

Filip Kulon

Instytut Badań Edukacyjnych w Warszawie

Katarzyna Paczuska

Instytut Badań Edukacyjnych w Warszawie

Od nowicjusza do eksperta – rola doświadczenia w ocenianiu egzaminów zewnętrznych¹

Wstęp

Przyzwyczajiliśmy się do istnienia wokół nas tzw. ekspertów. Za eksperta może uchodzić np. mechanik samochodowy, który słuchając dźwięków wydawanych przez silnik samochodu, może zdiagnozować usterkę, lub naukowiec trafnie przewidujący wynik wyborów na podstawie prowadzonych analiz. Bycie ekspertem nie jest ograniczone do zawodów praktycznych, jak przykładowy mechanik, ani do szeroko rozumianej nauki – eksperci istnieją w każdej dziedzinie życia. Jakie cechy świadczą jednak o tym, że ktoś jest ekspertem? Ekspert posiada wiedzę i umiejętności znacznie przekraczające ich przeciętny poziom w populacji (Ericsson, 2006a). Są one jednak ograniczone do danej dziedziny. Raczej wątpliwe, by nasz mechanik zastosował skomplikowany model do dostępnych danych i otrzymał trafną predykcję dotyczącą wyniku wyborów. Oczywiście może mu się poszczęścić i intuicyjnie przewidzi ten wynik lub po prostu zgadnie. Podobnie nasz naukowiec może mieć problem z trafną diagnozą usterki, bo nawet jeśli usłyszy inny niż zazwyczaj dźwięk silnika, to nie będzie w stanie określić jego źródła i przyczyny.

Psychologowie poznawczy wskazują, że wiedza ekspertów różni się od wiedzy nie-ekspertów nie tylko ilościowo, o czym była już mowa, ale też jakościowo. Jakościowo odmienne okazują się również sposoby korzystania z niej, co łącznie prowadzi do wysokiej skuteczności działania. Właściwościami wyróżniającymi wiedzę ekspertów jest m.in. specyficzna struktura organizacji wiedzy, wysoki stopień jej proceduralizacji i bogaty repertuar schematów rozwiązywania problemów specyficznych dla danej dziedziny (Nęcka, Orzechowski i Szymura, 2013). Uporządkowanie wiedzy na różnych poziomach, jej proceduralny charakter bez utraty dostępu do wiedzy deklaratywnej, schematy rozwiązywania problemów o umiarkowanym poziomie abstrakcyjności pozwalają na sprawne i efektywne wykorzystywanie wiedzy w zależności od etapu, na którym znajduje się rozwiązywany problem. Taka struktura i właściwości wiedzy dają jednak nie tylko korzyści. Eksperti są co prawda szybsi, popełniają mniej błędów oraz wykazują się dużą elastycznością w rozwiązywaniu rutynowych problemów, jednak w literaturze przedmiotu zwraca się uwagę na ich większą sztywność w rozwiązywaniu problemów nietypowych (Chi, 2006).

¹ Tekst powstał w wyniku realizacji projektu badawczego nr 2015/17/N/HS6/02961 finansowanego ze środków Narodowego Centrum Nauki.

Eksperci mogą mieć trudność z nowymi sposobami rozwiązywania problemu z uwagi na już posiadane schematy i automatyzację działania (Kossowska, 2000). Zjawisko to nazwano „efektem eksperta”.

