



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



EDUKACYJNA
WARTOŚĆ
DODANA

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



RAPORT TEMATYCZNY Z BADANIA

**Wykorzystanie przez szkoły metody
edukacyjnej wartości dodanej.
Analiza wyników badania
ilościowego 2013 i 2015**



Autor:
Piotr Zielonka

Wydawca:
Instytut Badań Edukacyjnych
ul. Górczewska 8
01-180 Warszawa
tel. (22) 241 71 00; www.ibe.edu.pl

© Copyright by: *Instytut Badań Edukacyjnych, Warszawa, październik 2015*

Publikacja opracowana w ramach projektu systemowego: *Badania dotyczące rozwoju metodologii szacowania wskaźnika edukacyjnej wartości dodanej*, współfinansowanego przez Unię Europejską ze środków Europejskiego Funduszu Społecznego, realizowanego przez Instytut Badań Edukacyjnych.

Egzemplarz bezpłatny

Spis Treści

Spis Treści	3
2. Wstęp	4
3. Metodologia badania.....	6
3.1. Dobór próby.....	6
3.2. Realizacja wywiadów	8
3.3. Błąd statystyczny.....	11
4. Opis próby	12
4.1. Reprezentatywność próby.....	13
4.2. Porównanie próby z 2013 i 2015 roku	17
4.3. Wnioski z porównania prób	18
5. Wyniki analiz.....	19
5.1. Odsetki wykorzystania EWD	19
5.2. Porównanie wskaźników wykorzystania EWD dla lat 2013 i 2015	21
5.2.1. Porównanie I	22
5.2.1. Porównanie II	23
5.2.2. Przedziały ufności z uwzględnieniem braków udziału	24
6. Wnioski	26

2. Wstęp

Kontekst i cele badania

Raport powstał jako element ewaluacji wdrażanego w latach 2013-2015 programu *Rozwój metody edukacyjnej wartości dodanej na potrzeby wzmocnienia ewaluacyjnej funkcji egzaminów zewnętrznych*. Celem programu było rozwijanie metodologii edukacyjnej wartości dodanej (dalej EWD¹) oraz upowszechnianie wykorzystania wskaźników. Głównym celem raportu jest oszacowanie poziomu wykorzystania wskaźników EWD przez gimnazja, licea i technika oraz odpowiedź na pytanie, czy wspomniany program przyczynił się do wzrostu wykorzystywania wskaźników EWD.

Wykorzystanie wskaźników EWD było już popularyzowane wcześniejszymi działaniami. Sam wspomniany program jest kontynuacją wcześniejszego projektu (*Badania dotyczące rozwoju metodologii szacowania wskaźników edukacyjnej wartości dodanej* (EWD), który był realizowany w latach 2007-2013 oraz wcześniejszych działań prowadzonych jeszcze w CKE. W ramach wcześniejszych działań udało się w znaczącej mierze spopularyzować wykorzystanie wskaźników EWD. Dlatego ocena skutków programu z lat 2013-2015 musi uwzględniać wcześniejsze działania, a zatem oprócz samego oszacowania odsetka szkół korzystających ze wskaźnika EWD pod koniec roku 2015 powinna również oszacować ten odsetek na koniec roku 2013 i porównać obie wartości.

W tym celu konieczne było wykorzystanie wyników dwóch badań ankietowych. Każde z nich zostało wykonane jako element ewaluacji wspomnianych programów:

- Pierwsze w roku 2013, jako część ewaluacji projektu z lat 2007-2013,
- Drugie w roku 2015, jako część ewaluacji projektu z lat 2013-2015.

Z punktu widzenia oceny wpływu interwencji jaką był wspomniany program pierwsze badanie stanowi pomiar stanu przed interwencją, zaś drugie stanu po interwencji. Łącznie pozwalają one zbadać zmiany w poziomie wykorzystania wskaźników EWD w latach 2013-2015.

Wcześniejsze ustalenia

Jako podsumowanie projektu z lat 2007-2013 powstał w 2014 roku *Raport z ewaluacji wewnętrznej projektu EWD*². Poświęcono w nim wiele uwagi zagadnieniu wykorzystania wskaźników EWD przez szkoły, w tym szacowaniu wskaźnika rezultatu, zwracając szczególną uwagę na złożony schemat doboru próby oraz problem braków udziału jaki napotkano przy realizacji badania. W niniejszym raporcie starano się wykorzystać zawarte tam ustalenia i kontynuować główne kierunki wytyczone wcześniej przy opisie i szacowaniu wskaźnika szkół wykorzystujących wskaźniki EWD.

¹ W raporcie nie będzie dokładniej omawiana metoda Edukacyjnej Wartości Dodanej. Informacje na jej temat można znaleźć na: www.ewd.edu.pl.

² Instytut Badań Edukacyjnych, *Raport z ewaluacji wewnętrznej projektu EWD*, Warszawa 2014, dostęp online: <http://eduentuzjasci.pl/pmbel/110-badanie/1016-ewaluacja-wewnetrzna-projektu-ewd.html>.

Przed wszystkim skorzystano z ustaleń na temat schematu doboru próby i ich konsekwencji dla szacowania wskaźników. Powtórzono również zaproponowany tam sposób operacjonalizacji wykorzystania wskaźników EWD oraz sposób poradzenia sobie ze szkołami, które zostały wylosowane do próby, lecz ostatecznie nie wzięły udziału w badaniu, w tym w szczególności z ich klasyfikacji na systematyczne i losowe.

Uwagi ogólne

W raporcie często używano pojęć „poziom nauczania” i „efektywność nauczania”. Przez „poziom nauczania” autor rozumiał bezwzględny, obiektywny poziom nauczania w szkole (czyli poziom umiejętności uczniów opuszczających szkołę), nawet jeżeli jest on w danych okolicznościach nieznany. W tym sensie wynik uzyskiwany przez uczniów szkoły na egzaminach zewnętrznych uważa się tylko za jego pomiar, nie zaś za właściwy poziom nauczania. Przez „efektywność nauczania” autor rozumiał wkład pracy dydaktycznej szkoły w uzyskany przez uczniów poziom umiejętności, czyli postęp jaki uczniowie dokonali w czasie nauki w placówce, ale tylko w zakresie wypracowanym przez szkołę. Pojęcie to, podobnie jak w przypadku poziomu nauczania rozumie się, jako raczej nieznana cecha szkoły, której wskaźniki EWD są tylko pomiarem, oszacowaniem. Autor raportu dołożył starań, aby owych pojęć używać ściśle, i rozdzielać poziom nauczania oraz efektywność nauczania od ich miar, takich jak wyniki egzaminacyjne i wskaźniki EWD. Warto zwrócić uwagę, że w tym sensie można używać pojęć „poziom nauczania” i „efektywność nauczania” również w odniesieniu do szkół, dla których ich pomiar nigdy nie został dokonany lub zwyczajnie nie jest możliwy.

W badaniu prowadzonym w 2015 roku badano cztery typy szkół: szkoły podstawowe, gimnazja, licea i technika. Spowodowane było to tym, że od 2014 roku dyrektorzy każdego z czterech rodzajów szkół mogli wykorzystywać wskaźniki EWD do ewaluacji pracy szkoły. Jednak w roku 2013, gdy realizowano pierwsze z dwóch porównywanych badań ewaluacyjnych, dyrektorzy szkół podstawowych jeszcze nie mieli możliwości wykorzystywania wskaźników EWD. Z tego powodu, pierwsze badanie nie obejmowało szkół podstawowych. Porównanie wyników badań z lat 2013 i 2015 nie będzie zatem obejmowało szkół podstawowych, a tylko gimnazja, licea i technika, gdyż tylko dla nich takie porównanie jest możliwe. Z tego powodu, w dalszej części raportu wszystkie elementy badania z 2015 roku dotyczące szkół podstawowych zostaną w raporcie pominięte, chociaż badanie to obejmowało również szkoły tego typu. Stąd o próbie zbadanej w 2015 mówi się, że liczyła ona 1252 szkoły (choć wliczając szkoły podstawowe byłoby ich więcej), na które składały się licea, technika i gimnazja (choć w istocie zbadano też szkoły podstawowe) oraz, że wszystkie typy szkół zostały zbadane tym samym kwestionariuszem (choć dla szkół podstawowych przygotowano osobny kwestionariusz).

Dla uproszczenia wywodu, w dalszej części raportu odnoszono się do dwóch badań nazywając je I falą badania oraz II falą badania. Jest to o tyle istotne, że najważniejsze pytania analizowane w raporcie również zadawano odnośnie różnych lat. Sprawia to, że mówienie w jednym zdaniu o różnych odpowiedziach w różnych latach badania na pytania dotyczące różnych lat jest bardzo kłopotliwe. Przyjęta konwencja poparwia przejrzystość raportu. I falą badania nazywane będzie badanie z roku 2013, zaś II – badanie z roku 2015.

3. Metodologia badania

3.1. Dobór próby

Głównym postulatem, jaki zrealizowano przy doborze próby, było jednocześnie zapewnienie reprezentatywności próby i zadowalającej dokładności oszacowań na dwóch poziomach:

1. na poziomie ogółu szkół w Polsce (tzn. na poziomie łącznej próby gimnazjów, techników i liceów),
2. na poziomie poszczególnych typów szkół (tzn. osobno dla podpróby szkół podstawowych, osobno dla podpróby gimnazjów itd.).

Za rozwiązanie optymalne uznano dobranie do próby po około 500 szkół każdego typu. Sprawilo to, że na poziomie każdego typu szkół liczebność próby pozwala na zadowalającą dokładność, a cała próba 1500 szkół zapewnia wystarczającą dokładność oszacowań na poziomie ogółu szkół ponadpodstawowych w kraju. Jednocześnie, oznacza to, że szkoły różnych typów miały różne prawdopodobieństwo znalezienia się w próbie. Wynika to z faktu, że liczebności szkół różnych typów są znacząco różne – w operacie z którego wylosowano próbę było przede wszystkim najwięcej szkół podstawowych, i znacząco mniej liceów oraz techników. Naturalnie wynika to z faktu, że licea i technika są bardziej wyspecjalizowanymi placówkami oraz przeciętnie jedna tego typu placówka skupia więcej uczniów niż gimnazjum. Liczebności szkół obu typów w operatach z dwóch fal badania zaprezentowano w Tabeli 1.

