

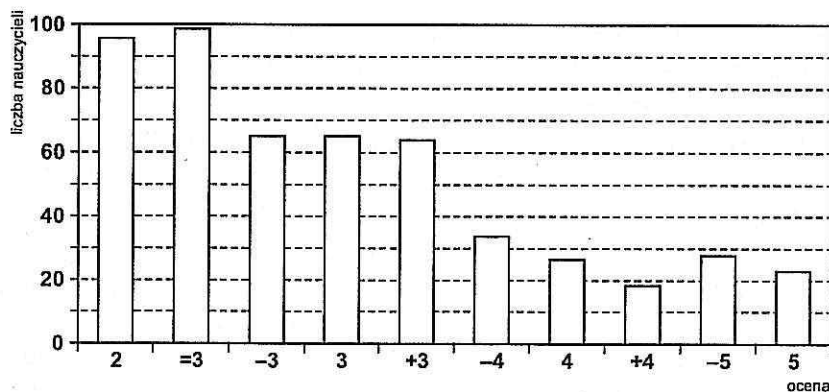
Jakub Lis

Wydział Badań i Analiz Okręgowej Komisji Egzaminacyjnej we Wrocławiu

Jak weryfikować pracę egzaminatorów w trakcie sesji egzaminacyjnej?

Jednym z najistotniejszych elementów systemu zewnętrznych egzaminów jest jakość pracy egzaminatora. Jak ważny jest to element systemu oceniania, świadczyć może artykuł Stanisława Jakubowicza *Poprawianie klasówki – sztuka, czy rzemiosło?* (1991) poświęcony ocenianiu prac przez nauczycieli. Autor prezentuje wyniki oceny tej samej pracy uczniowskiej przez 520 nauczycieli przedmiotu. Wyniki prezentuje poniższy wykres.

Wykres 1.
Rozkład wyników oceny jednej pracy przez 520 nauczycieli



Zaskakuje fakt, że jedno rozwiązanie zadania w oczach poprawiających uzyskało praktycznie każdą z możliwych ocen. A trzeba pamiętać, że w przypadku egzaminów zewnętrznych rozwiązań zadań może być tyle, ilu uczniów zdawało egzamin.

W OKE Wrocław, w czasie każdej sesji egzaminacyjnej sprawdzanych jest około 53 tys. prac uczniów szkoły podstawowej oraz około 56 tys. prac z każdej części egzaminu gimnazjalnego. Łącznie daje to liczbę około 165 tys. prac sprawdzanych i ocenianych w ciągu 6 dni. W ramach przygotowań do przeprowadzenia sprawdzianu i egzaminu gimnazjalnego w OKE Wrocław przeszkolonych zostało ponad 2 tys. egzaminatorów, a około 1,5 tys. sprawdzało prace uczniów. Przy takiej liczbie osób zatrudnionych do sprawdzania bardzo istotne staje się weryfikowanie pracy pojedynczych egzaminatorów. Najmniejszy błąd decyduje o wynikach uczniów z kilku, czasem kilkunastu szkół. Przewodniczący każdego, liczącego 20 osób, zespołu zobowiązany jest do weryfikacji merytorycznej 10% prac ocenionych przez swoich egzaminatorów. W OKE Wrocław zatrudnianych jest około 120 przewodniczących. Można przyjąć zatem, że każdy z nich weryfikuje średnio ponad 100 prac. Wydaje się, że istotnym ułatwieniem pracy przewodniczącego, z oczywistą korzyścią dla systemu egzaminacyjnego, byłoby opracowanie metody, która pozwalałaby automatycznie wskazywać egzaminatorów sprawdzających zadania niezgodnie z przyjętymi kryteriami, których pracę następnie można byłoby wnikliwie przeanalizować.

