowane jest zbadanie zależności efektu egzaminatora od strategii oceniania (ocenianie zadań w sekwencji całej pracy a ocenianie zadaniami) oraz od techniki oceniania – tradycyjnej i z wykorzystaniem Sieci (electronic marking). W trakcie pilotażu ocenianie tradycyjne z wykorzystaniem obydwu strategii przebiegać będzie w domu egzaminatora, natomiast e-ocenianie w ośrodkach egzaminacyjnych wyposażonych w pracownie komputerowe i przy wsparciu eksperta z DRS⁴. Dodatkowo e-ocenianie odbywać się będzie w systemie zadaniowym, a nie tak jak w dotychczasowej praktyce – całych prac. W pierwszym przypadku – ocenianie tradycyjne obejmować będzie bezpośredni kontakt egzaminatora z oryginalną pracą ucznia. Drugie rozwiązanie – nowe w polskim systemie oceniania – polegać będzie na ocenianiu wybranych zadań (wiązek zadań) przy korzystaniu z dostępu do obrazu zadania na ekranie monitora komputerowego. Rozwiązanie to pozwala na:

³ Projekt jest realizowany w ramach Programu Operacyjnego Kapitał Ludzki, Priorytet III, Działanie 3.2. Badania dotyczące podnoszenia jakości narzędzi systemu egzaminów zewnętrznych.

⁴ Data Service Limited jest wiodącym dostawcą systemów e-oceniania w Wielkiej Brytanii. Zapewnia serwis e-oceniania między innymi dla jednej z największych komisji egzaminacyjnych w Wielkiej Brytanii AQA.

- specjalizację egzaminatorów w ocenianiu wybranych działów tematycznych, a nawet na specjalizację w ocenianiu poszczególnych zadań,
- większą personalizację pracy i odpowiedzialność za wykonane zadanie – egzaminator identyfikuje się z pulą ocenianych zadań i nie jest anonimowy,
- zwiększenie komfortu pracy egzaminatora – docelowo praca we własnym domu, na własnym komputerze,
- zwiększenie zgodności punktowania – dzięki kontaktowi on-line (natychmiast i z egzaminatorami w całym kraju).

Dobór egzaminatorów i próby prac egzaminacyjnych do pilotażu

O zgłoszenie egzaminatorów do pilotażu poproszeni zostali dyrektorzy okręgowych komisji egzaminacyjnych. Każda komisja miała prawo zgłoszenia dwunastu egzaminatorów dla każdego z czterech egzaminów włączonego do pilotażu w 2008 roku. Egzaminatorzy powinni spełniać następujące warunki: posiadać kilkuletnie doświadczenie w ocenianiu oraz doświadczenie w funkcji przewodniczącego zespołu egzaminatorów, mieć za sobą praktykę w szkoleniu kandydatów na egzaminatorów i samych egzaminatorów. Zaproszeni do pilotażu egzaminatorzy od kwietnia uczestniczą w e-learningowym kursie, którego pierwszy moduł miał na celu nabycie podstawowych umiejętności w korzystaniu z platformy e-learningowej MOODLE.

Prace do oceny zostały wybrane z obszaru OKE Kraków z powiatów, których średni wynik był najbardziej zbliżony do średniej krajowej. Warunku takiego nie udało się zrealizować dla egzaminu maturalnego z matematyki ze względu na małą liczbę absolwentów wybierających ten przedmiot na poziomie rozszerzonym. Załączone poniżej tabele przedstawiają opis populacji prac egzaminacyjnych i próby prac planowanych i włączonych do pilotażu. Różnica pomiędzy próbą planowaną a zrealizowaną wynika z braku nawyków zdających do pisania rozwiązań w wyznaczonym miejscu na arkuszu egzaminacyjnym.