Nie tylko kryteria i właściwości wiedzy eksperckiej, lecz także droga jej rozwoju są od wielu lat przedmiotem dociekań badawczych. Klasyczne modele nabywania wiedzy eksperckiej zakładają, że jest to proces wieloetapowy, którego pierwsze fazy obejmują wsparcie zewnętrzne, w kolejnych etapach rozwijane są umiejętności monitorowania czynności i identyfikacji kryteriów coraz wyższego poziomu wykonywania zadań w danej dziedzinie. W dalszej kolejności ekspert zyskuje pełną kontrolę zarówno nad realizacją zadań w obrębie domeny specjalizacji, jak i kontrolę nad procesem nabywania wiedzy (Nęcka i in., 2013). O ile drogą rozwoju od nowicjusza do eksperta jest praktyka i doświadczenie, o tyle posiadanie doświadczenia niekoniecznie automatycznie oznacza posiadanie ponadprzeciętnej wiedzy i umiejętności. Badacze zwracają uwagę, że istotą stawania się i bycia ekspertem jest samoregulacja i umiejętność dalszego rozwoju. Doskonalenie, pogłębianie wiedzy i umiejętności odbywa się na drodze tzw. świadomej praktyki (*deliberate practice*), której kluczowymi właściwościami jest zorientowanie na cel – doskonalenie, w tym monitorowanie i samoocena, identyfikacja słabości i świadoma praca nad ich eliminacją (Ericsson i in., 2006b; Zimmerman, 2006).

Egzaminator w polskim systemie egzaminów zewnętrznych – założenia a rzeczywistość

Wysokiej jakości ocenianie efektów osiągnięć szkolnych to jeden z fundamentów właściwego funkcjonowania systemu egzaminów zewnętrznych w edukacji i zarazem jedno ze źródeł jego wiarygodności. Ocenianie jest czynnością angażującą złożoną aktywność umysłową, na którą składają się złożone procesy poznawcze, m.in. rozumowanie, rozwiązywanie problemów oraz podejmowanie decyzji, których podłoże stanowi wiedza specjalistyczna (Szaleniec i in., 2015). Z tych powodów system egzaminów zewnętrznych powierza to wymagające i odpowiedzialne zadanie wybranym i przygotowanym specjalistom – egzaminatorom.

Dopuszczenie do oceniania reguluje *Rozporządzenie Ministra Edukacji Narodowej z dnia 21 kwietnia 2009 r. w sprawie ramowego programu szkolenia kandydatów na egzaminatorów, sposobu prowadzenia ewidencji egzaminatorów oraz trybu wpisywania i skreślania egzaminatorów z ewidencji*. Egzaminatorem może zostać czynny zawodowo nauczyciel z co najmniej 3-letnim stażem², który ukończył szkolenie oraz uzyskał pozytywny wynik następującego po szkoleniu egzaminu. Zapisy rozporządzenia stanowią, że szkolenie dla kandydatów na egzaminatorów maturalnych w przypadku matematyki trwa 25 godzin, w tym 11 godzin poświęconych jest ocenianiu, a szkolenie z języka polskiego trwa 40 godzin, w tym, w zakresie oceniania – 20 godzin, z czego 7 godzin poświęcone ma być ocenianiu części ustnej egzaminu (Rozporządzenie, 2009). Około 85% (język polski) lub 90% (matematyka)

² Szczegółowe warunki i wymagania wobec kandydatów na egzaminatorów określa art. 9c ust. 3 lub 4 ustawy z dnia 7 września 1991 r. o systemie oświaty (Dz. U. z 2004 r. nr 256, poz. 2572, z późn. zm.).

czasu szkolenia poświęconego ocenianiu powinno odbyć się w formie ćwiczeń. W 2014 roku wprowadzono zmiany do rozporządzenia, zwiększając czas szkolenia dotyczącego oceniania z matematyki do 14 godzin, a z języka polskiego do 29 (z czego 10 godzin ma dotyczyć części ustnej egzaminu) oraz dodając zapisy o uwzględnieniu oceniania holistycznego (Rozporządzenie, 2014). Zdecydowana większość egzaminatorów została jednak wpisana do rejestru po przejściu szkoleń sprzed wprowadzonych zmian. Ramowe plany szkoleń dotyczących innych egzaminów są bardzo podobne. W przywoływanym rozporządzeniu, w załącznikach dotyczących egzaminów maturalnych z matematyki i języka polskiego (w odniesieniu do sprawdzania i oceniania odpowiedzi do zadań otwartych), podane są oczekiwane efekty szkolenia:

W wyniku realizacji bloku tematycznego uczestnik szkolenia powinien umieć:

1. opisać i stosować zasady oceniania odpowiedzi zdających do zadań otwartych według ustalonych kryteriów;
2. kwalifikować błędy popełniane przez zdających i oznaczać je;
3. stosować procedury związane ze sprawdzaniem i ocenianiem odpowiedzi zdających do zadań otwartych;
4. współpracować z innymi egzaminatorami (Rozporządzenie, 2009, s. 5537).