Tabela 1. Liczebności populacji, z których losowano próbę w dwóch falach badania.

	Gimnazja	Licea	Technika	Razem
Fala I	6 137	2 036	1 742	9 915
Fala II	6 103	2 001	1 692	9 796

Źródło: Opracowanie własne na podstawie wyników badania ankietowego.

Dodatkowo, dobór próby miał w szczególny sposób uwzględniać wielkość miejscowości, w której znajduje się szkoła oraz efektywność i poziom nauczania w szkołach. Uznano, że kontrola tych zmiennych pozwoli uniknąć skrzywień w próbie, a jednocześnie są to zmienne, które dało się kontrolować na poziomie doboru próby, gdyż ich wartości dla wszystkich szkół były znane już przed badaniem. Zachodziła bowiem obawa, że szkoły o niskim poziomie lub niskiej efektywności nauczania mniej chętnie korzystają ze wskaźników EWD. Jednocześnie można przypuszczać, że szkoły te mogą mniej chętnie brać udział w badaniu, gdyż dyrektorzy mogli by chcieć unikać przyznania się do faktu nie wykorzystywania wskaźników w ich szkołach. Gdyby przypuszczenie to okazało się prawdziwe, a w próbie znalazła by się nadreprezentacja szkół o wysokim poziomie nauczania lub o wysokiej efektywności, to oszacowanie odsetka szkół korzystających ze wskaźnika EWD zostałoby tym sposobem zawyżone. Aby uniknąć takiego ryzyka próbę dobrano w sposób gwarantujący, że znajdują się w niej szkoły zarówno o wysokiej, jak i niskiej efektywności nauczania, wysokim i niskim poziomie nauczania oraz z dużych i małych miejscowości. Udało się to osiągnąć przez zastosowanie doboru systematycznego z warstwowaniem implícite. Również dobór szkół rezerwowych, czyli takich, w których realizowano badanie dopiero w sytuacji, gdy nie udało się go zrealizować w odpowiadającej szkole z próby głównej, przeprowadzono w sposób gwarantujący, że szkoła rezerwowa dla danej szkoły posiada zbliżone do niej cechy.

Warstwowanie implícite w procedurze doboru systematycznego polega na uporządkowaniu operatu, z którego losowane będą jednostki, według zmiennych warstwujących. Dopiero z tak uporządkowanego operatu losowane są jednostki. Dzięki temu, gdy po określeniu interwału losowania i wylosowaniu punktu startowego do próby włączane są kolejne jednostki znajdujące się na pozycjach odpowiadających wartościom punktu startowego i wielokrotnościom interwału, do próby powinny być włączane jednostki z całego przekroju wartości, jakie przyjmują zmienne, według których uporządkowano zbiór. Naturalnie, warstwowanie to nie jest tak dokładne, jak warstwowanie explicito stosowane zwykle przy doborze warstwowym. W przeciwieństwie do losowania warstwowego proporcjonalnego, gdzie założony rozkład zmiennej w każdej próbie jest identyczny z populacyjnym (z dokładnością do zaokrągleń wynikających z konieczności losowania całych jednostek), w losowaniu systematycznym z warstwowaniem implícite tylko oczekiwany rozkład zmiennej w próbie odpowiada rozkładowi populacyjnemu. Należy jednak pamiętać, że wariancja rozkładu w próbie jest znacznie mniejsza niż w przypadku gdy nie wykonuje się warstwowania implícite, tj. np. przy doborze prostym lub systematycznym bez warstwowania implícite. Niemniej jednak warstwowanie implícite znacząco zwiększa efektywność doboru, zwłaszcza, gdy cechy według których uporządkowany jest operat są zmiennymi silnie determinującymi wartości zmiennych, które są mierzone w badaniu. W niniejszym raporcie nie zbadano korelacji zmiennych warstwujących z innymi zmiennymi, należy jednak przypuszczać, że owa korelacja zachodzi, gdyż wieloletnie doświadczenia badań edukacyjnych skłaniają do przekonania, że wielkość miejscowości, w której znajduje się szkoła, oraz poziom i efektywność nauczania silnie warunkują inne zmienne opisujące szkoły.

Specyfiką doboru systematycznego jest fakt, że przy danej liczebności operatu nie pozwala on dobrać prób o dowolnej liczebności oraz sama liczebność próby przy określonej procedurze doboru jest zmienną losową. Wynika to z faktu, że liczba jednostek w operacie zwykle nie jest podzielna przez interwał losowania. Wielkość interwału losowania jest liczbą jednostek w populacji podzieloną przez zakładaną liczebność próby i zaokrągloną w górę. Zaokrąglenie w górę jest konieczne, aby żadna jednostka z operatu nie miała prawdopodobieństwa znalezienia się w próbie równego zero³. Zjawisko to zwiększa się dodatkowo, gdy w doborze systematycznym losuje się kilka podprób dla możliwości oszacowania w sposób najbardziej precyzyjny wariancji estymatorów. W omawianym badaniu w każdym typie szkół wylosowano pięć podprób. Jest to w pewnej mierze przyczyną faktu, że liczba jednostek zbadanych w każdym typie szkoły odbiega od 500. Jest to również przyczyną różnej liczebności prób realizowanych w poszczególnych typach szkół. Należy jednak pamiętać, że w głównej mierze zróżnicowane liczebności szkół wynikają z niechęci ze strony szkół do udziału w badaniu i innych problemów realizacyjnych. Uzyskane liczebności prób wylosowanych i zbadanych zawiera Tabela 2.

³ Dobór systematyczny można co prawda wykonać w wersji dopuszczającej niecałkowite interwały losowania, jednak powoduje to obciążenie oszacowania wariancji estymatorów. Obciążenie to jest tym większe, im większą część populacji stanowi dobierana próba. W omawianym badaniu stosunek wielkości próby do populacji w przypadku techników wyniósł prawie ¼, co groziłoby obciążeniem wyników. Wydaje się, że właśnie to było przyczyną rezygnacji z tej wersji systematycznego doboru próby (zob. Carl-Erik Särndal, Bengt Swensson, Jan Wretman, *Model Assisted Survey Sampling*, New York 2003).

Tabela 2. Klasyfikacja przyczyn braku udziału na losowe i nielosowe z fali I badania.

		Gimnazja	Licea	Technika	Razem
Fala I	Wylosowane	497	480	492	1 469
	Zrealizowane	430	389	404	1 223
Fala II	Wylosowane	493	476	498	1 467
	Zrealizowane	435	403	414	1 252

Źródło: Raport z ewaluacji wewnętrznej projektu EWD.

Aby możliwie zminimalizować problem z odmowami udziału w badaniu oraz innymi przyczynami niemożliwości realizacji badania w części szkół, zastosowano zastępowanie jednostek z próby zasadniczej jednostkami z prób rezerwowych. Próby rezerwowe dobrano przy okazji systematycznego doboru próby zasadniczej: pierwszą próbę rezerwową stanowiły szkoły bezpośrednio poprzedzające szkoły z próby zasadniczej w uporządkowanym operacie, zaś drugą próbę rezerwową szkoły następujące po nich. Dzięki temu maksymalizowane jest podobieństwo szkół z prób rezerwowych do szkół z próby zasadniczej pod względem zmiennych za pomocą których posortowano zbiór.

Należy mieć na uwadze, że przy obliczaniu estymatorów punktowych oraz ich błędów nie brano pod uwagę informacji o tym, czy szkoła pochodzi z próby zasadniczej, czy też rezerwowej, lecz potraktowano je tak, jakby wszystkie pochodziły z próby zasadniczej. Inaczej mówiąc, można powiedzieć, że przyjęto, że szkoły z prób rezerwowych są identyczne ze szkołami z próby zasadniczej. Również nie wzięto pod uwagę faktu, że w części szkół nie udało się zrealizować wywiadu zarówno w szkole należącej do próby zasadniczej, jak i w szkołach rezerwowych. Założenia te przyjmowane są arbitralnie i powodują niedoszacowanie błędów oszacowań, są jednak powszechnie stosowane w praktyce badawczej na całym świecie i konieczne dla zastosowania większości procedur statystycznych. Ponadto, odsetek szkół, w których nie udało się zrealizować badania jest względnie niski (16,7% w I fali badania oraz 14,7% w II fali badania), a procedura doboru prób rezerwowych rzeczywiście zapewnia podobieństwo szkół rezerwowych do zasadniczych, co sprawia, że przyjmowane założenia nie odstają dalece od rzeczywistości. Wyjątkiem jest kluczowy pomiar, jaki ma na celu badanie, czyli oszacowanie wartości wskaźnika wykorzystania wskaźników EWD w szkołach – w przypadku tej analizy podjęto równoległe dwa sposoby oszacowania wskaźnika: z przyjęciem upraszczających założeń oraz bez ich przyjmowania.

3.2. Realizacja wywiadów

Badania realizowane były poprzez wywiady telefoniczne (tzw. CATI). W czasie realizacji napotkano na problemy typowe w badaniach ankietowych: respondenci odmawiali udziału w badaniu lub próby kontaktu ze szkołami kończyły się niepowodzeniem. Gdy mimo kilku prób wciąż nie udawało się skontaktować ze szkołą lub dyrekcja szkoły wyraźnie odmówiła udziału w badaniu, próbowano skontaktować się ze szkołą z próby rezerwowej. Dla każdej szkoły dobrane były dwie szkoły rezerwowe. Dokonano tego w sposób gwarantujący podobieństwo do szkoły z próby zasadniczej (zastosowaną w tym celu procedurę opisano w części 3.1). Mimo zastosowania procedury zastępowania szkół, nie udało się zrealizować badania we wszystkich założonych szkołach lub ich rezerwach.