Tabela 1. Statystyki dla populacji i próby wybranej do pilotażu dla sprawdzianu

Statystyki	Kraj	Powiat jasielski	Prace zakwalifikowane do e-oceny
Liczba prac	452 365	1 496	1 048
Średnia	26,57	26,55	26,86
Mediana	28	28	28
Dominanta	33	34	34
Odchylenie standardowe	7,82	7,78	7,87

Tabela 2. Statystyki dla populacji i próby wybranej do pilotażu dla egzaminu gimnazjalnego część humanistyczna

Statystyki	Kraj	Powiat tarnowski	Prace zakwalifikowane do e-ocenywania
Liczba prac	506 193	3 087	529
Średnia	31,48	31,49	32,54
Mediana	33	33	34
Dominanta	37	37	37
Odchylenie standardowe	9,78	9,28	8,67

Tabela 3. Statystyki dla populacji i próby wybranej do pilotażu dla egzaminu gimnazjalnego część matematyczno-przyrodnicza

Statystyki	Kraj	Powiat kolbuszowski	Prace zakwalifikowane do e-ocenywania
Liczba prac	505 759	994	744
Średnia	25,31	25,27	26,94
Mediana	23	23	26
Dominanta	18	17	16
Odchylenie standardowe	10,22	9,97	10,07

Tabela 4. Statystyki dla populacji i próby wybranej do pilotażu dla egzaminu maturalnego z matematyki na poziomie rozszerzonym

Statystyki	Kraj	Wybrane powiaty: puławski, Biała Podlaska, myślenicki, Kraków, Nowy Sącz, ropczycko-sędziszowski	Prace zakwalifikowane do e-ocenywania
Liczba prac	27538	2551	522
Średnia	24,9	26,20	21,95
Mediana	25	27	21
Dominanta	15	15	15
Odchylenie standardowe	11,88	12,12	12,43

Model badań i narzędzia badawcze

W trakcie badań pilotażowych zostaną zastosowane: kwestionariusz egzaminatora, kwestionariusz przewodniczącego zespołu egzaminatorów oraz prace egzaminowanych z sesji egzaminacyjnej 2007 roku. Do e-oceniaia zostało wybrane:

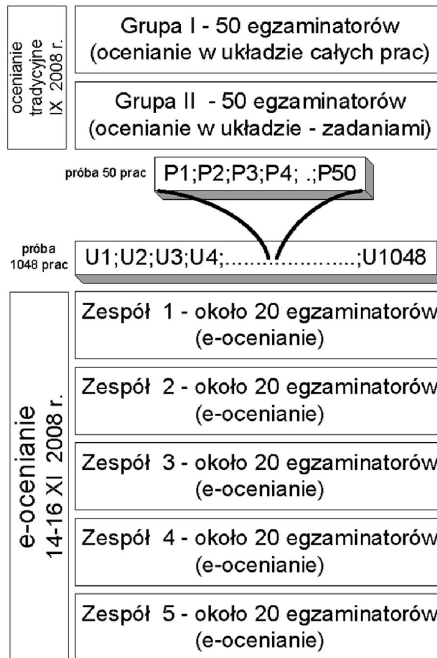
1. dla sprawdzianu – 1048 prac,
2. dla egzaminu gimnazjalnego w części humanistycznej – 529 prac,
3. dla egzaminu gimnazjalnego w części matematyczno-przyrodniczej – 744 prace,
4. dla egzaminu maturalnego z matematyki na poziomie rozszerzonym – 522 prace.

Zróznicowana liczba prac uwarunkowana była różnym czasem oceny zadań w poszczególnych egzaminach oraz techniczną możliwością skanowania prac. Prace, w których egzaminowani udzielali odpowiedzi poza obszarem wyznaczonym do tego celu, musiały być odrzucone z powodów technicznych.

Planuje się, że w trakcie pilotażu e-oceniaia każde zadanie będzie oceniane przez pięciu egzaminatorów.

Do oceniaia tradycyjnego zostało wybranych:

1. dla sprawdzianu – 50 prac spośród 1048,
2. dla egzaminu gimnazjalnego w części humanistycznej – 30 prac spośród 529,
3. dla egzaminu gimnazjalnego w części matematyczno-przyrodniczej – 36 prac spośród 744,
4. dla egzaminu maturalnego z matematyki – 30 prac spośród 522.