W trakcie roku szkolnego organizowane są szkolenia uzupełniające, głównie zdalne. Ponadto egzaminatorzy przechodzą także szkolenia każdorazowo przed przystąpieniem do sesji oceniania. Polegają one na próbnym ocenianiu arkuszy egzaminacyjnych i omawianiu otrzymanych wyników, każdorazowo dotyczą konkretnych zadań.

Można zakładać, że szkolenie dla kandydatów na egzaminatorów jest pierwszym krokiem na drodze stawania się ekspertem. Kolejne szkolenia i uczestnictwo w sesjach egzaminacyjnych powinny prowadzić do nabywania coraz bardziej rozległej wiedzy i doświadczenia, przekształcających się w wiedzę ekspercką, co powinno przekładać się na pracę coraz wyższej jakości. Okazuje się jednak, że niekoniecznie musi tak być. *Badania porównywalności oceniania i efektu egzaminatora* przeprowadzone w latach 2013–2014, w których po wstępnym szkoleniu egzaminatorzy oraz nauczyciele nie-egzaminatorzy oceniali prace maturalne z języka polskiego i matematyki, nie potwierdziły większej skuteczności egzaminatorów (Szaleniec i in., 2015). W przypadku języka polskiego nie zaobserwowano różnic w odsetku poprawnych klasyfikacji odpowiedzi uczniów pomiędzy egzaminatorami i nauczycielami. W przypadku matematyki istniały minimalne różnice w odsetkach poprawnych klasyfikacji pomiędzy egzaminatorami i nauczycielami na poziomie pojedynczych zadań, lecz na poziomie wszystkich zadań nie zaobserwowano różnicy. Nie zaobserwowano także związku pomiędzy doświadczeniem egzaminatorów, mierzonym jako liczba głównych sesji egzaminacyjnych, w których egzaminator brał udział, a odsetkiem poprawnych klasyfikacji. Większość egzaminatorów języka polskiego uczestniczyła w ośmiu lub dziewięciu sesjach, mieli oni zatem relatywnie duże doświadczenie. Egzaminatorzy matematyki byli nieco bardziej zróżnicowani pod tym względem, choć prawie jedna czwarta z nich uczestniczyła we wszystkich dziewięciu sesjach (Szaleniec i in., 2015).

Te nieco zaskakujące wyniki dają z pewnością do myślenia. Niewykluczone, że badanych egzaminatorów dotyczył problem przywoływanego „efektu eksperta” – sztywności w odniesieniu do nietypowych problemów mogącej powodować błędy w ocenianiu. Możemy także dopuścić sytuację, w której nauczyciele poprzez swoją praktykę szkolną nabywają kompetencji w zakresie oceniania na poziomie porównywalnym z poziomem specjalistów. Być może to system przygotowania i wsparcia egzaminatorów jest niefunkcjonalny – wymiar szkoleń dla egzaminatorów jest niewystarczający lub ich zakres nietrafny, co powoduje brak przełożenia na kompetencje i wysoką jakość pracy. Jednak nawet jeśli tak by było, jak wytłumaczyć brak różnic pomiędzy oceniającymi ze względu na doświadczenie? Czy doskonalenie i doświadczenie rzeczywiście nie mają znaczenia?