Zjawisko braków udziału, a w mniejszej mierze również procedura zastępowania szkół, może negatywnie wpływać na badanie powodując obciążenie. Jeżeli bowiem respondenci o pewnych cechach mniej chętnie brali udział w badaniu, a o innych chętniej, skutkować to może obciążeniem oszacowania innych zmiennych w próbie. Obciążenie to pojawia się tylko wtedy, gdy udział lub brak

udziału w badaniu zależy od cech respondenta lub cech reprezentowanej przez niego szkoły. Z taką sytuacją w szczególności mamy do czynienia, gdy udział lub brak udziału w badaniu zależy od woli respondenta. Gdy respondent zdecydowanie odmawia udziału w badaniu, zwłaszcza gdy przyznaje, że ma to związek z tematem badania (co czasami miało miejsce), należy oczekiwać, że ma to związek z jego cechami i cechami szkoły powiązanymi z przedmiotem badania, a w związku z tym może powodować obciążenie.

Oczywiście część braków udziału i konieczności zastępowania szkół nie było związanych z wolą respondenta. Gdy brak udziału wynika z sytuacji losowych, czyli takich, które mogą z podobnym prawdopodobieństwem dotknąć każdej szkoły, jak np. zmiana dyrektora czy choroba dyrektora, braki udziału nie są tak groźne, gdyż raczej nie są związane z cechami szkoły, a zatem wpływają na błąd tylko w ten sposób, że z powodu realizacji mniejszej liczby szkół, zwiększa się losowy składnik błędu, czyli błąd statystyczny.

Możliwa jest jeszcze trzecia sytuacja: jeżeli szkoła została zlikwidowana lub w wyniku reorganizacji przestała być samodzielną placówką, to oznacza to, że wylosowana z operatu jednostka w rzeczywistości nie należy do badanej zbiorowości. Wtedy można nie uwzględniać jej w dalszej procedurze estymacji bez żadnych konsekwencji dla wyników badania. Te braki udziału traktowane będą więc tak samo jak losowe braki udziału, i chociaż w istocie ich natura jest nieco odmienna, w dalszych zestawieniach zostaną ujęte razem z nimi.

Systematyczność lub losowość braków udziału, i odsetki jakie braki udziału każdego z dwóch typów stanowią w liczebności wylosowanej próby jest kluczowa dla oceny reprezentatywności próby. Z tego powodu podjęta została analiza zarejestrowanych braków udziału pod kątem charakteru błędów, jakie mogą powodować. W przypadku I fali badania kategoryzacja i analiza braków danych została już podjęta w *Raporcie z ewaluacji wewnętrznej projektu EWD*. Niniejszy raport stara się kontynuować rozstrzygnięcia jakie w tej kwestii poczynili autorzy raportu z 2014 roku. Za systematyczne braki udziału uznali oni sytuacje, gdy dyrektor nie chce brać udziału w badaniu, w szczególności gdy:

- nie wskazuje przyczyny odmowy,
- jako przyczynę wskazuje niewykorzystywanie wskaźników EWD w szkole lub nie orientowanie się w tematyce EWD,
- podaje przyczynę bardzo ogólną (jak brak czasu, nadmiar badań).

Wyjątkiem jest sytuacja, gdy dyrektor odmówił wzięcia udziału w badaniu, gdyż:

- nie odpowiada mu technika badania, lecz wyrażał on gotowość do wzięcia udziału w badaniu przeprowadzonym inną techniką (jako preferowaną formę ankiety często wskazywano ankietę internetową),
- nie chciał jeszcze raz przechodzić przez ten wywiad, gdyż wcześniej przeprowadzono z nim rozmowę na temat innej placówki.

Pozostałe rodzaje braków udziału, czyli sytuacje gdy z potencjalnym respondentem w ogóle nie udało się skontaktować lub kontakt nawiązano, ale wywiadu nie zrealizowano z powodu odkładania długotrwałego wywiadu skategoryzowano jako losowe braki udziału, gdyż przyczyna braku realizacji badania w szkole była inna niż brak woli respondenta. Kategoryzację przyczyn braku udziału w I fali badania prezentują Tabela 3

Tabela 3. Klasyfikacja przyczyn braku udziału na losowe i nielosowe z fali I badania.

Nielosowe braki udziału	Losowe braki udziału
– odmowa bez podania przyczyny. – Brak czasu lub chęci udziału w badaniu. – Nadmiar badań. – Szkoła nie korzysta z EWD. – Inne przyczyny odmowy.	– Z respondentem przeprowadzono już wywiad na temat innej placówki. – Nieobecność dyrektora. – Technika wywiadu. – Nowy dyrektor. – Długotrwałe przekładanie wywiadu. – Szkoła zlikwidowana lub zreorganizowana.

Źródło: Raport z ewaluacji wewnętrznej projektu EWD.

Niestety, w II fali badanie realizowane było przez innego wykonawcę, co sprawiło, że odmowy udziału w badaniu klasyfikowane były według innych kryteriów. Niemniej jednak podjęto próbę ich skategoryzowania kierując się podobnymi zasadami. Rezultat prezentuje Tabela 4.

Najliczniej pojawiały się w badaniu braki udziału opisane przez wykonawcę jako „brak możliwości realizacji badań w wyznaczonym terminie”. Z jednej strony były to sytuacje, gdy respondent z braku czasu przekładał na wywiad na termin późniejszy. Trudno w takiej sytuacji uważać, że respondent w ogóle nie chciał poddać się procedurze badawczej. Z drugiej zaś strony, postępowanie takie mogło być celowe i mieć na celu niewzięcie udziału w badaniu. Ostatecznie uznano, że połowa braków udziału z tej kategorii uznana będzie za losową, a połowa za systematyczną. Chociaż jest to postępowanie arbitralne, to wydaje się, że raczej zawyża ono liczbę wywiadów celowo odkładanych w czasie, a zatem raczej spowoduje to przeszacowanie, niż niedoszacowanie rozmiaru błędów badania. Dlatego nie spowoduje to wyciągnięcia nieuzasadnionych wniosków.

Tabela 4. Klasyfikacja przyczyn braku udziału na losowe i nielosowe z fali II badania.

Nielosowe braki udziału	Losowe braki udziału
– Brak czasu. – Brak obowiązku udziału w badaniu. – Szkoła nie prowadzi analiz EWD. – Szkoła nie udziela informacji przez telefon. – Niechęć do badania. – Nieznajomość problematyki EWD. – Odmowa podania przyczyny. – Sekretariat uniemożliwił kontakt z dyrekcją. – Szkoła jest niepubliczna i z tego powodu odmawia udziału w badaniu.	– Ankieta przerwana w trakcie realizacji. – Błędne dane szkoły. – Brak kontaktu w wyznaczonym terminie. – Brak kontaktu ze szkołą. – Likwidacja placówki. – Nowa dyrekcja. – Przeprowadzenie wywiadu na temat innej placówki w zespole szkół.

Źródło: opracowanie własne na podstawie wyników badania ankietowego.

Na podstawie klasyfikacji braków udziału policzono efektywne odsetki realizacji próby zasadniczej. Tabela 5 prezentuje uzyskane wartości dla fali I badania, zaś Tabela 6 dla II. Wyliczenia dotyczą tylko szkół z próby zasadniczej, gdyż traktując zagadnienie ściśle, tylko je można uważać za prawdziwą próbę losową. Zatem znajdujące się w tabelach liczebności dotyczą liczby zrealizowanych wywiadów z próby zasadniczej oraz liczby systematycznych braków udziału z próby zasadniczej. W tabeli dotyczącej II fali uzyskano liczebności ułamkowe, co wynika z uznania, że połowę wywiadów, dla których jako przyczynę braku realizacji wskazano „brak możliwości realizacji badań w wyznaczonym terminie” należy uznać za losową, a połowę za systematyczną. Wydaje się, że choć jest to nieintuicyjne, to nie przeszkadza w poprawnym szacowaniu efektywnego odsetka realizacji próby.

Tabela 5. Realizacja wywiadów z próby zasadniczej i efektywny poziom realizacji próby z fali I badania.

	Gimnazja	Licea	Technika	Razem
Zbadane szkoły z próby zasadniczej (A)	284	269	283	836
Nielosowane braki udziału (B)	99	90	78	267
Efektywny poziom realizacji badania: $A/(A+B)$	74,2%	74,9%	78,4%	75,8%

Źródło: Raport z ewaluacji wewnętrznej projektu EWD.

Tabela 6. Realizacja wywiadów z próby zasadniczej i efektywny poziom realizacji próby z fali II badania.

	Gimnazja	Licea	Technika	Razem
Zbadane szkoły z próby zasadniczej (A)	353	317	348	1018
Nielosowane braki udziału (B)	76,5	85,5	76,5	238,5
Efektywny poziom realizacji badania: $A/(A+B)$	82,2%	78,8%	82,0%	81,0%

Źródło: opracowanie własne na podstawie wyników badania ankietowego.

Uzyskane odsetki zostaną wzięte pod uwagę przy szacowaniu odsetka szkół wykorzystujących wskaźniki EWD w wersji pesymistycznej w rozdziale 5.2.3.

3.3. Błąd statystyczny

Procedura doboru próby w I i II fali badania była taka sama. Procedurę tę należy określić jako jednostopniową, niezależną, ze zróżnicowanymi prawdopodobieństwami, systematyczną w obrębie trzech warstw. Z założenia jest to próba reprezentatywna dla populacji gimnazjów, techników i liceów w Polsce. Z obu prób wyodrębnić można wiele podprób. W raporcie w pewnych analizach rozważane będą trzy podpróby:

- próba gimnazjów – reprezentatywna dla populacji gimnazjów w Polsce,
- próba liceów – reprezentatywna dla populacji liceów,
- próba techników – reprezentatywną dla populacji techników w Polsce.