Prezentowany komunikat opisuje jedną z prób znalezienia metody weryfikacji oceniania w trakcie sesji egzaminacyjnej. Zawiera on zarówno rozważania teoretyczne, jak i praktyczne wnioski do realizacji postawionych zadań. Taki podział jest wynikiem zainteresowania zarówno teorią oceniania, jak i technicznymi możliwościami wdrożenia prezentowanych rozwiązań.

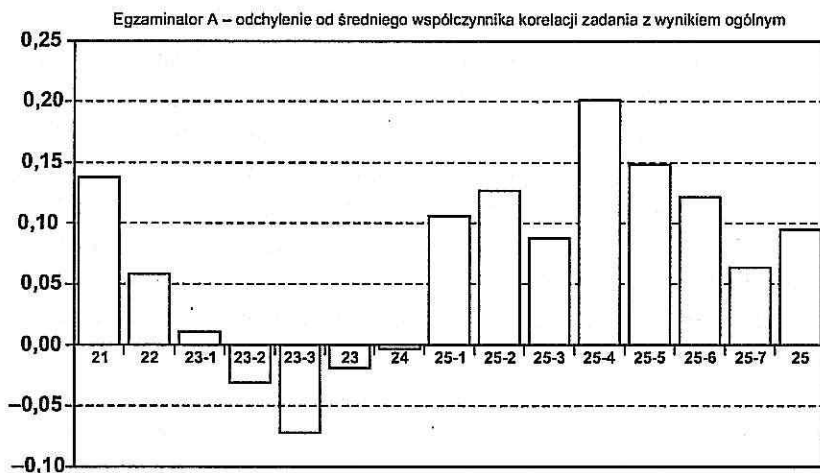
Badania przeprowadzono, wykorzystując wyniki uczniów piszących egzaminy w roku 2003, już po zakończeniu akcji egzaminacyjnej. Nie miały one na celu weryfikacji pracy egzaminatorów w trakcie sesji, a jedynie wypracowanie narzędzi na przyszłość.

Poszukiwanie wskaźnika najbardziej adekwatnego do postawionych celów obejmowało kilka różnych miar. Pierwsza miara obejmowała średnie wyniki prac ocenionych przez jednego egzaminatora. Przyjęto trzy wskaźniki: średni wynik egzaminu w województwie, średni wynik wszystkich prac sprawdzanych przez egzaminatora oraz średnią jednego ocenianego przez niego oddziału. Ponieważ każdy egzaminator statystycznie sprawdza ponad 100 prac, wydawało się, że porównanie to będzie miało sens ze względu na dość dużą próbę. Oszacowanie to było jednak na tyle grube, że na tej podstawie nie dało się wyciągnąć żadnych praktycznych wniosków. Przykładem może być tu jeden z egzaminatorów, którego średnie wyniki odbiegały o kilka punktów od przyjętego prognozy. Okazało się jednak, że egzaminator ten sprawdzał prace uczniów prawie wyłącznie z dużych miast. Dalsze obserwacje pokazały, że średnie wyniki są bardzo mocno skorelowane z wielkością miejscowości, z jakiej pochodzą uczniowie. We wszystkich dotychczasowych statystykach średni ogólny wynik z terenów wiejskich był o 2–3 punkty niższy od wyniku dużych aglomeracji miejskich. Tendencja ta zyskała sobie miano „specyfiki dużej zbiorowości”. Stąd wniosek, że chcąc rzetelnie oceniać pracę egzaminatorów, trzeba by za każdym razem brać pod