Metodologia

Odpowiedzi na powyższe pytanie mogą dostarczyć **dotatkowe analizy przeprowadzone** na danych z cytowanego badania, w którym poza procedurą wielokrotnego oceniania prac przez nauczycieli nie-egzaminatorów oraz egzaminatorów przeprowadzono procedurę wielokrotnego oceniania prac z udziałem koordynatorów oceniania na poziomie okręgowych komisji egzaminacyjnych (OKE). Koordynatorzy oceniania to egzaminatorzy, którzy nie tylko mają wieloletnie doświadczenie w ocenianiu prac, lecz rokrocznie odpowiadają za uzgadnianie schematu i kryteriów oceniania na poziomie krajowym, prowadzą szkolenia dla przewodniczących zespołów egzaminatorów i odpowiadają za organizację sesji oceniania (Szalencic i in., 2015), w tym ustalają sposoby postępowania w przypadku wątpliwości, których nie rozwiązano na niższym poziomie koordynacji podczas sesji oceniania. W opisywanym badaniu po trzech koordynatorów niezależnie oceniało po 121 prac z języka polskiego i matematyki, a na późniejszym etapie uzgadniali oni oceny podczas panelu. Wyniki uzyskane w procesie oceniania przez koordynatorów możemy porównać z wynikami oceniania przez egzaminatorów i nauczycieli. Powinniśmy w ten sposób rozstrzygnąć, czy istnieją różnice w jakości oceniania przez nauczycieli, egzaminatorów i koordynatorów i czy można którąś z tych grup uznać za ekspertów w ocenianiu.

Do porównania jakości oceniania przez poszczególne grupy oceniających, odmiennie niż w przywoływanych badaniach, użyto współczynników korelacji wewnątrzklasowej (*intraclass correlation coefficient*, ICC; Shrout i Fleiss, 1979). Jest to współczynnik oparty na analizie wariancji, często używany w celu obliczenia zgodności lub spójności wielokrotnych ocen. Jego wartość równa 0 oznacza całkowity brak zgodności lub spójności pomiędzy ocenami, a wartość 1 – pełną zgodność lub spójność. Zgodność następuje wtedy, kiedy oceniający przyznają dokładnie tyle samo punktów tej samej pracy. Spójność z kolei dotyczy sytuacji, kiedy przyznana punktacja nie jest identyczna, ale następuje jej systematyczne przesunięcie. Spójność pozwala na ignorowanie łagodności lub surowości oceniającego, co jest bardzo pożądane, choć można oczekiwać, że eksperci będą przyznawali nie tylko spójne, ale też zgodne oceny. Analizowane dane dotyczą dwóch sytuacji. Pierwsza ma miejsce, kiedy każda praca jest oceniana przez inny zestaw nauczycieli lub egzaminatorów wybranych losowo z populacji. Druga natomiast występuje, kiedy każdy z koordynatorów, wybrany losowo z populacji, ocenia każdą pracę. Nasze zainteresowanie

skupia się na spójności ocen pojedynczych oceniających, zatem odpowiednim współczynnikiem dla nauczycieli i egzaminatorów jest ICC(1, 1), a dla koordynatorów ICC(2,1) (Shrout i Fleiss, 1979).

W celu osiągnięcia jak największej porównywalności analiz wskazane jest przeprowadzanie obliczeń **na tych samych danych – pracach z tego samego roku** oraz tematu (język polski) lub poziomu (matematyka) egzaminu. Najbardziej ograniczone są dane oceniane przez koordynatorów, zatem pozostałe dane należałoby ograniczyć do tego samego zestawu prac. Niestety ze względu na złożoność schematu badawczego nauczycielom i koordynatorom oceniania nie przydzielono tych samych prac. W puli wspólnych prac z języka polskiego znalazło się ich zaledwie 31. Do prezentowanych analiz zdecydowano więc o ograniczeniu zakresu prac jedynie do tego samego roku, co dało 123 prace oceniane przez czterech nauczycieli każda (ograniczenie do tego samego tematu, którzy oceniali koordynatorzy dawało tylko 82 prace). Z matematyki w puli wspólnych prac ocenianych przez nauczycieli i koordynatorów było zaledwie 15. Przyjęto więc takie same reguły jak dla języka polskiego, otrzymując 182 prace oceniane przez czterech nauczycieli każda (ograniczenie do poziomu rozszerzonego dawało 58 prac). W przypadku języka polskiego wszystkie prace oceniane przez koordynatorów zostały ocenione przez komplet ośmiu egzaminatorów, zatem wspólnych prac było 121. W wyniku błędu koordynatorom oceniającym zadania z matematyki przydzielono część prac nieocenianych przez egzaminatorów, przez co wspólna liczba prac wyniosła ostatecznie 108 (przy czym jeden z koordynatorów nie ocenił indywidualnie jednej z prac, zatem niemożliwe było użycie jej do obliczenia współczynnika korelacji wewnątrzklasowej). Liczebność prac i oceniających wraz z uzyskanymi wartościami współczynników korelacji wewnątrzklasowej i ich 95% przedziałami ufności dla języka polskiego przedstawiono w tabeli 1, a dla matematyki – w tabeli 2.