Każdą z tych podprób można rozważać jako próbę dobraną w sposób jednostopniowy, niezależny, systematyczny, z równymi prawdopodobieństwami. Jak wskazuje literatura, zastosowanie warstwowania implícite sprawia, że pod względem efektywności próbę traktować można tak, jakby została dobrana w sposób prosty⁴. Oszacowania na poziomie podprób obejmujących jeden typ placówek z jednej fali badania traktowane będą w analizach jak próby proste niezależne. Z kolei na poziomie ogółu szkół biorących udział w jednej fali badania, z racji niezależnego losowania szkół w obrębie poszczególnych typów oraz nierównych prawdopodobieństw wylosowania do próby różnych typów szkół, właściwie jest traktowanie próby jako składającej się z trzech warstw:

- warstwy gimnazjów,

⁴ Cochran William G., *Sampling Techniques*, Wiley 1977.

- warstwy liceów,
- warstwy techników.

Prawdopodobieństwo znalezienia się szkoły określonego typu w próbie określone jest przez proporcję: liczby szkół tego typu, w których zrealizowano badanie do liczby szkół, które znajdowały się w operacie. Prawdopodobieństwa te są znacząco zróżnicowane, gdyż jak wskazywano w części

Powoduje to wzrost wariancji wag, a zatem również spadek efektywności doboru próby względem hipotetycznego doboru prostego (według popularnej procedury efektywną wielkość próby można oszacować na ok. 940-950 jednostek, choć właściwe liczebności próby wyniosły 1 223 w I fali oraz 1 252 w II).

Warstwowy schemat doboru próby, zróżnicowane prawdopodobieństwa znalezienia się szkół w próbie, konieczność porównywania dwóch prób szkół, a wreszcie specyficzność hipotez, które badano, sprawiły, że dokonanie porównań i uzyskanie poprawnych wartości testów statystycznych wymagało zastosowania nietypowych rozwiązań. Z tego powodu obliczenia wykonano w środowisku R w pakiecie „survey”⁵ dedykowanym do wnioskowania statystycznego przy złożonych schematach doboru próby. W większości sytuacji udało się wykonać analizy za pomocą funkcjonalności zawartych w pakiecie. Sytuacja taka miała miejsce w przypadku:

- analiz w obrębie próby z jednej fali badania,
- analiz porównawczych między dwiema falami badania.
- analiz w obrębie podpróby szkół jednego typu w obrębie badania,
- analiz porównawczych między dwiema podpróbami szkół jednego typu między dwiema falami badania.

Problematyczne były jednak analizy prowadzone w rozdziale 4.1, które wymagały postępowania nietypowego, tj. badania hipotez o zgodności rozkładu uzyskanego w próbie z rozkładem zmiennej zastanym w operacie. W celu przeprowadzenia tych analiz konieczna była zmodyfikacja funkcji z pakietu „survey”⁶. We wszystkich analizach zastosowano poprawki na skończoną wielkość badanej populacji.

4. Opis próby

Przed właściwymi analizami zbadano próby pod względem najbardziej ogólnych kryteriów. Wykonano trzy rodzaje porównań. W pierwszej kolejności porównano wylosowane próby z operatami, z których były one losowane. Analiza ta ma za zadanie odpowiedzieć na pytanie, czy nie następowała pewna autoselekcja respondentów do badania. W drugiej kolejności zbadano, czy respondenci deklarujący korzystanie z EWD jednocześnie reprezentują szkoły cechujące się niższymi wskaźnikami EWD oraz wynikami egzaminacyjnymi. Jeżeli odpowiedzi na oba te pytania byłyby twierdzące, to skłaniałoby to do wniosku, że oszacowania wskaźników wykorzystania pomiarów EWD są obciążone. W dalszej

⁵ T. Lumley (2012) "survey: analysis of complex survey samples". R package version 3.28-2.

⁶ Za pomoc w pracy nad tą częścią, podziękowania należą się Tomaszowi Żółtakowi, który służył wsparciem i wiedzą merytoryczną w wykonaniu najtrudniejszych modyfikacji.

kolejności porównano strukturę prób I i II fali badania pod kątem cech neutralnych dla pytań badawczych, ale takich których niezgodność między oboma badaniami mogłaby sugerować pewne problemy z danymi.

4.1. Reprezentatywność próby

Duży zakres informacji posiadanych o szkołach jeszcze przed ich wylosowaniem pozwala na ocenę *ex post* zrealizowanych prób pod kątem ich reprezentatywności. Informacje, o których mowa, znajdowały się w operatach, czyli w zbiorach, z których losowano próbę. Pozwala to na zbadanie parametrów rozkładów zmiennych w próbie i w operacie, i oceny reprezentatywności próby pod względem tych cech. Oczywiście, jest wiadome, że próba wylosowana jeżeli została dobrana w sposób losowy i zgodny ze sztuką musi być reprezentatywna dla operatu, z którego była losowana pod względem dowolnych cech. Problem polega jednak na tym, że próba wylosowana nie jest tożsama z próbą zrealizowaną: jak wskazano w rozdziale 3.2 wystąpiły typowe i zwykle niedające się uniknąć problemy z odmowami udziału w badaniu.

Należy jednak pamiętać, że problemy takie mogą wpłynąć na wynik badania tylko wtedy, gdy korelują z cechami szkół, których zbadanie jest przedmiotem badania. Przedmiotem badania było wykorzystywanie wskaźników EWD, o czym zresztą respondentów informowano. Naturalnym przypuszczeniem jest, że szkoły o słabych wynikach czy niskich wskaźnikach EWD mniej chętnie będą brać udział w badaniu. Przypuszczenie to sprawdzono w tym rozdziale.

Teoretycznie na temat szkół znajdujących się w operacie można pobrać z Systemu Informacji Oświatowej (dalej SIO) niemal dowolnie wiele zmiennych opisujących szkoły, takich jak liczba uczniów w szkole, czy charakter szkoły (publiczny/niepubliczny, specjalny/interakcyjny/ogólnodostępny). Niestety, nie ma prostego sposobu na przyłączenie tych danych do operatu, gdyż SIO i bazy CKE stanowią osobne systemy danych i nie ma klucza, który pozwala łączyć dane między nimi. Wydaje się jednak, że dane jakie znajdowały się w operatach i pochodziły z danych CKE wystarczą do zbadania najważniejszych hipotez dotyczących możliwego kierunku odchylenia wyników w próbie zbadanej względem populacji szkół. Ze zmiennych znajdujących się w operacie do analizy wybrano:

- informacje o wielkości miejscowości, w której znajduje się szkoła (dokładnie informacja, czy szkoła znajduje się na wsi oraz liczba mieszkańców miasta, jeżeli szkoła znajdowała się w mieście),
- średni wskaźnik EWD (uśredniony wskaźnik z przedmiotów humanistycznych i matematyczno-przyrodniczych),
- średni wynik egzaminów zewnętrznych (uśrednione wyskalowane wyniki z egzaminów zewnętrznych, na podstawie których liczone były wskaźniki EWD).

Przy tej okazji warto wyjaśnić, że operaty z I i II fali badania nieznacznie się różniły pod względem jednostek, które się na nie składały. Wynika to z faktu, że próbę za każdym razem losowano spośród szkół, dla których dostępne były trzyletnie wskaźniki EWD dla lat od 2011 do roku badania. Między rokiem 2013 i 2015 liczba takich szkół nieznacznie się zmniejszyła. Dlatego niewielkie różnice w rozkładach zmiennych w operatach dwóch edycji badania są naturalne. Należy również pamiętać, że wskaźniki EWD i wyskalowane wyniki egzaminacyjne nie są porównywalne między latami, stąd analiza porównująca te wartości między rokiem 2013 i 2015 nie miała by sensu. Należy też zwrócić uwagę na fakt, że chociaż średnie wyniki egzaminacyjne i wskaźniki EWD są skalowane odpowiednio do średnich 100 i 0, to nie znajduje odzwierciedlenia w operatach. Wynika z dwóch przyczyn:

- Po pierwsze, wyniki egzaminacyjne są skalowane, a wskaźniki EWD liczone dla szerszej zbiorowości szkół, niż znajdujące się w operatach, bo w operatach znajdują się tylko szkoły, dla których dostępne są wskaźniki dla lat od 2011 roku do roku badania. Wykluczono więc z niego szkoły dla których wskaźniki dostępne są dla krótszego okresu czasu, chociaż brały one udział w obliczeniach, i dopiero razem z nimi średnie były by zgodne z uzyskanymi w populacyjnych obliczeniach.
- Po drugie, modele skalujący i wykorzystywany do liczenia wskaźników EWD operują na poziomie ucznia i sprowadzają średnią do pożądanej wartości na zbiorowości uczniów, a zatem po zagregowaniu wyników do poziomu szkół, średnie na poziomie szkół mogą być odmienne.

Wielkość miejscowości, w której znajduje się szkoła

Zmienną mówiącą o liczbie mieszkańców miejscowości przekodowano na kategorie i na tak przygotowanej zmiennej porównano rozkład w operacie z rozkładem w próbie zrealizowanej. Wyniki prezentuje Tabela 7.

Rozkłady zmiennej są do siebie bardzo zbliżone między próbą zrealizowaną i operatem w obu edycjach badania. Dla I fali największe różnice odnotowano dla kategorii „miasto od 50 do 200 tys. mieszkańców” (1,2 p. p.), a dla II fali dla kategorii „Miasto od 200 do 500 tys. mieszkańców” (1,6 p.p). Różnice są zatem niewielkie. Aby uzyskać ścisłą odpowiedź na pytanie, czy różnice tej wielkości powinny powodować obawę o niereprezentatywność próby, wykonano specyficzny test statystyczny. Zbadano jakie jest prawdopodobieństwo wylosowania próby o odnotowanym lub większym odchyleniu od rozkładu z operatu (przy założeniach na temat doboru próby sformułowanych w rozdziale 3.3). Wynik testu to wartość p równa 0,713 ($\chi^2=2,91$; $df=5$) dla I fali oraz 0,166 ($\chi^2=7,83$; $df=5$) dla fali II. Oznacza to, że prawdopodobieństwo wylosowania próby w takiej mierze odbiegającej od operatu dla roku 2013 wyniosło 0,748, zaś dla roku 2015 0,153. Prawdopodobieństwa te są wysokie, należy więc przyjąć, że próby odpowiednio reprezentują populację ze względu na kategorię wielkości miejscowości w jakiej znajduje się szkoła.