Rysunek 1. Schemat modelu badań na przykładzie sprawdzianu

Planuje się, że każda praca punktowana w sposób tradycyjny zostanie oceniona przez 100 egzaminatorów uczestniczących w pilotażu. Ponadto każdy 100-osobowy zespół został podzielony losowo na dwie grupy. Pierwsza grupa otrzyma prace przygotowane do oceny w dotychczasowym układzie (kolejne prace poszczególnych zdających). Druga grupa otrzyma do oceny materiał uporządkowany zadaniami (osobne broszury z rozwiązaniami poszczególnych zadań). Prace oceniane w tradycyjny sposób stanowiąc będą zakotwiczenie do analiz efektu egzaminatora w e-oceniu.

Monitorowanie efektu egzaminatora w pilotażu e-ocenia

Na podstawie danych uzyskanych z pilotażu do e-ocenia przeprowadzona zostanie ocena efektu egzaminatora. Celem tego przedsięwzięcia jest weryfikacja hipotezy głoszącej, że efekt egzaminatora będzie jednakowo silny w ocenianiu tradycyjnym i e-oceniu. Rezultaty dotychczas przeprowadzonych prac nad efektem egzaminatora w Polsce (por. Dubiecka, Szaleniec, Węziak, 2006; Węziak, 2008) pozwalają oczekiwać, że efekt egzaminatora w e-oceniu będzie mniejszy niż w ocenianiu tradycyjnym. Tym samym pozwoli to na odrzucenie powyżej przedstawionej hipotezy badawczej. Taki rezultat stanowiłby poważny argument za wprowadzeniem systemu e-ocenia do egzaminów zewnętrznych w Polsce.

Procedura badawcza

W celu weryfikacji postawionej hipotezy badawczej niezbędne jest zmierzenie efektu egzaminatora występującego w ocenianiu tradycyjnym i w e-oceniu, a następnie porównanie wielkości tych dwóch efektów.

Efekt egzaminatora obejmuje szeroką kategorię efektów generujących wariancję wyników oceniania, której źródłem są egzaminatorzy, a nie zmienność poziomu ocenianej cechy. Do najczęściej spotykanych efektów należą:

1. *łagodność lub surowość oceniania,*
2. *efekt halo,*
3. *tendencja centralna w ocenianiu,*
4. *tendencja ocen ekstremalnych* (Dubiecka, Szaleniec, Węziak, 2006, s. 100).

Niemniej w ramach projektu e-ocenie analizą objęta zostanie tylko pierwsza grupa efektów.

Łagodność oceniania definiuje się jako stałą cechę związaną z osobowością egzaminatora i występującą niezależnie od sytuacji egzaminacyjnej. Ocena łagodnego egzaminatora jest przeciętnie wyższa niż wynik wynikający z ustalonych kryteriów zewnętrznych. O surowości oceniania mówi się w przypadku, gdy przeciętny wynik oceniania jest poniżej średniej wartości ocenianej zgodnie z ustalonymi zewnętrznymi kryteriami cechy (Dubiecka, Szaleniec, Węziak, 2006, s. 100).

Do oceny poziomu surowości egzaminatorów zastosowane zostanie wieloaspektowe skalowanie Rascha (Many-Facet Rasch Model). Metoda ta została już przetestowana z pozytywnym skutkiem w poprzednich projektach prowadzonych przez CKE (por. Węziak, 2005; Dubiecka, Szaleniec, Węziak, 2006, Węziak, 2008),

zaś jej szczegółowy opis zamieszczono w: Węziak (2005); Dubiecka, Szaleniec, Węziak (2006). Warto jednak przypomnieć, że metoda ta pozwala na kwantyfikację cechy nieobserwowalnej bezpośrednio, jaką jest surowość oceniania egzaminatora. Kwantyfikacja przeprowadzana jest na podstawie ocen (punktów) przyznanych przez danego egzaminatora za określone zadania określonym uczniom. Ponadto obejmuje przypisanie każdemu egzaminatorowi wartości liczbowej na przedziałowej skali odpowiadającej charakteryzującemu go poziomowi surowości oceniania. Liczby pozycje na skali wyrażone są w jednostkach zwanych logitami i zwykle zawierają się w przedziale $<-3; 3>$. Dzięki możliwości zmierzenia i wyrażenia za pomocą skali surowości oceniania możliwe staje się porównanie egzaminatorów między sobą.