Choć możliwe jest analizowanie spójności na poziomie pojedynczych zadań, zdecydowano się na analizę sumarycznych wyników prac. Jest to podejście liberalne, które uwzględnia ewentualne znoszenie się błędów oceniania pojedynczych zadań. Było to możliwe, ponieważ w badaniu cała praca była oceniana przez jedną osobę i nie wprowadzono specjalizacji w ocenianiu poszczególnych zadań. Zasadne wydaje się analizowanie wyników sumarycznych, gdyż ogólne szkolenie kandydatów na egzaminatorów powinno zapewniać spójność na tym poziomie. Niespójność na poziomie zadań mogłaby wynikać ze szkolenia bezpośrednio przed ocenianiem, w odniesieniu do konkretnego arkusza i zadań.

Tabela 1. Wartości współczynników korelacji wewnątrzklasowej dla prac z języka polskiego (wraz z 95% przedziałami ufności)

	Liczba prac	Liczba ocen każdej pracy	Dolna granica przedziału ufności ICC	Korelacja wewnątrzklasowa (ICC)	Górna granica przedziału ufności ICC
Koordynatorzy	121	3	0,883	0,912	0,935
Egzaminatorzy	121	8	0,728	0,778	0,824
Nauczyciele	123	4	0,757	0,808	0,852

Tabela 2. Wartości współczynników korelacji wewnątrzklasowej dla prac z matematyki (wraz z 95% przedziałami ufności)

	Liczba prac	Liczba ocen każdej pracy	Dolna granica przedziału ufności ICC	Korelacja wewnątrzklasowa (ICC)	Górna granica przedziału ufności ICC
Koordynatorzy	107	3	0,984	0,988	0,992
Egzaminatorzy	108	4	0,960	0,970	0,978
Nauczyciele	182	4	0,983	0,986	0,989

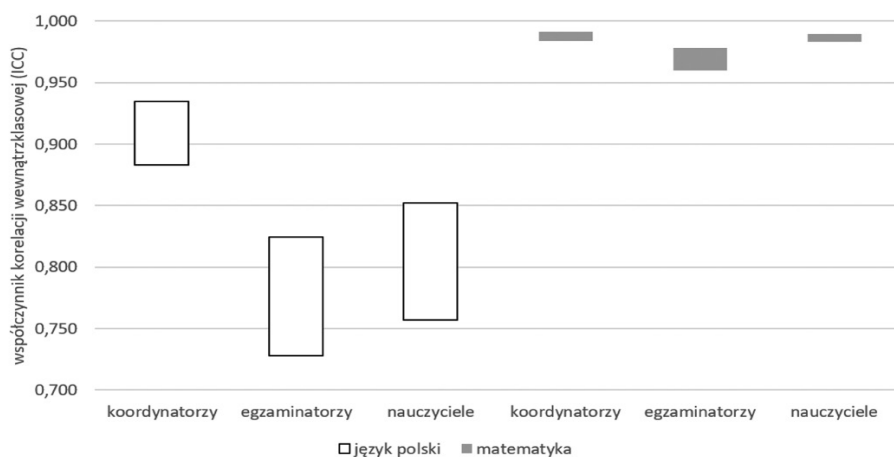
Spójność oceniania w różnych grupach oceniających

W pierwszej kolejności sprawdzono, czy pomiędzy nauczycielami i egzaminatorami rzeczywiście nie ma różnic w jakości oceniania. W tym celu obliczono odpowiednie współczynniki korelacji wewnątrzklasowej dla obydwu grup. W przypadku języka polskiego potwierdzono wnioski z przytoczonych badań – 95% przedziały ufności współczynników korelacji wewnątrzklasowej częściowo się pokrywały (por. tab. 1).