Tabela 7. Porównanie rozkładu kategorii wielkości miejscowości w próbach i operatach.

	Kategoria wielkości miejscowości	w operacie	w próbie
Badanie 2013	Wieś	38,9%	39,0%
	Miasto do 20 tys. mieszkańców	17,6%	17,8%
	Miasto od 20 do 50 tys. mieszkańców	11,8%	12,0%
	Miasto od 50 do 200 tys. mieszkańców	16,2%	15,0%
	Miasto od 200 do 500 tys. mieszkańców	6,9%	7,9%
	Miasta powyżej 500 tys. mieszkańców	8,5%	8,4%
Badanie 2015	Wieś	41,8%	39,9%
	Miasto do 20 tys. mieszkańców	17,7%	16,7%
	Miasto od 20 do 50 tys. mieszkańców	11,2%	11,8%
	Miasto od 50 do 200 tys. mieszkańców	16,1%	15,7%
	Miasto od 200 do 500 tys. mieszkańców	5,9%	7,5%
	Miasta powyżej 500 tys. mieszkańców	7,4%	8,4%

Źródło: opracowanie własne na podstawie wyników badań ankietowych.

Analogiczne obliczenia wykonano dla średnich wyników egzaminacyjnych szkół oraz wskaźników EWD. Wyniki prezentuje Tabela 8. Dla II fali badania różnice zarejestrowane między operatem i

zrealizowaną próbą są minimalne. Oszacowane prawdopodobieństwa otrzymania odchyleń równej wielkości lub większych wyniosło aż 0,842 w przypadku średnich wyników egzaminacyjnych oraz 0,946 w przypadku wskaźników EWD. Niestety, sytuacja jest odmienna w przypadku I fali badania. Różnice były na tyle duże, że oszacowane prawdopodobieństwo w przypadku średnich wyników egzaminacyjnych wyniosło 0,020, zaś w przypadku wskaźników EWD 0,042. Są to niskie prawdopodobieństwa, które przy zwyczajowo przyjmowanym poziomie istotności 0,05 należy uznać za nieprzypadkowe. Sugeruje to, że procesy mające miejsce przy realizacji badania sprawiające, że pewne szkoły rezygnowały z udziału w badaniu mogły mieć charakter systematyczny. Wyższe wskaźniki poziomu umiejętności i efektywności nauczania odnotowano w szkołach zbadanych. Skłania to do wniosku, że w I fali badania mniej chętnie brały udział szkoły o niskich wskaźnikach EWD.

Tabela 8. Porównanie średnich wyników egzaminacyjnych oraz wskaźników EWD między operatem a zbadaną próbą.

	fala		średnia	bł. std.	t	wartość p
Średnie wyniki egzaminacyjne	I	operat	98,708	-	2,330	0,020
		próba	99,106	0,171		
	II	operat	98,642	-	0,190	0,842
		próba	98,682	0,202		
Wskaźniki EWD	I	operat	-0,275	-	2,038	0,042
		próba	-0,126	0,083		
	II	operat	-0,250	-	0,068	0,946
		próba	-0,244	0,073		

Źródło: opracowanie własne na podstawie wyników badań ankietowych.

Wykazano, że zrealizowana próba znacząco odbiega od operatu pod względem mierników poziomu nauczania i efektywności nauczania w szkole. Wynik ten uzyskano na połączonej zbiorowości gimnazjów, liceów i techników. Aby pogłębić analizę analogiczne testy przeprowadzono osobno dla każdego typu placówek. Wyniki prezentują Tabela 9 (średnie wyniki) i 10 (wskaźniki EWD).

Odchyłeń nie obserwuje się w zbiorowości gimnazjów i techników. W tych typach placówek różnice między operatem a próbą są niewielkie zarówno pod względem średnich wyników, jak i wskaźników EWD. Prawdopodobieństwa uzyskania zaobserwowanych odchyleń oszacowane zostały na raczej wysokie (od 0,256 do 0,355). Duże różnice odnotowano natomiast w przypadku liceów. Prawdopodobieństwo ich uzyskania w przypadku wyników egzaminów oszacowano na 0,005 zaś w przypadku wskaźników EWD na 0,051. Można wyciągnąć wniosek, że to procesy realizacji w obrębie liceów spowodowały niższy poziom średnich egzaminów oraz wskaźników EWD w zbadanych szkołach.

Tabela 9. Porównanie średnich wyników egzaminacyjnych między operatem i próbą dla roku 2013.

Zmienna	operat	próba	t	wartość p
Gimnazja	99,705	99,922	0,924	0,355
Licea	95,948	97,016	2,787	0,005
Technika	98,425	98,672	1,050	0,294

Źródło: opracowanie własne na podstawie wyników badań ankietowych.

Tabela 10. Porównanie średnich wskaźników EWD między operatem i próbą dla roku 2013.

Zmienna	operat	próba	t	wartość p
Gimnazja	0,189	0,279	0,926	0,354
Licea	-1,381	-1,061	1,952	0,051
Technika	-0,621	-0,463	1,135	0,256

Źródło: opracowanie własne na podstawie wyników badań ankietowych.

4.2. Wartość wskaźników EWD, a korzystanie z nich w ewaluacji

W poprzedniej części wykazano, że szkoły, w których udało się zrealizować badanie cechowały się niższymi wartościami wskaźników EWD oraz średnich wyników egzaminacyjnych. Dotyczyło to liceów zbadanych w I fali badania. W tej części podjęto próbę odpowiedzi na pytanie, czy szkoły, których dyrektorzy deklarują wykorzystywanie wskaźników EWD posiadają jednocześnie wyższe wskaźniki EWD.

W tym celu przeprowadzono porównanie średnich wartości wskaźników EWD oraz wyników egzaminacyjnych w grupie szkół korzystających z EWD i niekorzystających wraz z testami istotności. Porównanie przeprowadzono oddzielnie dla każdego typu szkół. Gdyby porównano odsetki odnotowane na próbie wszystkich szkół, ewentualna różnica w średnich wartościach między szkołami wykorzystującymi EWD i niewykorzystującymi mogłaby wynikać – nie z samej różnicy między korzystającymi z EWD, a niekorzystającymi w obrębie jednego typu placówek – ale z różnego odsetka szkół korzystających z EWD w obrębie szkół różnego typu. A przecież nie chodzi nam o zbadanie zależności wskaźników EWD od typu szkoły, lecz zależności wartości wskaźników od korzystania z nich w ewaluacji. Każdego z dyrektorów pytano kilkakrotnie o korzystanie ze wskaźników EWD za każdym razem w odniesieniu do innego roku. Najwłaściwsze wydało się takie porównanie, które spełniało by dwa warunki:

- porównanie średnich wartości wskaźników EWD dla danego roku powinno odbywać się między grupami wyróżnionymi na podstawie pytania o korzystanie z EWD w tym samym roku (dla którego policzone były wskaźniki),
- inne czynniki (takie jak odległość w czasie czy kontekst badania) powinny zostać tak dobrane, aby maksymalizować dokładność porównania.

Z tego powodu dla I fali badania wybrano pytanie o korzystanie ze wskaźników EWD w roku 2012, zaś dla drugiej – w 2014. Poczynano tak dlatego, że w poszczególnych falach właśnie dla tych lat wszyscy dyrektorzy powinni posiadać najświeższe informacje o ich wartościach. Dla I fali badania nie wybrano roku 2013, gdyż wielu dyrektorów w czasie badania jeszcze nie zdążyło przeprowadzić analiz efektywności nauczania w tym roku, co mogłoby zakłócić wynik porównania. Wyniki porównania zaprezentowano w Tabeli 11.

W przypadku wszystkich czterech porównań dotyczących gimnazjów wynik okazał się nieistotny statystycznie. Inaczej było jednak w przypadku liceów i techników – w przypadku tych placówek, tam gdzie korzystano ze wskaźników EWD szkoła jednocześnie uzyskiwała wyższe wskaźniki EWD oraz wyższe wyniki egzaminacyjne. Dla czterech testów dotyczących liceów i dla czterech testów dotyczących techników testy statystyczne były istotne na poziomie istotności 0,001. Dla wszystkich porównań średnich wyższe wartości odnotowano w szkołach korzystających ze wskaźników.

Należy przyjąć, że w przypadku gimnazjów korzystanie ze wskaźników EWD nie koreluje z efektywnością nauczania i poziomem nauczania, zaś w przypadku liceów i techników korelacja taka zachodzi i jest znacząca: różnica między szkołami korzystającymi a niekorzystającymi wynosi między 1/3 a 1/2 odchylenia standardowego.

Tabela 11. Porównanie średnich wskaźników EWD i średnich wyników egzaminacyjnych między szkołami korzystającymi i niekorzystającymi ze wskaźników EWD.

	Fala	typ szkoły	korzystające	niekorzystające	stat. t	wartość p
Wskaźniki EWD	I	Gimnazja	0,367	-0,090	1,865	0,062
		Licea	-0,584	-2,316	4,689	0,000
		Technika	-0,058	-1,162	4,012	0,000
	II	Gimnazja	0,179	0,124	0,180	0,858
		Licea	-0,984	-2,716	3,886	0,000
		Technika	-0,235	-1,513	4,410	0,000
Średnie wyniki	I	Gimnazja	100,029	99,468	1,062	0,288
		Licea	98,150	94,028	4,979	0,000
		Technika	99,249	97,673	3,387	0,001
	II	Gimnazja	99,426	98,871	0,592	0,554
		Licea	97,311	93,136	4,168	0,000
		Technika	99,360	96,677	5,388	0,000

Źródło: opracowanie własne na podstawie wyników badań ankietowych.

4.3. Porównanie próby z 2013 i 2015 roku

Między badaniami zrealizowanymi w roku 2013 i 2015 porównano również rozkład płci, zajmowanego przez respondentów stanowiska i średni staż pracy oraz staż na stanowisku dyrektora.

