

Paulina Skórska

Maciej Koniewski

Instytut Socjologii UJ

Przemysław Majkut

Okręgowa Komisja Egzaminacyjna w Krakowie

Zespół EWD

Obiektywność oceny stopnia spełnienia wymagań stawianych przez państwo szkołom w ramach Systemu Ewaluacji Oświaty

Prezentowana praca jest poświęcona rozważeniu statusu, jaki mają oceny stopnia spełnienia wymagań, tzw. „liter” przyznawane szkołom w ramach Systemu Ewaluacji Oświaty (SEO). W ramach artykułu zostanie rozważone pytanie, czy przyznanie „literom” funkcji obiektywnej informacji o ocenie stopnia spełnienia wymagań jest zasadne. W tym celu przeprowadzona zostanie analiza teoretyczna statusu „liter” jako wskaźników w odniesieniu do założeń metodologicznych koncepcji „wskaźnika” w naukach społecznych i ewaluacji. Zostanie także przeprowadzone porównanie ocen stawianych przez wizytatorów ds. ewaluacji szkołom w ramach obszaru efekty z innymi wskaźnikami pozwalającymi ocenić osiągnięcia szkolne uczniów.

Ewaluacja zewnętrzna w polskim systemie szkolnictwa

Ewaluacja w polskim systemie oświaty jest obecnie rozwijana w ramach Systemu Ewaluacji Oświaty. Został on wprowadzony Rozporządzeniem Ministra Edukacji Narodowej z dnia 7 października 2009 r. w sprawie nadzoru pedagogicznego (Dz. U. z dnia 9 października 2009 r.). Na mocy Rozporządzenia ewaluacja ma być jedną z form nadzoru pedagogicznego - obok kontroli i wspomaganie.

System Ewaluacji Oświaty opiera się na dwóch filarach. Ewaluacja wewnętrzna (autoewaluacja) stawia sobie za główny cel doskonalenie jakości pracy szkoły, co ma służyć wsparciu rozwoju szkół i zwiększaniu jakości edukacji. Za jej prowadzenie w szkołach odpowiedzialni są dyrektorzy. Celem ewaluacji zewnętrznej, rozwijanej w ramach SEO, jest ocena stopnia spełniania przez szkoły wymagań określonych przez państwo. Wymagania, kryteria do nich oraz sama procedura ewaluacji zostały dokładnie opisane w załączniku do przytoczonego powyżej Rozporządzenia.

Proces ewaluacji zewnętrznej w szkole jest prowadzony przez kuratoria oświaty. Osoba realizująca ewaluację nazywana jest ewaluatorem lub wizytatorem (musi ona spełniać wszystkie warunki formalne opisane w Rozporządzeniu). Osoba zatrudniona jako ewaluator jest zobowiązana do ukończenia, przynajmniej raz na dwa lata, doskonalenia w zakresie ewaluacji. Po wyznaczeniu zespołu lub osoby do realizacji ewaluacji zewnętrznej szkoły kuratorium

zawiadania dyrektora szkoły oraz organ prowadzący szkołę o zamiarze przeprowadzenia planowej ewaluacji zewnętrznej i jej zakresie w terminie 30 dni przed planowanym rozpoczęciem ewaluacji. Następnie osoba(y) prowadząca ewaluację informuje radę pedagogiczną o harmonogramie badania ewaluacyjnego. Ewaluacja ma być prowadzona w sposób niezakłócający pracy szkoły w ciągu nie dłużej niż 5 dni roboczych. Po jej zakończeniu, w terminie do 7 dni, osoba(y) prowadząca ewaluację przedstawia wyniki i wnioski z badania oraz, w terminie do 7 dni od daty spotkania z dyrektorem i radą pedagogiczną, opracowuje raport, którego elementem są oceny poziomu spełnienia wymagań, przyznane przez ewaluatora.

Siedemnaście wymagań będących przedmiotem oceny grupuje się w czterech strategicznych obszarach:

- **efekty** - rozumiane jako wszelkie pożądane rezultaty pracy dydaktycznej, wychowawczej i opiekuńczej szkoły (lub innej działalności statutowej);
- **procesy** - rozumiane jako system następujących po sobie zmian, mających prowadzić do wytworzenia efektów;
- **środowisko** - obejmujące wszelkie aspekty współpracy szkół ze środowiskiem lokalnym, a zwłaszcza z rodzicami uczniów;
- **zarządzanie** - rozumiane jako wszystkie procesy zarządcze i kierownicze, wspierające trzy powyższe obszary.

Według Rozporządzenia każde z wymagań podlega ocenie na pięciostopniowej skali:

- **poziom A** - bardzo wysoki stopień wypełniania wymagania przez szkołę;
- **poziom B** - wysoki stopień wypełniania wymagania przez szkołę;
- **poziom C** - średni stopień wypełniania wymagania przez szkołę;
- **poziom D** - podstawowy stopień wypełniania wymagania przez szkołę;
- **poziom E** - niski stopień wypełniania wymagania przez szkołę.