W przypadku matematyki 95% przedziały ufności współczynników korelacji wewnątrzklasowej okazały się rozłączne, co nie potwierdza wniosku o braku różnic w jakości oceniania przez nauczycieli i egzaminatorów. Warto zwrócić uwagę na to, że w przytaczanych badaniach uzyskane wyniki nie były jednoznaczne, choć różnice w odsetkach poprawnych klasyfikacji były minimalne. Również w przypadku współczynników korelacji wewnątrzklasowej różnice okazały się znikome (por. tab. 2). Ciekawe jest jednak to, że wyższy wynik dla matematyki (a więc większą spójność) uzyskali nauczyciele niebędący egzaminatorami, podobnie jak w pierwotnych badaniach. Potencjalnym wytłumaczeniem mniejszej spójności egzaminatorów mógłby być „efekt eksperta” i powiązana z nim sztywność prowadząca do błędów w sytuacjach nietypowych. Nauczyciele niebędący egzaminatorami nie wykazywaliby tego efektu. Z uwagi na brak doświadczenia nauczyciele w przypadku niepewności co do prawidłowości przyznania oceny mogli także częściej niż egzaminatorzy konsultować się z przewodniczącym zespołu w celu ustalenia wyniku. Potwierdzenie tych hipotez wymagałoby jednak odrębnych badań, skoncentrowanych na tym aspekcie.

Następnym etapem analizy było obliczenie współczynników korelacji wewnątrzklasowej dla koordynatorów. Okazały się one wyższe niż współczynniki dla egzaminatorów zarówno dla języka polskiego (por. tab. 1), jak i dla matematyki (por. tab. 2), a ich 95% przedziały ufności były rozłączne. Przedziały ufności współczynników korelacji wewnątrzklasowej dla wszystkich grup oceniających dla języka polskiego i matematyki przedstawiono graficznie na rysunku 1.

Z uwagi na dysproporcje liczebności koordynatorów i egzaminatorów oceniających każdą pracę postanowiono zweryfikować otrzymane wyniki, stosując procedurę próbkowania (*subsampling*) podobną do metody *bootstrap* (Politis i Romano, 1994). W przypadku istnienia odstających egzaminatorów, zwiększających wariację ocen pracy, współczynniki korelacji wewnątrzklasowej oszacowane dla wszystkich egzaminatorów jednocześnie mogłyby być zaniżone (zwłaszcza dla języka polskiego).



Rysunek 1. Przedziały ufności współczynników korelacji wewnątrzklasowej dla trzech grup oceniających (język polski i matematyka)

Zastosowana procedura polegała na wylosowaniu 100 tys. razy po trzech egzaminatorów z wszystkich dostępnych (ośmiu dla języka polskiego i czterech dla matematyki) i obliczeniu dla nich współczynników korelacji wewnątrzklasowej. Następnie uśredniono oszacowania punktowe współczynników, a 95% przedziały ufności wyznaczono jako 0,025 i 0,975 kwantyle rozkładu oszacowań punktowych. Otrzymane w ten sposób wartości punktowe nie różnią się od tych oszacowanych dla wszystkich egzaminatorów jednocześnie, a przedziały ufności są nieco węższe (0,743–0,811 dla języka polskiego i 0,966–0,975 dla matematyki). Nie powoduje to jednak odmiennej interpretacji niż podana wyżej – rozłączność wskazanych wcześniej przedziałów ufności pozostaje zachowana.