Pod względem rozkładu płci w obu badaniach nie ma znaczących różnic. Różnica w odsetku kobiet i mężczyzn wyniosła 0,8 p.p. i jest nieistotna statystycznie ($\chi^2=0,161$; $df=1$; $p=0,708$). Dokładne wyniki prezentuje tabela 12.

Tabela 12. Porównanie rozkładu płci respondentów według badań z lat 2013 i 2015.

Płeć respondenta	2013	2015
Mężczyzna	33,4%	34,2%
Kobieta	66,6%	65,8%

Źródło: opracowanie własne na podstawie wyników badań ankietowych.

Jeżeli chodzi o stanowisko pełnione przez respondenta, to w II fali badania uczestniczyło mniej dyrektorów niż w I fali – ich odsetek był niższy o 6 p.p. Naturalnie, w II fali odsetek większy niż w I fali stanowili wicedyrektorzy. Różnica jest istotna statystycznie ($\chi^2=9,722$; $df=1$; $p=0,004$). Wyniki prezentuje tabela 13.

W II fali badania wywiad częściej odbywał się bezpośrednio z dyrektorem odpowiedzialnym za konkretną placówkę, nie zaś z dyrektorem całego zespołu szkół. Trudno jednak określić jaka mogła być tego przyczyna.

Tabela 13. Porównanie rozkładu stanowiska zajmowanego przez respondentów według badań z lat 2013 i 2015.

Zajmowane stanowisko	2013	2015
Dyrektor	74,0%	68,3%
Wicedyrektor	26,0%	31,7%

Źródło: opracowanie własne na podstawie wyników badań ankietowych.

Średni staż na stanowisku pracy wyniósł wśród zbadanych w I fali 8,24 lat, zaś wśród zbadanych w II – 9,10. Różnica jest istotna statystycznie. Podobną różnicę zaobserwować można na zmiennej staż pracy (25,64 względem 26,55 co daje 0,9 roku różnicy). Obie różnice są istotne statystycznie. Dokładne wyniki prezentuje tabela 14.

Różnica w średnich tej wielkości wydaje się naturalna, skoro społeczeństwo polskie się starzeje, i wydłużany jest wiek emerytalny. Zmiana średniego wieku mieszczą się w zakresie wartości od 0 do 2 lat daje się wytłumaczyć czasem jaki minął między I a II falą badania oraz zjawiskami ogólnospołecznymi.

Tabela 14. Porównanie średniego stażu pracy respondentów według badań z lat 2013 i 2015.

Zmienna	2013	2015	t	p
Staż na stanowisku	8,24	9,10	3,237	0,001
Staż pracy	25,64	26,55	3,009	0,003

Źródło: opracowanie własne na podstawie wyników badań ankietowych.

4.4. Wnioski z porównania prób

Analizy dowodzą, że zrealizowana próba liceów z roku 2013 wykazuje skrzywienie względem operatu, z którego została wylosowana. Przekłada się to również na skrzywienie próby gimnazjów, liceów, techników z I fali badania rozważanej jako całość. Skrzywienie to może mieć jedną z dwóch przyczyn:

- Procedura losowania: test istotności dał wartość p równą 0,02, co oznacza, że w 2% prób może zwyczajnie w wyniku losowania pojawić się takie odchylenie. Skrzywienie to może po prostu wynikać z „nieszczęśliwego” wylosowania próby.
- Autoselekcja szkół do badania: można przypuszczać, że dyrektorzy o niskich wskaźnikach EWD, a być może również – nie wykorzystujący EWD – mniej chętnie brali udział w badaniu. Może być to źródłem obciążenia dla przewidywanego w raporcie wskaźnika.

Aby wykluczyć sytuację, w której autoselekcja szkół do badania powodowałaby nadmiernie optymistyczne oszacowanie wskaźników, oprócz tradycyjnego wyliczania przedziałów ufności wykonana zostanie dodatkowa analiza uwzględniająca szkoły do których nie udało się dotrzeć w ramach procedury badawczej.

Problem ten w szczególności dotyczy analiz na liceach w roku 2013. W ich przypadku jednocześnie okazało się, że szkoły zbadane odstają pod względem poziomu i efektywności nauczania oraz szkoły, których dyrektorzy deklarują korzystanie ze wskaźników EWD cechują się wyższymi wskaźnikami EWD. Należy mieć to na względzie przy analizie wyników zawartych w części 5. Należy pamiętać, że wpływa to negatywnie również na oszacowania wskaźników cechujących całą próbę trzech typów szkół.

5. Wyniki analiz

W *Raporcie z ewaluacji wewnętrznej projektu EWD*, w którym analizowano wyniki z fali I badania wskaźnik wykorzystania wskaźników EWD w szkołach zoperacjonalizowano jako odsetek szkół, w których kiedykolwiek wykorzystano wskaźniki EWD. Podyktowane było to głównie faktem, że badanie realizowane było w październiku 2013 roku i wielu respondentów stwierdziło, że w tym roku jeszcze nie prowadzili analiz efektywności nauczania, dlatego odpowiedzi dotyczące samego roku 2013 nie mogły adekwatnie opisywać skali wykorzystania wskaźników EWD pod koniec projektu. Z drugiej strony, badanie wyników z roku 2012 uważano za mniej adekwatne, bo ewaluowany program trwał jeszcze cały rok 2012 i 2013 obejmując również liczne działania popularyzujące wykorzystanie wskaźników EWD. Z tego powodu uznano, że adekwatniej będzie jako wskaźnik wykorzystania wskaźników EWD wykorzystać odsetek szkół, które kiedykolwiek wykorzystywały wskaźniki EWD. Dla zachowania ciągłości w tym raporcie również policzono tak zdefiniowany wskaźnik.

Problematyczny okazał się jednak fakt, że w fali II badania, która odbywała się w 2015 roku nie pytano o korzystanie ze wskaźników w roku 2015, dlatego porównywalność tak liczonych wskaźników jest ograniczona. Ponadto uznano, że gdy porównywane będą odpowiedzi respondentów w I fali, gdzie zadawano pytanie o trzy okresy, w których prowadzono analizy z wynikami z II fali, gdzie pytano o cztery okresy, porównanie może okazać się nieadekwatne. Z tego powodu, podjęto próbę policzenia również odsetka wykorzystywania wskaźnika EWD w postaci, która w obrębie każdej fali badania opierała by się na deklaracjach respondentów co do jednego roku. W tym celu w tym rozdziale w części 5.1 dokładnie przeanalizowano odpowiedzi respondentów na pytania dotyczące wykorzystywania wskaźników EWD.

5.1. Odsetki wykorzystania EWD

Pytanie o wykorzystywanie wskaźników EWD zadane zostało w dosyć złożonej postaci. Po pierwsze, zadawane było oddzielnie odnośnie każdego roku: w badaniu z roku 2013 zadawane było odnośnie roku 2013, 2012 oraz zadano ogólniejsze pytanie o to, czy wykorzystywano wskaźniki EWD przed rokiem 2012; zaś w roku 2015 zapytano dodatkowo o wykorzystywanie wskaźników EWD w roku 2014. Co więcej respondenci udzielali też odpowiedzi, iż nie orientują się co do tego, czy w szkole korzystano w określonym roku ze wskaźników. Pewnym problemem jest również fakt, że I fala badania odbywała się w październiku 2013 roku, a wobec tego wielu dyrektorów wskazywało, że w roku 2013 jeszcze w ich szkołach nie prowadzono analiz efektywności nauczania.

Warto też zwrócić uwagę, że pytanie o wykorzystywanie EWD w szkołach zadawano tylko wtedy, jeżeli dyrektor stwierdził, że nie orientuje się czym są wskaźniki EWD. Na szczęście odpowiedzi takiej udzieliło bardzo mało dyrektorów – w badaniu z roku 2013 było ich 4, zaś w roku 2015 – 3. Uznano, że jeżeli dyrektor nie wie czym jest EWD, to w szkole wskaźniki EWD nie mogą być wykorzystywane do oceny efektywności nauczania. Z tego powodu, w dalszych analizach uznano, że w szkołach prowadzonych przez tych respondentów nie wykorzystuje się wskaźników EWD.

Tabela 15 prezentuje odsetki respondentów deklarujących korzystanie ze wskaźników EWD w określonym roku w I i II fali badania. Jak widać, rozbieżności między deklaracjami dotyczącymi poszczególnych lat są znaczące, są one również istotne statystycznie. Dla pytania o wykorzystywanie wskaźników przed rokiem 2012 różnica między badaniem z 2013 i 2015 roku jest znacząca – wynosi aż 17%. Dla roku 2012 jest nieco mniejsza – wyniosła 7%. Dla roku 2013 jest bardzo duża – aż 40%.

Tabela 15. Wykorzystanie wskaźników EWD według badań z lat 2013 i 2015 (odsetek odpowiedzi „tak”).

Zmienna	I fala	II fala	t	Wartość p
Wykorzystanie EWD przed 2012 rokiem	64,1%	47,0%	8,245	<0,001
Wykorzystanie EWD w 2012 roku	76,1%	69,1%	3,775	<0,001
Wykorzystanie EWD w 2013 roku	41,7%	81,1%	20,36	<0,001
Wykorzystanie EWD w 2014 roku	-	82,9%	-	-

Źródło: opracowanie własne na podstawie wyników badań ankietowych.

Jak wskazuje dalsza analiza wyników, duże rozbieżności mogą wynikać z udzielanych przez respondentów odpowiedzi innych niż „tak” i „nie”. Na pytania o wykorzystywanie wskaźników EWD możliwe były również odpowiedzi, iż respondent nie orientuje się co do tego, czy w szkole korzystano ze wskaźników, oraz w przypadku pytania dotyczącego roku 2013 z I fali badania możliwe było, że respondent wskazał, że w tym roku jeszcze nie prowadzono analiz efektywności nauczania. Odsetki innych odpowiedzi na tę serię pytań zaprezentowano w tabeli 16.