Charakterystyki wypełniania poszczególnych wymagań opracowano dla poziomu B i D, tworząc listę kryteriów oceny szkoły osobno dla każdego wymagania na poziomie B i D. Spełnienie przez szkołę wymagania lepiej niż na poziomie B skutkuje przyznaniem oceny A. Spełnienie przez szkołę wymagania lepiej niż na poziomie D, ale nie tak dobrze jak na poziomie B, skutkuje przyznaniem oceny C. Spełnienie przez szkołę wymagania gorzej niż na poziomie D skutkuje przyznaniem oceny E.

Na podstawie opisanej w Rozporządzeniu procedury ewaluacji można zrekonstruować bezpośrednie rezultaty „nowego” nadzoru, oczekiwane przez ustawodawcę. Są nimi:

- zebranie informacji od różnych osób związanych ze szkołą (dyrektor, nauczyciele, uczniowie, pracownicy niepedagogiczni, przedstawiciele Jednostek Samorządu Terytorialnego),
- ich analiza - i na tej podstawie
- ocena stopnia spełnienia przez szkołę poszczególnych wymagań.

Należy zauważyć, że jedynym konkretnym działaniem wobec szkół poddanych ewaluacji zewnętrznej jest zobligowanie dyrektorów szkół, które otrzymały oceną najgorszą, tj. E, do opracowania i wdrożenia programu poprawy efektywności kształcenia lub wychowania.

Poza bezpośrednim celem Systemu Ewaluacji Oświaty, jakim jest ocena stopnia spełnienia wymagań przez szkoły, dokonana w procesie opisanym w Rozporządzeniu jako ewaluacja zewnętrzna, można sformułować dodatkowe dwa cele¹ niebezpośrednie. Pierwszym z nich jest **rozwój całego systemu oświatowego** - obejmujący nie tylko wyznaczenie możliwości rozwoju szkół, ale także monitorowanie tego rozwoju, co umożliwi opracowanie przez MEN informacji o stanie oświaty. Drugim z celów jest **rozwój systemów zarządzania oświatą** - podejmowanie decyzji i strategiczne zarządzanie oświatą na podstawie zgromadzonych danych i wyników prowadzonych analiz. Obydwa z celów niebezpośrednich czynią pewne założenia co do statusu ocen wymagań. Była możliwość wykorzystania ich w monitorowaniu systemu oświatowego oraz w systemach zarządzania oświatą, oceny dokonywane przez wizytatorów w ramach ewaluacji zewnętrznej muszą być porównywalne i obiektywne. Inaczej mówiąc, ocena stopnia spełnienia wymagań przez szkoły musi dawać nam taki sam obraz szkoły względem wymagań, stawianych szkole przez państwo niezależnie od osoby przeprowadzającej ewaluację. Tylko wtedy możliwe będzie wykorzystanie w pełni potencjału informacyjnego systematycznej ewaluacji zewnętrznej szkół.

Status ocen w ramach SEO

Za podstawowy cel nowego nadzoru pedagogicznego, można uznać odpowiedź na pytania: „Jak funkcjonuje szkoła w wymienionych w Rozporządzeniu obszarach?” oraz „W jakim stopniu (na jakim poziomie) szkoła spełnia wymagania opisane w Rozporządzeniu?”. Odpowiedź na te pytania ma się odbyć w toku realizacji procedury opisanej w Rozporządzeniu (streszczonej wcześniej) i dzięki procedurom, narzędziom, zasobom przygotowanym przez podmioty obecnie realizujące „Program wzmocnienia efektywności systemu nadzoru pedagogicznego i oceny jakości pracy szkoły etap III”, czyli Ośrodek Rozwoju Edukacji (ORE), Instytut Spraw Publicznych Uniwersytetu Jagiellońskiego (ISP UJ) i firmę Era Ewaluacji Sp. z o.o. Przez procedury rozumiemy konkretne działania i zalecenia dla ewaluatorów (jak zbierać dane, jak analizować dane, jak pisać raport, jak przyznawać oceny - „litery”); narzędzia badawcze (pytania powiązane z poszczególnymi kryteriami, zapisane w kwestionariuszach ankiet, scenariuszach wywiadów indywidualnych i grupowych wykorzystywanych w ewaluacji zewnętrznej szkół); zasoby (wyszkolenie ewaluatorów w zakresie prowadzenia ewaluacji).

Zgodnie z zapisami Rozporządzenia można stwierdzić, że ustawodawca chce traktować oceny („litery”) jako wskaźnik poziomu wypełniania przez szkoły wymagań, jakie stawia im państwo. W szczególności zaś jako narzędzie do identyfikacji szkół nierealizujących wymagań co najmniej na podstawowym poziomie.