Model analiz

Na podstawie ocen (liczby punktów) przyznanych przez egzaminatorów w ocenianiu tradycyjnym, jak i w e-ocenianiu planuje się przeprowadzenie następujących analiz:

1. porównanie poziomu surowości oceniania danego egzaminatora i zespołu egzaminatorów ze średnią,
2. przeprowadzenie analizy wariancji poziomów surowości oceniania egzaminatorów,
3. przeprowadzenie analizy asymetrii rozkładu poziomów surowości oceniania,
4. zbadanie, czy istnieje znaczący wpływ przynależności egzaminatora do zespołu egzaminatorów na poziom jego surowości oceniania,
5. zbadanie, czy cechy charakteryzujące egzaminatora wpływają na jego poziom surowości oceniania.

Na podstawie wyników procesu monitorowania przeprowadzonego przez CKE w roku 2007 przyjęto, że w projekcie rejestracji podlegać będą następujące cechy charakteryzujące egzaminatorów:

1. stopień awansu zawodowego,
2. przedmiot, którego uczy egzaminator,
3. liczba lat pracy w zawodzie nauczyciela,
4. krotność pracy na stanowisku egzaminatora,
5. pełniona funkcja w procesie oceniania,
6. wielkość miejscowości, w której pracuje (główne miejsce pracy) egzaminator,
7. przynależność do zespołu egzaminatorów w trakcie e-oceniania.

Ponadto dodane zostanie pytanie dotyczące stosunku egzaminatora do korzystania z technologii IT, jak również rejestrowaniu podlegać będzie czas oceniania każdego zadania w e-ocenianiu i czas oceniania zadania w ocenianiu tradycyjnym.

W trakcie badań podjęta zostanie także próba określenia:

1. w jakim stopniu uczestniczący w pilotażu egzaminatorzy akceptują przechodzenie od oceniania tradycyjnego do e-oceniania,
2. w jakim stopniu uczestniczący w pilotażu przewodniczący zespołów egzaminatorów akceptują nowy sposób monitorowania i zapewniania jakości oceniania,
3. czy dostarczanie zadań na ekran komputera egzaminatora jest płynne i zapewnia komfort pracy,
4. czy uzgadnianie wyników oceny pracy jednego egzaminowanego przez dwóch lub więcej egzaminatorów nie jest trudniejsze niż w przypadku dotychczasowej praktyki.

Taki wybór zmiennych i zagadnień wynikał z dotychczasowych doświadczeń autorów projektu, jak również uwzględniał wyniki badań prowadzonych w innych krajach. W przypadku cechy *stosunek do korzystania z technologii IT* jej rejestracji podjęto się ze względu na zastosowanie nowego narzędzia (oceniania w formie on-line).

W pierwszym etapie prac na podstawie wyników oceniania tradycyjnego oszacowane zostaną poziomy umiejętności uczniów, poziomy trudności zadań i poziomy surowości oceniania egzaminatorów. W kroku drugim te same wielkości oszacowane zostaną na podstawie wyników e-oceniania. Należy zaznaczyć, że w tym etapie dla zapewnienia porównywalności wyników odnoszących się do poziomów surowości oceniania egzaminatorów, poziomy umiejętności uczniów, których prace ocenianie będą w ocenianiu tradycyjnym, jak również poziomy trudności zadań wykorzystanych w ocenianiu tradycyjnym przyjęte zostaną jako znane (na poziomie wynikającym z analiz etapu pierwszego). Takie postępowanie pozwoli na porównanie efektów egzaminatora w ocenianiu tradycyjnym i w e-ocenianiu, czyli porównanie poziomów surowości oceniania w obu podejściach.