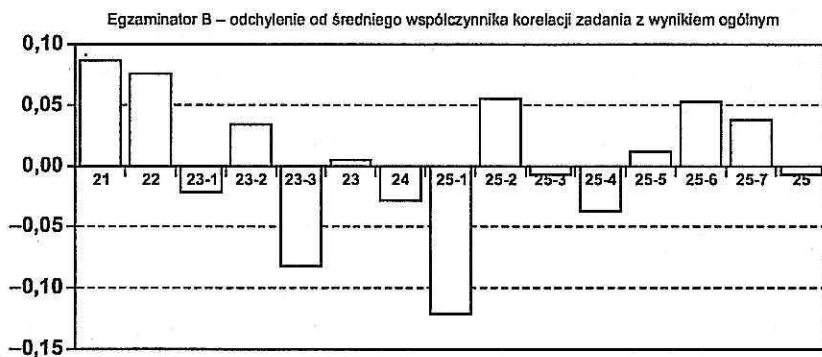
uwagę uwarunkowania społeczne grupy uczniów ocenianych przez egzaminatora. Pociąga to za sobą konieczność jednorodnego doboru prac dla zespołu egzaminatorów lub każdorazowe szczegółowe opisywanie sprawdzanej społeczności. Drugie rozwiązanie, ze względu na zamierzony cel elektronicznej weryfikacji, jest praktycznie niewykonalne, pierwsze natomiast pociąga za sobą ogromne trudności i koszty związane z doбором prac dla zespołu egzaminatorów. Jednak nawet jednorodny, ze względu na warstwy, dobór prac nie gwarantuje, że wyniki poszczególnych kryteriów, jak również całych zadań powinny korelować z ogólnymi wynikami uczniów. W 2002 r. OKE współtworzyła wraz z Dolnośląską Szkołą Wyższą Edukacji raport pt. *Obraz gimnazjów wrocławskich po trzech latach ich funkcjonowania* (Gołębnik, 2002). Jednym z wniosków tego raportu był fakt, że nie można porównywać statystycznych wyników szkół z różnych dzielnic Wrocławia, bez uwzględnienia innych aspektów pracy szkoły. Analogicznie, chcąc rzetelnie oceniać wyniki pracy egzaminatorów, powinno się rozbić całą populację na podpopulacje, a najlepiej na poszczególnych uczniów.

Ponieważ średnie wyniki nie pozwalały na rzetelną ocenę pracy egzaminatora, podjęto próbę spojrzenia na problem pod kątem mocy różnicującej poszczególnych zadań jako wskaźnika niezależnego od ogólnych wyników uczniów. Ponieważ egzaminator otrzymywał do oceny prace uczniów z losowo wybranych oddziałów, można założyć, że oceniał prace zarówno dobrych, jak i słabych uczniów. Zatem sposób doboru prac nie powinien mieć zbytniego wpływu na różnicowanie wyników uczniów. Arkusz zadań złożony jest z pojedynczych zadań, które następnie są oceniane według różnych kryteriów. W celu analizy pracy egzaminatorów każdemu z nich przypisano zbiór złożony ze wskaźników mocy różnicującej każdego kryterium, wyliczonych na podstawie ocenionych przez niego prac. Za punkt odniesienia przyjęto średni wojewódzki wskaźnik mocy różnicującej poszczególnych kryteriów. Różnica pomiędzy wynikiem wzorcowym a wynikiem egzaminatora stanowiła cechę obrazującą odchylenia od przyjętego progu. Otrzymano w ten sposób zbiór wartości charakteryzujących każdego egzaminatora. Ponieważ część kryteriów w zadaniach otwartych była wielopunktowa, do oceny mocy różnicującej wykorzystano współczynnik korelacji liniowej Pearsona, korelując wyniki za poszczególne kryteria z ogólnym wynikiem ucznia. Pomimo iż około połowa z możliwych do zdobycia przez ucznia punktów pochodzi z zadań zamkniętych, na które nie ma wpływu egzaminator, przyjęto liczenie mocy różnicującej kryteriów i zadań w odniesieniu do ogólnego wyniku ucznia. Rozwiązanie to miało na celu uwzględnienie rzeczywistego wyniku ucznia, a nie tylko oceny wystawionej przez egzaminatora. W celu łatwiejszej analizy zgromadzonego materiału przedstawiono graficznie odchylenia wskaźników mocy różnicującej zadań ocenianych przez egzaminatora. Przyjęto również, że przewodniczący zespołu egzaminatorów otrzyma jedynie graficzne przedstawienie „charakterystyki” każdego ze swoich pracowników i oceni jej przydatność do oceny ich pracy. Przykładowe „charakterystyki” dwóch egzaminatorów prezentują wykresy 2 i 3.