Należy w tym miejscu krótko rozważyć definicję wskaźnika w naukach społecznych i w analizach polityk publicznych. **Wskaźnik** to mierzalna zmienna,

¹ Mazurkiewicz G., Berdzik J., *System Ewaluacji Oświaty: model i procedura ewaluacji zewnętrznej* [w:] Mazurkiewicz G. (red.), *Ewaluacja w nadzorze pedagogicznym. Odpowiedzialność*, Kraków 2010.

która niejako „zastępuje” coś, co jest trudniej mierzalne². OECD definiuje³ wskaźnik jako czynnik ilościowy, jakościowy lub zmienną, za pomocą której można zmierzyć osiągnięcie, odzwierciedlić zmiany dokonane przez interwencję lub wspomóc ocenę efektywności rozwoju. Inna definicja opisuje⁴ wskaźnik jako zmienną poddającą się pomiarowi i **pozostającą w związku** z interesującymi nas cechami analizowanych obiektów czy stanów rzeczy, przez co wykorzystuje się je do stwierdzenia, **czy dany stan rzeczy zachodzi** lub, gdy mamy do czynienia z charakterystykami dającymi się ująć ilościowo, **jaki jest jego rozmiar**. Powyżej przytoczone definicje akcentują związek wskaźnika z interesującymi badacza aspektami obiektów (zwykle trudno mierzalnymi). W przypadku realizacji polityk publicznych i wdrażania interwencji współfinansowanych ze środków unijnych szczególnie nacisk jest położony na **związek wskaźnika z celami** wdrażanej interwencji. W tym ujęciu wskaźnik służy określeniu, **czy oraz w jakim stopniu** założone cele projektu / programu / interwencji zostały zrealizowane. Dokonując takiej oceny, musimy oczywiście określić, czy zostały osiągnięte produkty interwencji (określone w celach operacyjnych). Przykładem wskaźnika produktu może być liczba szkoleń kierowanych do osób, które mają pełnić funkcję ewaluatorów. Oczywiście zwykle bardziej interesują nas wskaźniki rezultatu, odzwierciedlające zmiany, jakie projekt przynosi bezpośrednim beneficjentom, przykładowo wzrost wskaźnika EWD szkoły, która wdrożyła rekomendacje z przeprowadzonej ewaluacji zewnętrznej. Dopelnieniem oceny skutków wdrażanej interwencji powinna być analiza wskaźników oddziaływania, oceniających efekty interwencji w dłuższej perspektywie. Posługując się naszym przykładem, takim wskaźnikiem mogłyby być wzrost liczby ewaluacji wewnętrznych, które były stymulowane przeprowadzeniem ewaluacji zewnętrznych w szkołach⁵.

Należy zatem rozważyć, w jakim stopniu możemy uznać oceny stopnia spełnienia wymagań jako wskaźniki, w myśl definicji obowiązujących w naukach społecznych, tj. mierzalne zjawiska, stany rzeczy, zdarzenia, zachowania, fakty, procesy, które stanowią przesłankę, że w określonych przypadkach z pewnością lub z pewnym prawdopodobieństwem dane zdarzenie, czy własność wystąpiła⁶. Wydaje się, że kluczowe jest tutaj rozważenie dwóch aspektów wskaźnika, trafności i rzetelności. Ocenę tego, czy „litera” są wskaźnikami, zależy od tego, czy:

- wymagania, kryteria, wskaźniki dla poszczególnych kryteriów i narzędzia do ich pomiaru są zdefiniowane w sposób jednoznaczny oraz że rozumiane są tak samo przez ich twórców, ewaluatorów i „dostawców” informacji, tj. respondentów (trafność wskaźnika);

² Jupp, V. (red.) *The SAGE dictionary of social research methods*. London: SAGE Publications, 2006.

³ Dostępna: http://www.funduszestrukturalne.gov.pl/PFS_Wzorce/Wzorce/PFS_Podstawowy.aspx?NRMO-DE=Published&NRNODEGUID=%7bA988C8C1-99BD-4000-A95D-137973871A63%7d&NRORIGINALURL=%2fewaluacja_2004_2006%2fslowniczek%2f&NRCACHEHINT=NoModifyGuest#w

⁴ Górniak J., Keler K., *Wskaźniki w ewaluacji ex-post programów publicznych* [w:] Haber A. (red.), *Ewaluacja ex-post. Teoria i praktyka badawcza*, PARP, Warszawa 2007.

⁵ Olejniczak K., *Mechanizmy wykorzystania ewaluacji*, Scholar, Warszawa 2008.

⁶ Nowak S., *Metodologia badań społecznych*, Wydawnictwo Naukowe PWN Warszawa, 2007.

- procedury, w wyniku których szkołom nadawane są „litery”, są opracowane w taki sposób, aby niezależnie od indywidualnych różnic między ewaluatorami, sposobu prowadzenia badania ewaluacyjnego (gromadzenia danych), obróbki i analizy danych, wpływu czynników zewnętrznych, np. w postaci nacisków dyrektora, czy rady pedagogicznej, przyznana szkole „litera” była taka sama, gdyby nadawana była przez różnych ewaluatorów, w różnych okolicznościach, ale dla tej samej szkoły (rzetelność wskaźnika).