Koordinatorzy oceniania w zakresie języka polskiego, w porównaniu z egzaminatorami, zakwalifikowali więcej prac jako zawierające błędy kardynalne. Błąd kardynalny to poważny błąd rzeczowy, który powoduje przyznanie 0 punktów za wszystkie kryteria oceny pracy, bez względu na pozostałą wartość wypracowania. Nie musi on być związany bezpośrednio z tematem wypracowania czy prowadzoną w nim narracją. W puli 121 prac ocenianych przez koordynatorów, aż 15 z nich uznali oni ostatecznie za zawierające błąd kardynalny. W przypadku egzaminatorów jedynie w przypadku trzech z tych prac któryś z egzaminatorów stwierdził taki błąd. W całych *Badaniach porównywalności oceniania i efektu egzaminatora* egzaminatorzy uznali, że praca zawiera błąd kardynalny zaledwie w 1,1% przypadków, a koordynatorzy aż w 12% przypadków (Szalenciec i in., 2015). Gdyby wszyscy koordynatorzy uznali te same prace za zawierające błędy kardynalne i tym samym przypisali im 0 punktów, spowodowałoby to podwyższenie ich spójności. Obliczone współczynniki korelacji wewnątrzklasowej nie są jednak z tego powodu zawyżone, gdyż w żadnym przypadku wszyscy koordynatorzy oceniania indywidualnie nie przypisali pracy 0 punktów ze względu na błąd kardynalny. Jeden z koordynatorów okazał się bardzo surowy i uznał aż 16 prac za zawierające błędy kardynalne, z czego 14 zostało uznane

za takie podczas panelu. Pozostali koordynatorzy nie byli tak surowi – jeden z nich uznał pięć prac za zawierające taki błąd (cztery wspólne z surowym koordynatorem), a drugi zaledwie jedną (wspólną z surowym koordynatorem).

Podsumowanie

Przeprowadzone analizy zasadniczo potwierdziły brak różnicowania w jakości oceniania pomiędzy nauczycielami nie-egzaminatorami a egzaminatorami. Wątpliwości co do eksperckiego charakteru oceniania przez egzaminatorów mogą być dwojakiego rodzaju. Z jednej strony można mieć wątpliwości, czy szkolenie wstępne, szkolenia w trakcie pracy, a także doświadczenie są wystarczające, aby uznać ich za ekspertów. Tego typu wątpliwości wskazywali w początkowym okresie przeprowadzania zewnętrznego egzaminu maturalnego pracownicy OKE na podstawie badań ankietowych egzaminatorów (Bulska-Leśniak i Kozak, 2007). Z drugiej strony, jeśli egzaminatorzy są ekspertami, to możemy się zastanawiać nad tym, czy „efekt eksperta” nie powoduje spadku jakości oceniania.

Patrząc na przedstawione wyżej wyniki, łatwo można zauważyć znaczne różnice pomiędzy spójnością oceniania w języku polskim i matematyce (por. rys. 1). Egzaminacje te zdecydowanie się różnią i z pewnością ocenianie wypowiedzi pisemnej z języka ojczystego jest zadaniem bardzo złożonym i szczególnie podatnym na błędy (Szalenciec i in., 2015). Najprawdopodobniej jest to przyczyną niższych wartości i dużo szerszych przedziałów ufności współczynników korelacji wewnątrzklasowej w przypadku języka polskiego niż matematyki. Bardzo interesująca jest odmienna spójność oceniania przez nauczycieli w zależności od analizowanego przedmiotu. Spójność niższa od koordynatorów w przypadku języka polskiego wydaje się zrozumiała, zwłaszcza w kontekście złożoności procesów poznawczych towarzyszących ocenianiu długiej wypowiedzi pisemnej. Zastanawiająca jest jednak wyższa spójność u nauczycieli niż u egzaminatorów w przypadku matematyki (na podobnym poziomie jak u koordynatorów oceniania) oraz niewielkie, choć istotne statystycznie różnice pomiędzy tymi trzema grupami (zwłaszcza w porównaniu z językiem polskim).

Niezależnie jednak od specyfiki egzaminu, koordynatorzy oceniania z okręgowych komisji egzaminacyjnych oceniali prace bardziej spójnie niż egzaminatorzy. Takie wyniki sugerują, że być może nie sama długość doświadczenia, lecz jego różnorodność ma większe znaczenie. Zakres obowiązków koordynatorów obejmuje wiele zadań związanych z ocenianiem, których egzaminatorzy nie muszą wykonywać. Należą do nich między innymi wypracowanie kryteriów ocen, rozstrzyganie wątpliwości podczas oceniania, prowadzenie szkoleń. Niewykluczone, że to poszerza i wzmacnia struktury wiedzy i ma przełożenie na jakość samego oceniania. Niewykluczone także, że zadania, które wykonują koordynatorzy oceniania, poprzez swoją złożoność sprzyjają świadomej praktyce.