Przewiduje się, że przeprowadzone analizy pozwolą uzyskać odpowiedzi na następujące pytania badawcze:

- Czy przeciętna surowość oceniania w ocenianiu tradycyjnym i w e-ocenianiu jest taka sama?
- Czy przeciętna surowość oceniania w ocenianiu tradycyjnym z wykorzystaniem różnych strategii (ocenianie całymi pracami w porównaniu z ocenianiem zadaniami) jest taka sama?
- Czy w e-ocenianiu występuje efekt wpływu przynależności egzaminatora do zespołu egzaminatorów na poziom surowości oceniania (efekt wpływu przewodniczącego zespołu egzaminatorów – PZE na poziom surowości oceniania członków jego zespołu)?
- Czy te same cechy charakteryzujące egzaminatorów (w tym zwłaszcza wielkość miejscowości zamieszkania) wpływają na poziom surowości oceniania w ocenianiu tradycyjnym i w e-ocenianiu?
- Czy postawa egzaminatora związana z zastosowaniem narzędzi informatycznych wpływa na sposób oceniania w e-ocenianiu?
- Jaki jest wpływ sposobu oceniania na czas trwania tego procesu?

- W jakim stopniu uczestniczący w pilotażu e-oceniający egzaminatorzy akceptują przechodzenie od oceniania tradycyjnego do e-oceniającego?
- W jakim stopniu uczestniczący w pilotażu przewodniczący zespołów egzaminatorów akceptują nowy sposób monitorowania i zapewniania jakości oceniania?
- Czy dostarczanie zadań na ekran komputera egzaminatora jest płynne i zapewnia komfort pracy?
- Czy uzgadnianie wyników oceny pracy jednego egzaminowanego przez dwóch lub więcej egzaminatorów nie jest trudniejsze niż w przypadku dotychczasowej praktyki?

Wyniki

Projekt jest w fazie realizacji. Planuje się, że ocenianie w sposób tradycyjny przeprowadzone zostanie we wrześniu 2008 roku, natomiast e-oceniającego w listopadzie. Jak dotąd, przygotowano ośrodki e-oceniającego w pięciu szkołach na terenie Krakowa. Ponadto wybrane zostały prace do e-oceniającego i próba prac do oceniania tradycyjnego w każdym z czterech egzaminów. Oszacowano również poziomy trudności zadań dla każdego z egzaminów oraz określono poziomy umiejętności uczniów, których prace wybrane zostały do pilotażu (szczegółowy opis analiz przedstawiono w: Węziak, 2008). Poziomy trudności zadań oszacowane zostały na podstawie 10% próby prostej losowanej z populacji prac wszystkich uczniów piszących dany egzamin, natomiast poziomy umiejętności uczniów, których prace zostały włączone do pilotażu, oszacowano przy przyjęciu parametrów zadań oszacowanych na podstawie próby krajowej. Uzyskane wyniki dotyczące parametrów zadań zestawiono w tabelach 5 – 7.

Tabela 5. Zadania otwarte arkusza standardowego sprawdzianu 2007 i ich punktacja

Kod czynności	Liczba punktów	Trudność w jednostkach logit	Kod czynności	Liczba punktów	Trudność w jednostkach logit
Z21_1	0 – 1	-1,3	Z24	0 – 1	0,59
Z21_2	0 – 1	-1,73	Z25_1	0 – 1	1,96
Z21_3	0 – 1	-1,33	Z25_2	0 – 1	-1,89
Z22_1	0 – 1	-0,43	Z26_1	0 – 1 – 2 – 3	-0,07
Z22_2	0 – 1	-1,33	Z26_2	0 – 1	1,18
Z22_3	0 – 1	-0,65	Z26_3	0 – 1 – 2	-0,9
Z22_4	0 – 1	-0,51	Z26_4	0 – 1	-0,34
Z23	0 – 1 – 2	-1,25	Z26_5	0 – 1	-0,38

Źródło: Węziak, 2008.