Analiza

W celu określenia tego, w jakim stopniu oceny stawiane szkołom w ramach ewaluacji zewnętrznej są wskaźnikami stopnia spełnienia wymagań przez państwo, można wykorzystać materiał zgromadzony w ramach przeprowadzonych ewaluacji. Porównanie ocen i komentarzy ewaluatorów może dać nam odpowiedź na pytanie, w jakim stopniu można traktować „litery” jako wskaźniki.

W celu odpowiedzi na postawione powyżej pytanie badawcze wylosowano próbę badawczą gimnazjów. Operat do losowania stanowiły wszystkie gimnazja publiczne, w których zrealizowano ewaluację zewnętrzną w terminie od 01.01.2011 do 31.07.2012 i które wypełniają wymaganie „Analizuje się wyniki egzaminu gimnazjalnego” w obszarze „Efekty działalności dydaktycznej, wychowawczej i opiekuńczej oraz innej działalności” na poziomie A (Grupa A) oraz D lub E (Grupa D+E). Operat do losowania pobrano dnia 01.08.2012 z platformy seo2.npseo.pl. W każdej z grup wylosowano, losowaniem prostym bez zwracania, po 15 szkół wraz z komentarzem do wymagania. Łącznie więc szczegółowej analizie poddano 30 szkół i komentarzy.

Analiza komentarzy polegała na ocenie argumentów, stawianych przez ewaluatorów na potwierdzenie swoich ocen, wraz z przykładami opartymi na wypowiedziach respondentów (nauczycieli, dyrektorów, partnerów z samorządu). Lista argumentów została opracowana na podstawie kryteriów do analizowanego wymagania oraz rekonstrukcji pytań postawionych respondentów. Lista argumentów znajduje się w załączniku (Załącznik 1).

Poniżej prezentujemy wyniki krótkiej analizy, dotyczącej wszystkich gimnazjów publicznych, poddanych ewaluacji w okresie od 01.01.2011 do 31.07.2012; zarówno tych w operacie, z którego losowaliśmy przypadki do bardziej szczegółowych analiz (szkoły z A i D+E), jak i pozostałe szkoły z B i C na analizowanym wymaganiu. Wyniki tej analizy odpowiadają na pytania: Ile szkół wypełnia analizowane wymaganie na określonym poziomie?, Jaka jest proporcja nadawanych ocen?, Jakie jest zróżnicowanie w proporcji poszczególnych ocen?, Jakie występują różnice między województwami?

W tabeli poniżej zaprezentowano liczbę poszczególnych „liter” przyznanych, objętym analizą szkołom, w wartościach absolutnych, jak i procentowych. Intensywność koloru wskazuje, jak bardzo dane województwo odchyła się od średniej wartości dla Polski. Na przykład województwami, w których przyznano stosunkowo najwięcej ocen A dla analizowanych danych, są kujawsko-pomorskie i podkarpackie; najmniej zaś w pomorskim, a w ogóle nie przyznano w łódzkim i zachodniopomorskim.

O zróżnicowaniu między województwami mówi wartość wariancji. Wariancja dla E jest zerowa, ponieważ ewaluatorzy bardzo niechętnie wystawiali taką ocenę we wszystkich województwach. Województwa są najbardziej zróżnicowane, jeśli chodzi o liczbę szkół, które otrzymały B.

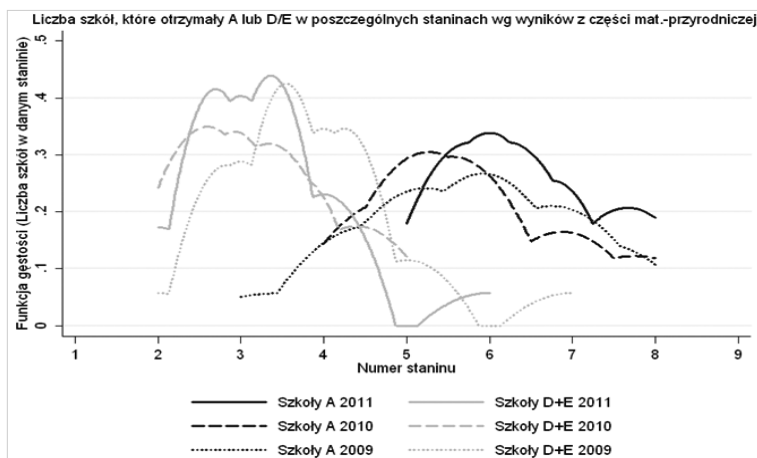
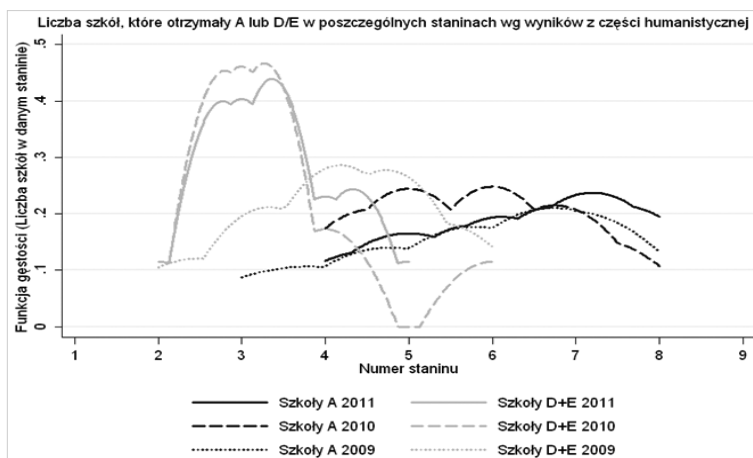
Podsumowanie ocen spełnienia wymagań wystawionych w ramach analizowanego kryterium we wszystkich szkołach w Polsce poddanych ewaluacji w badanym okresie