Wykres 2.
Przykładowy wykres obrazujący odchylenia
od średniego współczynnika mocy różnicującej



Wykres 3.
Przykładowy wykres obrazujący odchylenia
od średniego współczynnika mocy różnicującej



Na wykresie 2 możemy zaobserwować, że egzaminator A ma ogólnie wyższe współczynniki mocy różnicującej, niż to ma miejsce w województwie. Świadczy to o znacznie większym niż przeciętnie skorelowaniu wyników za poszczególne kryteria z wynikami ogólnymi uczniów. Można by zatem wysnuć wniosek, że egzaminator ten łagodniej ocenia uczniów, którzy uzyskali ogólnie wysokie wyniki, a surowiej uczniów ze słabymi wynikami.

W celu głębszej analizy tej sytuacji należałoby porównać powyższą „charakterystykę” z kartoteką arkusza oraz wykształceniem egzaminatora. Możliwe, iż okazałoby się, że egzaminator ten jest bardziej liberalny w ocenie zadań z jednego przedmiotu (np. 21, 22, 25), a jednocześnie bardziej rygorystyczny w zakresie zadań z innego przedmiotu, którego jest nauczycielem (np. 23, 24). Możliwe jednak, że egzaminator ten otrzymał skrajnie różne prace uczniów bardzo dobrych i bardzo słabych. Każde z tych stwierdzeń można przyjąć lub odrzucić jedynie po wnikliwej analizie prac uczniów i budowy arkusza egzaminacyjnego.

Analizując „charakterystykę” egzaminatora B, trudno zauważyć jakąś tendencję w ocenie prac uczniów. Można jednak podejrzewać, że egzaminator dość dowolnie stosuje ogólnie przyjęte kryteria. Zwraca uwagę dość duże ujemne odchylenie współczynnika mocy różnicującej w kryterium 25-1. Świadczy to o odwrotnym różnicowaniu uczniów, tzn. uczniowie o ogólnie dobrych wynikach otrzymywali za to kryterium statystycznie mniej punktów niż uczniowie ogólnie słabi. Analizując tę sytuację, można postawić dwie tezy. Albo egzaminator błędnie stosuje kryterium, albo przypadek sprawił, że takie niestandardowe prace otrzymał do oceny.

Analizując około 100 graficznych „charakterystyk” egzaminatorów, stwierdzono, że warto również zająć się określeniem progu odchyień, które nie są uznawane za błędy, inaczej mówiąc nad określeniem wartości odchylenia od założenia, które dałoby podstawę do uznania egzaminatora za „podejrzanego”. Ponadto nasuwają się pytania, czy jedno, dwa, a może więcej kryteriów, odbiegających od przyjętego założenia, stanowi podstawę do przyjrzenia się pracy egzaminatora. Odpowiedzi można jednak uzyskać dopiero po wnikliwej analizie większej liczby „charakterystyk”, porównaniu ich z pracami uczniów oraz licznych konsultacjach z przewodniczącymi zespołów egzaminatorów.

Przedstawione rozważania wyraźnie pokazują zarówno potrzebę weryfikacji pracy egzaminatorów, jak również dużą złożoność tego problemu. Wielość uwarunkowań wpływających na ocenę pracy egzaminatora, a zatem również na potencjalne wskaźniki tej oceny, rodzą pytanie, czy taka ocena jest w ogóle możliwa w trakcie sesji egzaminacyjnej. Być może wszystkie obecne i przyszłe rozwiązania będą mogły posłużyć jedynie do oceny egzaminatorów dopiero po zakończeniu sesji.