Tabela 6. Zadania otwarte arkusza standardowego egzaminu gimnazjalnego 2007 i ich punktacja

Część humanistyczna			Część matematyczno-przyrodnicza		
Kod czynności	Liczba punktów	Trudność w jednostkach logit	Kod czynności	Liczba punktów	Trudność w jednostkach logit
z21	0 – 1	0,69	z26	0 – 1	-3,51
z22	0 – 1	0,11	z27	0 – 1 – 2	-0,69
z23	0 – 1	0,71	z28	0 – 1 – 2	0,27
z24	0 – 1	-0,36	z29.1	0 – 1	-0,50
z25	0 – 1	1,12	z29.2	0 – 1	0,46
z26	0 – 1	0,45	z30.1	0 – 1	-0,02
z27	0 – 1	1,68	z30.2	0 – 1	1,61
z28	0 – 1	0,92	z30.3	0 – 1	0,10
z29.1.1	0 – 1	1,02	z30.4	0 – 1	2,23
z29.1.2	0 – 1	-3,05	z31.1	0 – 1	-3,43
z29.2	0 – 1	-2,7	z31.2	0 – 1	-0,67
z29.3	0 – 1	-0,13	z31.3	0 – 1	1,95
z29.4	0 – 1	0,71	z32.1	0 – 1	3,50
z30.1.1	0 – 1	-1,45	z32.2	0 – 1	0,24
z30.1.2	0 – 1	0,41	z32.3	0 – 1	0,88
z30.1.3	0 – 1	1,68	z32.4	0 – 1	4,00
z30.1.4	0 – 1	1,24	z33.1	0 – 1	0,98
z30.1.5	0 – 1	2,05	z33.2	0 – 1	0,99
z30.1.6	0 – 1	0,37	z33.3	0 – 1	1,16
z30.2.1	0 – 1	-0,43	z33.4	0 – 1	1,56
z30.2.2	0 – 1	-0,7	z34.1	0 – 1	-0,40
z30.2.3	0 – 1	-0,5	z34.2	0 – 1	-0,25
z30.3.1	0–1–2–3	2,18	z34.3	0 – 1	-1,92
z30.3.2	0 – 1	-0,53			
z30.4.1	0–1–2	2,08			
z30.4.2	0 – 1	2,42			

Źródło: Węziak, 2008.

Tabela 7. Poziom trudności zadań otwartych arkusza standardowego egzaminu maturalnego z matematyki 2007 na poziomie rozszerzonym

Kod czynności	Trudność w jednostkach logit	Kod czynności	Trudność w jednostkach logit
z1_1	-1,01	z5_3	0,97
z1_2	-0,44	z5_4	1,8
z1_3	-0,95	z5_5	2,52
z1_4	-1,24	z5_6	3,28
z1_5	-0,5	z5_7	3,22
z2_1	-0,85	z8_1	-0,46
z2_2	-2,24	z8_2	0,1
z2_3	-1,05	z8_3	-0,86
z2_4	0,13	z10_1	-3,39
z2_5	-0,13	z10_2	-1,39
z5_1	-2,08	z10_3	-0,03
z5_2	-1,41	z10_4	-0,11

Źródło: Węziak, 2008.