	A	B	C	D	E	Σ	A	B	C	D	E	Wskaźnik jakości
dolnośląskie	4	12	15	5	0	36	11%	33%	42%	14%	0%	0,3
kujawsko-pomorskie	8	14	8	3	0	33	24%	42%	24%	9%	0%	0,6
lubelskie	1	9	8	1	0	19	5%	47%	42%	5%	0%	0,5
lubuskie	1	11	7	1	0	20	5%	55%	35%	5%	0%	0,6
łódzkie	0	26	5	1	0	32	0%	81%	16%	3%	0%	0,8
małopolskie	1	12	5	0	0	18	6%	67%	28%	0%	0%	0,7
mazowieckie	6	19	14	3	2	44	14%	43%	32%	7%	5%	0,5
opolskie	1	3	1	0	0	5	20%	60%	20%	0%	0%	0,8
podkarpackie	11	18	13	3	0	45	24%	40%	29%	7%	0%	0,6
podlaskie	1	9	6	1	0	17	6%	53%	35%	6%	0%	0,5
pomorskie	1	11	18	1	0	31	3%	35%	58%	3%	0%	0,4
śląskie	9	24	10	1	0	44	20%	55%	23%	2%	0%	0,7
świętokrzyskie	1	7	5	2	1	16	6%	44%	31%	13%	6%	0,3
warmińsko-mazurskie	1	6	10	2	0	19	5%	32%	53%	11%	0%	0,3
wielkopolskie	3	12	12	2	0	29	10%	41%	41%	7%	0%	0,4
zachodnio-pomorskie	0	3	4	2	0	9	0%	33%	44%	22%	0%	0,1
Polska	49	196	141	28	3	417	12%	47%	34%	7%	1%	0,5
						wariancja	0,7%	1,8%	1,4%	0,3%	0,0%	

Obliczono też „wskaźnik jakości”, mówiący, w których województwach szkoły lepiej wypełniają analizowane wymaganie; o ile literki są dobrym wskaźnikiem jakości pracy szkoły. Wskaźnik został obliczony jako różnica między odsetkiem szkół, które otrzymały A lub B a odsetkiem szkół, które otrzymały D lub E. Wskaźnik może przyjmować wartości od -1 do 1. Wartości dodatnie oznaczają, że w danym województwie jest przewaga szkół „lepszyc”. Na podstawie wskaźnika można powiedzieć, że najwięcej szkół wypełniających analizowane wymaganie na najlepszym poziomie jest w województwach łódzkim i opolskim; najmniej zaś w zachodnio-pomorskim.

Możemy zastanowić się, czy istnieje związek między surowymi wynikami egzaminacyjnymi analizowanych szkół i ocenami, jakie zostały im postawione podczas ewaluacji w ramach obszarów efekty. Związek ten został przedstawiony na dwóch wykresach poniżej. Prezentują one funkcję gęstości (Kernel Density Estimation) analizowanych 30 szkół w poszczególnych staninach w latach 2009-2011 ze względu na wyniki uczniów na egzaminie z części humanistycznej i matematyczno-przyrodniczej. Jak widać szkół, które otrzymały ocenę A, jest więcej w górnych staninach (zaznaczone na czarno). Analogicznie szkół, które otrzymały ocenę D lub E, jest więcej w dolnych staninach (zaznaczone

na szaro). Szkoły, którym przyznano ocenę D lub E, pogorszyły swoje wyniki zarówno z części humanistycznej, jak i matematyczno-przyrodniczej, część z nich, które w 2009 plasowały się w staniach 4-7, w latach 2010-2011 „spadło” do stanów 2-4. Natomiast szkoły ocenione na A utrzymują wyniki przez ostatnie trzy lata na podobnym poziomie. Oceny wypełniania analizowanego wymagania bardziej pokrywają się z wynikami egzaminów z części matematyczno-przyrodniczej, niż humanistycznej.