Niewątpliwie można zostać ekspertem w ocenianiu, o czym świadczą wyniki uzyskane w grupie koordynatorów oceniania. Porównywalna (bądź większa w przypadku matematyki) spójność oceniania między nauczycielami nie-egzaminatorami a egzaminatorami z doświadczeniem, skłania ku zastanowieniu nad koniecznością zmian w systemie nadawania uprawnień i szkolenia egzaminatorów egzaminu maturalnego.

Bibliografia

- Bulska-Leśniak, T., Kozak, B. (2007). Pytania o ocenianie, czyli egzamin po egzaminie. W: B. Niemiernko i M.K. Szmigel (red.). *Uczenie się i egzamin w oczach uczniów* (s. 561–566), Kraków: GRUPA TOMAMI.
- Chi, M. (2006). Two approaches to the study of experts' characteristics. W: K.A. Ericsson, N. Chairness, P.J. Feltovich i R.R. Hoffman (red.). *The Cambridge Handbook of Expertise and Expert Performance* (s. 21–31). New York: Cambridge University Press.
- Ericsson, K. A. (2006a). An introduction to Cambridge Handbook of Expertise and Expert Performance: Its development, organization, and content. W: K.A. Ericsson, N. Chairness, P.J. Feltovich i R.R. Hoffman (red.). *The Cambridge Handbook of Expertise and Expert Performance* (s. 3–19). New York: Cambridge University Press.
- Ericsson, K.A. (2006b). The influence of experience and deliberate practice in the development of the superior expert performance. W: K.A. Ericsson, N. Chairness, P.J. Feltovich i R.R. Hoffman (red.). *The Cambridge Handbook of Expertise and Expert Performance* (s. 683–704). New York: Cambridge University Press.
- Kossowska, M. (2000). Sztywność a sposób funkcjonowania poznawczego. *Przegląd Psychologiczny*, 43(2), 217–240.
- Nęcka, E., Orzechowski, J. i Szymura, B. (2013) *Psychologia poznawcza*. Warszawa: Wydawnictwo Naukowe PWN.
- Politis, D.N. i Romano, J.P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, 22(4), 2031–2050.
- Rozporządzenie Ministra Edukacji Narodowej z dnia 21 kwietnia 2009 r. w sprawie ramowego programu szkolenia kandydatów na egzaminatorów, sposobu prowadzenia ewidencji egzaminatorów oraz trybu wpisywania i skreślania egzaminatorów z ewidencji, Dz.U. z 2009 r. nr 70, poz. 600.
- Rozporządzenie Ministra Edukacji Narodowej z dnia 28 kwietnia 2014 r. zmieniające rozporządzenie w sprawie ramowego programu szkolenia kandydatów na egzaminatorów, sposobu prowadzenia ewidencji egzaminatorów oraz trybu wpisywania i skreślania egzaminatorów z ewidencji, Dz.U. z 2014 r. nr 0, poz. 675.
- Shrout, P.E. i Fleiss, J.L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, 86(2), 420–428.
- Szaleniec, H., Kondratek, B., Kulon, F., Pokropek, A., Skórska, P., Świst, K., Wołodźko, T. i Żółtak, M. (2015). *Efekt egzaminatora w ocenianiu prac maturalnych z języka polskiego i matematyki*. Warszawa: Instytut Badań Edukacyjnych.
- Zimmerman, B.J. (2006). Development and adaptation of expertise: The role of self-regulatory processes and beliefs. W: K.A. Ericsson, N. Chairness, P.J. Feltovich i R.R. Hoffman (red.). *The Cambridge Handbook of Expertise and Expert Performance* (s. 705–723). New York: Cambridge University Press.