Podsumowanie

Od 2002 roku, kiedy to odbyły się pierwsze w naszym kraju egzaminy zewnętrzne, wiele zmieniło się zarówno w teorii, jak i praktyce egzaminowania. Zmiany te były wynikiem szybkiej ekspansji technologii informatycznej we wszystkich obszarach indywidualnej i społecznej aktywności. 4 marca 2008 roku w 400 szkołach został przeprowadzony przez Wyższą Szkołę Humanistyczno-Ekonomiczną w Łodzi oraz Wydawnictwa Szkolne i Pedagogiczne pierwszy próbny e-egzamin na poziomie gimnazjalnym w części matematyczno-przyrodniczej (Stańdo, Szalaniec, 2008). Wprawdzie testy przeprowadzane komputerowo (Computer-Based Testing – CBT), a szczególnie komputerowe testy adaptacyjne (computerized adaptive test – CAT) nie są jeszcze powszechnie wykorzystywane w egzaminach doniosłych, ale są coraz powszechniej stosowane w badaniach edukacyjnych i egzaminach o niskiej doniosłości dla ucznia, a mających istotne znaczenie diagnostyczne. Warto zauważyć, że systemy komputerowych egzaminów skorzystały nie tylko z najnowszej technologii informatycznej, ale także wzięły z tradycyjnych egzaminów to, co było w nich najlepsze. W egzaminach przeprowadzanych komputerowo (CBT) a także coraz częściej w tradycyjnych egzaminach pisanych na papierze zadania otwarte rozszerzonej odpowiedzi oceniane są przez egzaminatorów (ekspertów) na ekranie komputera. W takim rozwiązaniu, które w naszym projekcie nazwaliśmy e-ocenianiem, egzaminator jest nadal kluczowym ogniwem w zapewnianiu jakości oceniania. Niezależnie od przyjętego rozwiązania dążenie do minimalizacji efektu egzaminatora pozostaje ważnym zadaniem w aspekcie zwiększania rzetelności egzaminowania. Przygotowując się do wdrożenia e-oceniania w polskim systemie, warto zadbać,

aby ta innowacja z jednej strony zapewniła postęp w egzaminowaniu, a z drugiej – w pełni wykorzystwała to, co najcenniejsze w dotychczasowej praktyce. Zanim zostanie przygotowany wstępny projekt wdrażania e-oceniań w całym systemie egzaminacyjnym na rok 2009 planowany pilotaż e-oceniań w zakresie przedmiotów maturalnych, w którym zostaną wykorzystane rezultaty badań prowadzonych w roku 2008.

Bibliografia:

1. Czarnotta-Mączyńska J., Lipska M., *Zgodność punktowania zadań otwartych przez egzaminatorów egzaminu gimnazjalnego 2007. Raport z monitoringu procesu sprawdzania*, 2007.
2. Dubiecka A., Szalaniec H., Węziak D., *Efekt egzaminatora w egzaminach zewnętrznych*, [w:] Niemierko B., Szmigel M. K. [red.], *O wyższą jakość egzaminów zewnętrznych. Część I. Etyka egzaminacyjna i zagadnienia ogólne*. XII krajowa konferencja z cyklu „Diagnostyka Edukacyjna”, PTDE 2006.
3. Fowles D., Adams C., 2005, *Does assessment differ when e-marking replaces paper-based marking?*, IAEA paper v3(cra+def).doc.
4. Hudson G., 2005 *Implementing Large-scale Electronic Marking in UK*, IAEA conference, Abuja.
5. Meadows M., Billington L., 2007, *NAA Enhancing the Quality of Project: Final Report for Research on Marker Selection*, National Assessment Agency, UK.
6. Mytkowski D., *Zgodność punktowania zadań otwartych przez egzaminatorów matury 2007. Raport z monitoringu procesu sprawdzania i oceniania arkuszy egzaminacyjnych*, Centralna Komisja Egzaminacyjna, Warszawa 2007.
7. Szalaniec H., Stańdo J., *E-egzamin 2008, Raport z projektu edukacyjnego*. http://www.wsip.com.pl/index.php/wsip_site/o_firmie/aktualnosci/e_egzamin_raport
8. Węziak D., *Zastosowanie wieloczynnikowego skalowania Rascha do porównania sposobu oceniania egzaminatorów*, [w:] Niemierko B., Szmigel M. K. [red.], *Holistyczne i analityczne metody diagnostyki edukacyjnej, perspektywy informatyczne egzaminów szkolnych*. XI krajowa konferencja z cyklu „Diagnostyka Edukacyjna”, PTDE 2005.
9. Węziak D., *Monitorowanie oceniania w sesji egzaminacyjnej 2007*, materiały konferencyjne, Konferencja CKE, Od oceniania tradycyjnego do e-oceniań, 21 lutego 2008, Warszawa 2008.