Wyniki analiz

Pierwszym wnioskiem z analizy jest to, że ilościowa informacja z komentarzy opisujących spełnienie kryterium na poziomie najwyższym (litera A) jest zbliżona do tej z komentarzy opisujących spełnienie kryterium na poziomie najniższym (litera D oraz litera E). W przypadku uzasadnień przyznania poziomu A spełnienia wymagania, średnia liczba słów wynosiła 1120, natomiast w przypadku przyznania ocen najniższych - 1178 (w tym przypadku była nawet nieco wyższa).

Jeśli jednak spojrzymy na jakość wartości informacyjnej związanej z tymi uzasadnieniami można zauważyć, że była ona znacząco różna dla uzasadnień oceny najwyższej i najniższej. Spośród 41 wyróżnionych, możliwych argumentów (wymienione w załączniku) mogących składać się na uzasadnienie oceny, na jedno uzasadnienie oceny najniższej przypada średnio 16 argumentów. W przypadku uzasadnień ocen najwyższych pojawia się więcej argumentów uzasadniających ocenę - na jeden komentarz przypada ich średnio 18. Jeszcze bardziej skrajna sytuacja pojawia się, kiedy weźmiemy pod uwagę nie tylko fakt pojawienia się (bądź nie) argumentu, ale również podanie konkretnego przykładu(ów) i/lub uzasadnienia argumentu. W przypadku komentarzy odnoszących się do przyznania ocen najniższych, jeden komentarz zawiera średnio 5 takich pogłębionych informacji. W przypadku komentarzy odnoszących się do przyznania ocen najwyższych liczba ta jest aż dwukrotnie wyższa - na jeden komentarz przypada średnio 10 informacji zawierającej oprócz argumentu, jego przykład(y) i/lub uzasadnienie. Dodatkowo, jeśli spojrzymy na informacje ambiwalentne (a więc sytuację, w której w komentarzu występuje sprzeczność, wynikająca np. z rozbieżności między deklaracjami nauczycieli i dyrektora, niekonsekwencją nauczycieli w odpowiedzi na pytania lub błędami autora raportu), to w przypadku komentarzy odnoszących się do przyznania ocen najniższych było dwukrotnie więcej niż w przypadku komentarzy odnoszących się do przyznania ocen najwyższych.

Z powyższych danych można wysnuć wniosek, że komentarze uzasadniające przyznanie najniższej oceny (D lub E) charakteryzują się znacząco niższą wartością informacyjną i są znacznie słabiej uzasadnione w porównaniu z komentarzami uzasadniającymi przyznanie najwyższej oceny spełnienia wymagania. Komentarz dotyczący przyznania niskiej oceny powinien być nasycony argumentami, ilustrującymi je przykładami (jak najbardziej skonkretyzowanymi) oraz ich uzasadnieniami. Zwłaszcza jeśli ma stanowić dla szkół wartościową informację pod kątem planowanych do wdrożenia zmian i planów naprawczych.

Kolejny problem z komentarzami uzasadniającymi oceny dotyczy tego, o co pytani są nauczyciele i dyrektorzy. Przede wszystkim sprawdza się (na podstawie ich deklaracji), czy i jakie działania oraz inicjatywy są podejmowane przez szkołę. Jedynym kryterium oceny prawdziwości takiej informacji jest sprawdzenie spójności deklaracji dyrektora i nauczycieli. Ewaluatorowi brak procedur działania w sytuacji niespójności, gdyż rozstrzygnięcie sprzeczności deklaracji nauczycieli i dyrektora wymaga bazowania na innych źródłach (np. analizie danych zastanych), co (jak wynika z niniejszej analizy) nie jest powszechną praktyką przy pisaniu komentarzy. Pytanie o to, czy prowadzona jest analiza i na jej podstawie wdrażane są wnioski, jest powierzchnym wymaganiem, gdyż nie odnosi się do skuteczności (nie mówiąc o efektywności) podejmowanych działań. Zamiast pytać nauczycieli i dyrektora o fakt prowadzenia analiz i o to, czy nauczyciele znają ich wnioski, lepiej skupić się na monitorowaniu wdrożonych działań. Brak odniesienia wprowadzanych na podstawie analiz działań do ich efektów po pierwsze nie pozwala ocenić, czy zmierzamy w odpowiednim kierunku (zblizamy się do celu). Po drugie, nie dostarcza uzasadnienia dla ponoszenia (różnorodnych: finansowych i pozafinansowych) kosztów prowadzenia analiz, gdyż nie można zweryfikować ich użyteczności.

Dodatkowo w komentarzach brakuje odniesień ewaluatora do jakości prowadzonych analiz (nie tylko do użyteczności, która jest trudniejsza w ocenie, ale nawet do poprawności i celowości wykonywanych czynności analitycznych).