Dużą różnicę dla roku 2013 należy tłumaczyć faktem, że w przypadku dawniejszego badania pytanie dotyczyło roku, w którym odbywało się badanie, a zatem część dyrektorów wskazywała, że jeszcze w ich szkole nie prowadzono analiz efektywności nauczania, co w naturalny sposób obniżyło odsetek odpowiedzi twierdzących na to pytanie. Okazuje się, że był to bardzo znaczący odsetek – aż 38,8% dyrektorów wskazało tę odpowiedź. Sprawia to, że odsetek wykorzystywania wskaźników EWD w roku 2013 jest nieporównywalny z innymi latami z przyczyn metodologicznych – pytanie to w roku 2013 zostało zadane w zupełnie innych warunkach. Sprawia to, że porównanie między rokiem 2013 badania z I fali badania i rokiem 2014 z II fali badania byłoby nieuzasadnione i należy z niego zrezygnować.

Tabela 16. Wykorzystanie wskaźników EWD według badań z lat 2013 i 2015 (odsetek odpowiedzi „tak”).

	Nie orientuję się”		„w tym roku nie prowadziliśmy jeszcze analiz efektywności nauczania”	
	2013	2015	2013	2015
Wykorzystanie EWD przed 2012 rokiem	4,7%	17,4%	-	-
Wykorzystanie EWD w 2012 roku	2,5%	9,2%	-	-
Wykorzystanie EWD w 2013 roku	1,2%	5,2%	38,8%	-
Wykorzystanie EWD w 2014 roku	-	3,7%	-	-

Źródło: opracowanie własne na podstawie wyników badań ankietowych.

Należy zatem rozważyć możliwość porównania wyniku z 2014 roku z II fali badania z wynikami z roku 2012 i wcześniejszymi latami. W przypadku porównania z pytaniem o okres przed 2012 rokiem natrafiono na dwa problemy:

- Po pierwsze, pytanie dotyczy kilku lat, a zatem trudno uznać je za porównywalne z pytaniem o rok 2014 z II fali badania, które dotyczyło wykorzystywania wskaźnika EWD w czasie jednego roku.

- Po drugie, gdy zadano to samo pytanie ponownie w II fali badania, uzyskano odmienne odsetki odpowiedzi.

W odpowiedzi na pytanie o korzystanie ze wskaźników EWD przed rokiem 2012 w I fali badania uzyskano 64,1% odpowiedzi twierdzących, zaś w II fali – 47,0%. Prawdopodobnie należy tłumaczyć te różnice tym, że w II fali badania dyrektorzy gorzej pamiętali tak odległe w czasie fakty, skoro między I a II falą badania minęły prawie pełne dwa lata. Potwierdza to też większy odsetek odpowiedzi „nie orientuję się” w II fali badania (17,4% względem 4,7% w I fali). W tabeli 17 wskazano odsetki odpowiedzi twierdzących z wyłączeniem odpowiedzi „nie orientuję się” i „w tym roku jeszcze nie prowadzono analiz efektywności nauczania”. Niestety, nawet przy wyłączeniu odpowiedzi „nie orientuję się” w I i II fali badania uzyskuje się dalece różne odsetki odpowiedzi twierdzących (67,4% w I fali względem 56,9% w II).

Tabela 17. Wykorzystanie wskaźników EWD według badań z lat 2013 i 2015 (odsetek odpowiedzi „tak” z wyłączeniem odpowiedzi „nie orientuję się” oraz „analizy w tym roku jeszcze nie były prowadzone”).

Zmienna	2013	2015	t	Wartość p
Wykorzystanie EWD przed 2011 rokiem	67,4%	56,9%	4,905	<0,001
Wykorzystanie EWD w 2012 roku	78,3%	76,4%	1,056	0,291
Wykorzystanie EWD w 2013 roku	69,8%	85,9%	7,796	<0,001
Wykorzystanie EWD w 2014 roku	-	86,3%	-	-

Źródło: opracowanie własne na podstawie wyników badań ankietowych.

Pozostaje jeszcze rozważyć wykorzystanie pytania o rok 2012. W I fali badania uzyskano większe odsetki odpowiedzi twierdzących na to pytanie – 76,1% względem 69,1%. Jednak po wyłączeniu respondentów, którzy w II fali badania zadeklarowali, że „nie orientuję się” odsetki są już bardzo zbliżone wynosząc 78,3% w I fali badania oraz 76,4% w II. Jest to również jedyne porównanie odsetków dla pytania o ten sam rok między I i II falą, które daje nieistotny statystycznie wynik ($t=1,056$; $p=0,291$), co znaczy, że różnica odsetków jest niewielka. Co więcej, problem wysokiego odsetka respondentów, którzy udzielili odpowiedzi „nie orientuję się” dotyczy tylko II fali badania (9,2% w II fali, 2,5% w I). Świadczy to o tym, że wynik z fali I dotyczący roku 2012 można uznać za porównywalny z wynikiem z roku 2014 z fali II. Warto pamiętać, że pomiar dla roku 2012 z fali I jest o tyle doskonalszy od pomiaru z fali II, że respondenci w I fali odpowiadając na pytanie odnosili się do faktów mniej odległych w czasie. Fakt ten znajduje również odzwierciedlenie w mniejszym odsetku odpowiedzi „nie orientuję się” w I fali badania.

5.2. Porównanie wskaźników wykorzystania EWD dla lat 2013 i 2015

Rozważania z wcześniejszego rozdziału skłaniają do wniosku, że jeżeli chodzi o porównywanie wykorzystywania wskaźników EWD w roku 2014 według wyników z II fali z danymi z innego roku z fali I, to adekwatne może być tylko porównanie z wynikiem z roku 2012. Jednocześnie, warto porównać wartość wskaźnika, tak jak ją zdefiniowano w raporcie z ewaluacji, która miała miejsce w 2013 roku, czyli poprzez porównanie odsetka dyrektorów deklarujących, że w ich szkole przynajmniej w jednym z wymienionych lat korzystano ze wskaźnika EWD. W związku z tym, dokonane zostaną dwa porównania – każde według innej operacjonalizacji wskaźnika wykorzystywania EWD w szkołach:

- I. Porównanie odsetka szkół wykorzystujących kiedykolwiek wskaźniki EWD między I i II falą badania.

II. Porównanie wyniku z roku 2012 z fali I z wynikiem z roku 2014 z fali II.

Zaletą porównania według operacjonalizacji I na pewno jest to, że pierwszy punkt pomiaru lepiej odzwierciedla koniec wcześniejszego projektu i początek drugiego. Dzięki temu pomiar według operacjonalizacji pomiar jest bardziej adekwatny ze względu na zastosowany schemat. Porównanie według operacjonalizacji II w istocie rzeczy dotyczy raczej lat 2012-2014. Z drugiej strony, operacjonalizacja I może zawyżać uzyskane wartości, gdyż za wykorzystywanie wskaźników EWD uznaje się według niej również sytuację, gdy w szkole korzystano ze wskaźników w roku np. 2012 nawet jeżeli później tego zaprzestano.

W obu operacjonalizacjach drugi punkt pomiaru w istocie dotyczy końca roku 2014, czy też roku szkolnego 2014-15, a zatem nie obejmuje części programu realizowanej w drugiej połowie 2015 roku i wykorzystania uzyskanych wtedy wskaźników EWD. Wynika to oczywiście z braku pytań dotyczących 2015 roku w ankiecie z II fali badania.

Za główną podstawę oceny poziomu wykorzystania wskaźników EWD pod koniec analizowanego programu roku i jego efektu obrana została I operacjonalizacja, a operacjonalizacja II wykorzystana zostanie jako pomiar pomocniczy mający na celu walidację wyniku uzyskanego za pomocą głównego pomiaru. Jeżeli wartość pomiaru pomocniczego jest znacząco niższa niż głównego, to należy mieć na względzie, że wartość uzyskana za pomocą I operacjonalizacji może być nieznacznie zawyżona. Dla obu operacjonalizacji wskaźników wyznaczone zostaną również przedziały ufności. Za podstawowe źródło na temat wartości wskaźnika powinny zostać uznane przedziały ufności wyliczone według I operacjonalizacji.

5.2.1. Porównanie przy operacjonalizacji I

Porównanie wskaźnika wykorzystywania wskaźników EWD w I operacjonalizacji daje wynik istotny statystycznie. Szczegóły zaprezentowano w tabeli 18. Średnia w I fali badania wyniosła 85,1%, zaś w drugiej 90,5%. Daje to 5,4% postępu między rokiem 2013 i 2014. Nie jest to wynik imponujący, jednak nie należy oczekiwać aby kiedykolwiek odsetek szkół wykorzystujących wskaźniki EWD wyniósł 100%, a ponadto przy wartościach bliskich 100% ulega on jakby nasyceniu, co znaczy, że im wskaźnik jest bliższy 100%, tym nieproporcjonalnie więcej pracy i przekonania coraz bardziej niechętnych wymaga jego zwiększenie. Należy pamiętać, że wynik może być nieco zawyżony, gdyż jak dowiedziono w rozdziale 4, połączona próba trzech rodzajów szkół może być obciążona z powodu nadreprezentacji liceów o wysokim poziomie i efektywności nauczania.

Tabela 18. Porównanie wskaźnika korzystania z EWD w fali I i II.

	Średnia	Bł. std.
Odsetek szkół korzystających przynajmniej raz w latach 2011-2013 ze wskaźników EWD według fali I	85,1%	0,010
Odsetek szkół korzystających przynajmniej raz w latach 2011-2014 ze wskaźników EWD według fali II	90,5%	0,009
	Statystyka t	Wartość p
Wynik testu statystycznego	4,137	<0,001

Źródło: opracowanie własne na podstawie wyników badań ankietowych.

W tabeli 19 zaprezentowano wyniki testów statystycznych na różnicę odsetka szkół korzystających ze wskaźników EWD dla ogółu szkół i trzech badanych typów szkół. Dodatkowo w tabeli ujęto

oszacowania punktowe średnich oraz krańce przedziałów ufności dla każdego typu szkoły oraz dla łącznej zbiorowości szkół.