Na zakończenie warto jeszcze wspomnieć o problemach natury technicznej, związanych z ewentualnym wdrożeniem proponowanej weryfikacji pracy egzaminatorów. Pierwszym problemem, na jaki się natknęto, był sposób wprowadzenia do systemu przetwarzania danych egzaminacyjnych identyfikatorów egzaminatorów. Pomimo iż badania prowadzone były na populacji jednego tylko województwa, ręczne wprowadzenie identyfikatorów egzaminatorów do informatycznego systemu przetwarzania danych zajęło około 4 dni. Alternatywą byłoby skanowanie tych identyfikatorów wraz z pracami uczniów. Jednak aktualna budowa karty nie przewiduje możliwości elektronicznego skanowania identyfikatorów. Co prawda egzaminator umieszcza swój kod na karcie odpowiedzi, jednak robi to odręcznie, a poza tym często na odwrotnej stronie karty. Doświadczenie uczy, że zarówno komputerowe rozpoznawanie odręcznego

pisma, jak również jakość dwustronnego sczytywania pozostawia bardzo wiele do życzenia. Rozwiązaniem może być ręczne wpisywanie kodów egzaminatorów do systemu. Jest ono jednak czasochłonne, podczas gdy jeszcze w trakcie sesji egzaminacyjnej najbardziej zależy nam na omawianych wynikach. Wydaje się, że najlepszym rozwiązaniem byłoby wyposażenie egzaminatorów w nalepki z własnym kodem kreskowym, który umieszczaliby albo na karcie odpowiedzi (w przypadku OKE Wrocław jest to około 200 tys. nalepek), albo na protokołach pracy egzaminatorów przeznaczonych do późniejszego sczytania. Obecna kryterialna budowa zadań powoduje konieczność umieszczania wielu pól wyboru na kartach odpowiedzi. Konsekwencją tego jest budowa karty technicznie mieszczącej się na granicy wymagań producentów systemów elektronicznego przetwarzania formularzy. Umieszczenie dodatkowego pola spowodowałoby jeszcze większe zagęszczenie, co mogłoby generować większą ilość pomyłek przy nanoszeniu danych na kartę, a w skrajnych przypadkach mogłoby prowadzić do niemożności elektronicznego przetwarzania. Drugie rozwiązanie pociąga za sobą jednak konieczność drukowania protokołów dla każdego egzaminatora z dokładnością do oddziału szkolnego i późniejsze nanoszenie jedynie korekt. Powoduje to sporo problemów logistycznych związanych z przygotowaniem pracy egzaminatorów.

Bardzo poważnym problemem są jeszcze ciągle niestabilne zasady organizacji egzaminów, nie sprzyjające przygotowaniu się do akcji egzaminacyjnej pod względem technicznym. Przekazywanie do OKE kartotek testów i sposobu przeliczania wyników punktowych w dniu egzaminu, daje bardzo mało czasu na dostosowanie istniejących aplikacji do parametrów bieżącej sesji. Problemem jest również brak znormalizowanej karty odpowiedzi, co pociąga za sobą każdorazowo żmudną weryfikację poprawności szablonu do elektronicznego sczytywania oraz skracanie i tak bardzo napiętych terminów weryfikacji danych egzaminacyjnych. Wprowadzenie jednolitego szablonu karty i wcześniejsze przekazywanie do OKE kartotek testów, ułatwiłoby wypracowanie skutecznych narzędzi elektronicznej weryfikacji pracy egzaminatora. Jest to tym ważniejsze, że tego typu weryfikacja traci swoje znaczenie w momencie, gdy uczeń otrzyma już wynik swojego egzaminu.

Bibliografia

- Gołębiak B. D. (red.) (2002), *Obraz gimnazjów wrocławskich po trzech latach ich funkcjonowania*, Dolnośląska Szkoła Wyższa Edukacji Towarzystwa Wiedzy Powszechnej we Wrocławiu, Wrocław.
- Jakubowicz S. (1991), *Poprawianie klasówki – sztuka, czy rzemiosło? Studium rezultatów pewnego egzaminu z dydaktyki fizyk*, „Kwartalnik Pedagogiczny”, nr 1.