Podsumowując, można stwierdzić, że głównie komentarze skupiają się na deklaracjach tego, czy podejmuje się działania (i jakie) na podstawie analizy wniosków, ale przed i po poziomie działania, występuje poważna luka informacyjna. Odnosząc się do analizy (podstawy działań), pomija się kwestię jakości prowadzonych analiz. „Luka po” odnosi się do braku odniesienia wprowadzanych działań do ich późniejszych efektów, co sprawia, że nie można trafnie ocenić zarówno użyteczności wprowadzanych działań, jak i prowadzonych analiz. Obrazuje to poniższy schemat.

Luki informacyjne w komentarzach uzasadniających oceny



Ostatnie uwagi odnoszą się do trafności komentarzy. „Literę” są w dużej mierze wystawianie na podstawie oceny tego, czy działania podjęte na podstawie analiz idą w parze z wysokimi (lub wzrastającymi) osiągnięciami uczniów. W przypadku części najniższych ocen, gdyby nie fakt niskich i/lub spadających wyników, ilość działań analitycznych jest tak rozległa, że ocena spełnienia wymagania powinna być wyższa. Większe znaczenie w przyznawaniu ocen mają zatem osiągnięcia uczniów niż samo wdrażanie wniosków oparte na systematycznym monitoringu i wynikach prowadzonych analiz. Dodatkowo, odnosząc się w ocenach do wyników uczniów, częściej bierze się pod uwagę surowe wyniki egzaminacyjne, a nie wartości wskaźnika Edukacyjnej Wartości Dodanej. Wskaźnik EWD mówi o przyroście wiedzy uczniów, czyli wkładzie szkoły w osiągi uczniów. Jest więc bardziej trafną miarą oceny efektywności działań szkół. Zwłaszcza zaś dla szkół, które pracują z mniej uzdolnionymi uczniami lub uczniami z trudnych środowisk.

Powyżej celowo użyto stwierdzenia, że ocenia się, czy wprowadzane działania idą w parze z osiągnięciami uczniów, gdyż nie ma pewności, że konkretnie analizowane działania wpływają na podnoszenie wyników uczniów (możemy mieć do czynienia wyłącznie z koincydencją, tzw. efekt deadweight, polegający na tym, że efekty byłyby osiągnięte nawet wówczas, gdyby nie wprowadzano określonego działania naprawczego). W związku z tym wysokie (podnoszące się) wyniki nie mogą być głównym kryterium oceny spełnienia wymagania (i poziomu spełnienia). Zwłaszcza, że efekty wprowadzanych działań (w tym lepsze wyniki uczniów) ujawniają się stopniowo i mają charakter odroczonego

w czasie od momentu wdrożenia interwencji. W większości komentarzy ewaluatorzy nie odnieśli się jednak wystarczająco do perspektywy czasowej, nie ujmując potencjalnych efektów wprowadzanych działań w kontekście czasu.

Podsumowanie

Ewaluacja zewnętrzna prowadzona w ramach Systemu Ewaluacji Oświaty miała w założeniu twórców dać wszystkim zainteresowanym przejrzystą informację dotyczącą stopnia spełnienia wymagań stawianych szkołom przez państwo. Przeprowadzona przez nas analiza wskazuje, że w obecnie istniejących warunkach, przy obecnie wykorzystywanych narzędziach badawczych i procedurach ewaluacji, nie można zagwarantować obiektywności ocen dawanych przez ewaluatorów w ramach obszaru „efekty”. Problem wynika głównie z oparcia ewaluacji w tym zakresie o deklaracje dyrektorów i nauczycieli, bez odniesienia się do oceny jakości i celowości działań podjętych przez szkołę. Analiza komentarzy wskazuje, że szkoły o najniższych i najwyższych ocenach nie różnią się znacząco, jeśli chodzi o zakres przeprowadzanych analiz i wdrażane na ich podstawie wnioski. Różnią się w zakresie surowych wyników egzaminacyjnych.

Obecnie należy uznać, że istnieje problem obiektywności ocen w ramach analizowanego kryterium. Wymagania, kryteria i pytania na nich oparte nie są zdefiniowane w sposób jednoznaczny i co z tym się wiąże nie zostały opracowane zestawy precyzyjnych wskaźników, z nimi powiązanych. Przy projektach zakrojonych na tak szeroką skalę konieczne jest zdefiniowanie bardzo precyzyjnych procedur i wskaźników, które umożliwiłyby obiektywność przy angażowaniu dużej liczby ewaluatorów i prowadzeniu przez nich badań ewaluacyjnych w tak różnych kontekstach. Procedury, w wyniku których szkołom nadawane są „litery”, nie dają łatwych rozwiązań ewaluatorom, pozwalającej na obiektywność oceny.

System Ewaluacji Oświaty jest ciągle nowym aspektem systemu szkolnictwa w Polsce. Nie można przecenić potencjału, jaki ma względem pozytywnych zmian w szkołach. Potrzebna jest jednak dalsza praca nad narzędziami badawczymi i procedurami w ramach SEO, aby lepiej można wykorzystywać informacje, które zbierane są w jego ramach, jak również przekuć entuzjazm osób zaangażowanych w ten program na efekty jeszcze lepiej służące polskiemu systemowi edukacji w Polsce.

Załącznik:

Argumenty, na które powołuje się ewaluator w uzasadnieniu (komentarzu)

1	potwierdzenie faktu wypełniania wymagania przez powołanie się na inne źródła niż tylko wywiady (np. dokumentacja)
2	wskazano kto dokonuje analizy (nauczyciele indywidualnie, zespoły przedmiotowe, zespoły językowe, zespoły ds. ewaluacji)
3	opisano procedurę analizy (kto analizuje, kto formułuje wnioski, jaki jest obieg informacji o wnioskach)
4	prowadzi się analizę na poziomie indywidualnym, klas, szkoły
5	porównuje się wyniki egzaminu z ocenianiem wewnętrznym

6	porównuje się wyniki egzaminu na tle gminy, województwa, kraju
7	prowadzi się analizy z wykorzystaniem średnich wyników egzaminu
8	prowadzi się analizy z wykorzystaniem skali staninowej
9	prowadzi się analizy z wykorzystaniem EWD
10	do analiz wykorzystuje się „Kalkulator EWD”
11	prowadzi się analizy w czasie
12	prowadzi się analizę kontekstową
13	analizuje się stopień opanowania konkretnych umiejętności
14	analizuje się trudność zadań egzaminacyjnych
15	analizuje się wyniki w odniesieniu do standardów wymagań egzaminacyjnych
16	przekazuje się informację zwrotną rodzicom i uczniom
17	nauczyciele znają wnioski z analiz prowadzonych w szkole
18	na podstawie analiz opracowywane są programy naprawcze
19	wprowadza się zmiany w metodach i formach pracy na lekcjach
20	doskonalą się umiejętności, które sprawiają uczniom najwięcej problemów na egzaminie
21	zwiększa się wymiar godzin przedmiotów o najgorszych wynikach
22	prowadzi się zajęcia wyrównawcze i inne pozalekcyjne
23	prowadzi się formy indywidualizacji nauczania i wsparcia ucznia
24	prowadzi się formy pracy zespołowej
25	współpracuje się z instytucjami kultury, organizuje się formy edukacji pozaszkolnej
26	modyfikuje się plany, programy nauczania, dobór podręczników na podstawie wyników analiz
27	synchronizuje się plany nauczania z różnych przedmiotów
28	prowadzi się ćwiczenia rozwiązywania testów egzaminacyjnych
29	organizuje się egzaminy próbne
30	prowadzi się diagnozę wiedzy i umiejętności uczniów „na wejściu”
31	prowadzi się zewnętrzne formy diagnozy wiedzy i kompetencji uczniów inne niż egzamin
32	monitoruje się sytuację dydaktyczną i wychowawczą w klasach
33	realizuje się szkolne projekty edukacyjne
34	prowadzi się indywidualną diagnozę potrzeb edukacyjnych ucznia
35	prowadzi się monitoring skuteczności poszczególnych działań szkoły w podnoszeniu efektywności nauczania
36	o skuteczności wdrożenia wniosków z analiz świadczy uczestnictwo uczniów w konkursach i olimpiadach
37	o skuteczności działań świadczą sukcesy absolwentów
38	o skuteczności działań świadczą poprawiające się wyniki na egzaminach
39	o skuteczności działań świadczą poprawiające się wartości EWD
40	o skuteczności działań świadczą wysokie aspiracje edukacyjne uczniów
41	o skuteczności działań świadczą przyznane szkole wyróżnienia i nagrody

Bibliografia:

1. Górniak J., Keler K., *Wskaźniki w ewaluacji ex-post programów publicznych* [w:] Haber A. (red.), *Ewaluacja ex-post. Teoria i praktyka badawcza*, PARP, Warszawa 2007.
2. Jupp V. (red.), *The SAGE dictionary of social research methods*. London: SAGE Publications, 2006.

3. Mazurkiewicz G., Berdzik J., *System Ewaluacji Oświaty: model i procedura ewaluacji zewnętrznej* [w:] Mazurkiewicz G. (red.), *Ewaluacja w nadzorze pedagogicznym. Odpowiedzialność*, Kraków 2010.
4. Nowak S., *Metodologia badań społecznych*, Wydawnictwo Naukowe PWN, Warszawa 2007.
5. Olejniczak K., *Mechanizmy wykorzystania ewaluacji*, Scholar, Warszawa 2008.
6. Rozporządzenie Ministra Edukacji Narodowej z dnia 7 października 2009 r. w sprawie nadzoru pedagogicznego (Dz. U. z dnia 9 października 2009 r.).