Według oszacowań wskaźnik ma największą wartość dla gimnazjów (90,9% w I fali i 92,2% w II), lecz jednocześnie uzyskano dla nich najniższą poprawę między dwoma punktami pomiaru (wzrost wykorzystania wyniósł tylko 1,3%). Poziom p testu wyniósł aż 0,492, należy zatem uznać, że różnica jest przypadkowa. W przypadku liceów wzrost wyniósł 8,9% uzyskując wartość 89,6% w II fali badania, zaś w przypadku techników – 15,7% uzyskując poziom 85,3%. Różnice dla liceów i techników są istotne statystycznie. Dokładne wyniki oraz przedziały ufności prezentuje tabela 19. Warto pamiętać, że wedle wcześniejszych ustaleń, uzyskany wynik dla liceów może być zbyt optymistyczny.

Niższą poprawę w gimnazjach należy tłumaczyć wspomnianym efektem nasycenia. Udało się natomiast poczynić duże postępy w liceach i technikach. Jednak na globalny wskaźnik (liczony dla ogółu szkół) największy wpływ mają właśnie gimnazja, gdyż w zbiorowości liceów, techników i gimnazjów stanowią one 63,3% placówek, dla których dostępne są wskaźniki EWD (wskaźniki są również dostępne dla niektórych podstawówek, ale zostały one pominięte, gdyż nie są przedmiotem tego badania, ich liczba nie została uwzględniona przy wyliczaniu wspomnianego odsetka).

Tabela 19. Porównanie wskaźnika korzystania z EWD z badań w latach 2013 i 2015.

	badanie	2,5 %	osz. pkt.	97,5 %	Statystyka t	Wartość p
Razem	2013	83,2%	85,1%	87,0%	4,137	<0,001
	2015	88,8%	90,5%	92,1%		
Gimnazjum	2013	88,3%	90,9%	93,6%	0,687	0,492
	2015	89,8%	92,2%	94,6%		
Liceum	2013	77,2%	80,7%	84,3%	3,923	<0,001
	2015	86,9%	89,6%	92,2%		
Technikum	2013	65,6%	69,6%	73,5%	6,243	<0,001
	2015	82,3%	85,3%	88,2%		

Źródło: opracowanie własne na podstawie wyników badań ankietowych.

5.2.2. Porównanie przy operacjonalizacji II

Wskaźniki liczone na drugi sposób są naturalnie niższe. Wynika to z faktu, że do wskaźnika w wersji drugiej wliczono tylko szkoły, które skorzystały ze wskaźników EWD w jednym konkretnym roku, a nie w dowolnym z tych, o które zadawano pytanie. Według II operacjonalizacji poziom wykorzystania wskaźników EWD w roku 2014 wyniósł 82,8%, co oznacza postęp o 7,9% od roku 2012. Wartość ta jest istotna statycznie. Również tutaj warto zwrócić uwagę, że wynik ten może być zbyt optymistyczny z powodu obciążenia próby. Dla pogłębienia analizy przeprowadzono osobne porównanie dla każdego typu szkół.

Tabela 20. Porównanie wskaźnika korzystania z EWD z badań w latach 2013 i 2015.

	Średnia	Bł. std.
Odsetek szkół korzystających ze wskaźników EWD w 2012 roku w I fali	76,1%	0,013
Odsetek szkół korzystających ze wskaźników EWD w 2014 roku w II fali	82,8%	0,012
	Statystyka t	Wartość p
Wynik testu statystycznego	3,979	<0,001

Źródło: opracowanie własne na podstawie wyników badań ankietowych.

Tabela 21 prezentuje porównanie wyników na poziomie typów szkół. Uzyskano podobne rezultaty jak w przypadku I operacjonalizacji wskaźnika: najwyższy poziom wykorzystywania wskaźników EWD dla 2012 roku odnotowano dla gimnazjów – 80,9%. Dla liceów i techników odnotowane poziomy były niższe (72,5% oraz 63,4%), jednocześnie dla tych placówek odnotowano wyższe wzrosty. W liceach wzrost wskaźnika można oszacować na 11,6 p.p., zaś w technikach o 12,7 p.p. Są to wartości istotne statystycznie. Inaczej jest w przypadku gimnazjów: uzyskana poprawa to 3,5p.p. Jest to poprawa tylko nieco wyższa niż w przypadku I operacjonalizacji, i ponownie jest ona nieistotna statystycznie. Oznacza to, że uzyskany efekt w gimnazjach jest tak mały, że może wynikać z przypadkowego wylosowania do próby większej liczby szkół korzystających ze wskaźników (prawdopodobieństwo, że wynika to z losowania wyniosło 0,166). Należy przyjąć, że wskaźnik wykorzystania wskaźników EWD w gimnazjach nie wzrósł, choć prawdopodobne wydaje się przypuszczenie, że po prostu wzrósł, lecz nieznacznie. Niemniej jednak pozostaje to tylko przypuszczeniem. Do oszacowań uzyskanych w tym rozdziale w dalszych częściach raportu autor odnosić się będzie, jako do wariantu podstawowego.

Tabela 21. Porównanie wskaźnika korzystania z EWD w latach 2012 (z fali I) i 2014 (z fali II).

	badanie	2,5 %	osz. pkt.	97,5 %	Statystyka t	Wartość p
Razem	fala I	73,6%	76,1%	78,6%	3,979	<0,001
	fala II	80,6%	82,9%	85,1%		
Gimnazjum	fala I	77,3%	80,9%	84,5%	1,384	0,166
	fala II	81,1%	84,4%	87,7%		
Liceum	fala I	68,5%	72,5%	76,5%	4,455	<0,001
	fala II	80,9%	84,1%	87,3%		
Technikum	fala I	59,2%	63,4%	67,5%	4,569	<0,001
	fala II	72,5%	76,1%	79,7%		

Źródło: opracowanie własne na podstawie wyników badań ankietowych.

5.2.3. Przedziały ufności z uwzględnieniem braków udziału

W rozdziale oszacowane zostaną jeszcze raz powyższe wartości, ale tylko na podstawie próby podstawowej i z uwzględnieniem szkół, które nie zgodziły się wziąć udziału w badaniu. W sytuacji, w której znacząca część badanej populacji okazała się niedostępna dla badania można podjąć próbę oszacowania tzw. pseudo-przedziału ufności. Procedura ta opisana jest w literaturze⁷. Uwzględnia ona przy wnioskowaniu statystycznym niepewność związaną z brakiem wiedzy na temat wartości analizowanych zmiennych dla części wylosowanej próby. Kluczowe dla jej wykonania było oszacowanie efektywnego odsetka realizacji próby, które zostało opisane w rozdziale 3.3. Metoda ta zazwyczaj daje dużo szersze przedziały ufności. Ich rozmiar naturalnie zależy od frakcji jednostek niedostępnych oszacowanej na podstawie badania. Odnotowane frakcje wyniosły one dla poszczególnych fal badania i typów placówek od 18 do 26% (por. tabela 6). W praktyce oznacza to poszerzenie szerokości przedziału ufności w każdą stronę o połowę tego odsetka. Dodatkowo im większa frakcja jednostek niedostępnych, tym oszacowanie punktowe bardziej przesuwają się w kierunku wartości 0,5.

Analiza ta jest nawiązaniem do analiz, które zostały wykonane w *Raport z ewaluacji wewnętrznej projektu EWD*. Uznano, że warto aby oszacowane dla I fali badania wskaźniki oszacować również dla

⁷ Lissowski G., Problem jednostek niedostępnych w reprezentacyjnych badaniach socjologicznych, W: Szaniawski K. (red.), Metody matematyczne w socjologii, Ossolineum 1971.

fali II. Interpretować je można jako wartości wskaźników wykorzystania EWD w pesymistycznym wariancie, czyli w takim gdy wszystkich szkołach z próby zasadniczej, w których nie udało się zrealizować badania, nie wykorzystuje się wskaźników EWD.

Tabela 22 prezentuje oszacowania wskaźników EWD oraz przedziały ufności w jakich mieszczą się one z prawdopodobieństwem 95% przy uwzględnieniu wyłącznie próby zasadniczej i uznaniu, że nie wiemy niczego o poziomie wykorzystania wskaźników EWD w niezrealizowanej części próby zasadniczej. Najistotniejsza informacja, jaką można zaczerpnąć z tej tabeli, to dolna krawędź przedziału ufności. Należy ją interpretować jako wartość I operacjonalizacji wskaźnika wykorzystania wskaźników EWD w możliwe pesymistycznym wariancie, tj. w takim, w których wszystkie szkoły z próby zasadniczej, do których nie udało się dotrzeć z procedurą badania nie wykorzystują wskaźników EWD.

Warto zwrócić uwagę, że na podstawie tych tabel pozornie można by wyciągnąć wniosek, że poprawa między latami 2013 i 2014 w najbardziej pesymistycznym wariancie jest większa, niż w wariancie podstawowym. Z porównania wartości 55,6% z wartością 65,7% w 3 kolumnie tabeli efekt, jakby się wydawało, wyniósłby prawie 10,0%. Byłby to jednak wniosek niesłuszny, a większa wartość tej różnicy wynika z większej niepewności z jaką mamy do czynienia w przypadku badania z roku 2013, gdzie zarejestrowano większy odsetek braków udziału. Oszacowanie efektu programu odbywać się powinno raczej na podstawie wariantu podstawowego, a w ostateczności na podstawie 4 kolumny tabeli określającej oszacowania punktowe wskaźnika. W tym przypadku efekt wyniósł by 6,2%, czyli niewiele więcej niż w wariancie podstawowym.

Tabela 22. Porównanie I wskaźnika wykorzystania EWD z badań w latach 2013 i 2015 z przedziałami ufności w wariancie pesymistycznym (z uwzględnieniem systematycznych braków udziału).

	Fala	2,5 %	osz. pkt.	97,5 %
Razem	I	55,6%	70,1%	84,6%
	II	65,7%	76,3%	86,9%
Gimnazjum	I	58,5%	73,7%	89,0%
	II	63,2%	76,2%	89,3%
Liceum	I	55,6%	69,7%	83,7%
	II	65,6%	77,3%	88,9%
Technikum	I	44,6%	60,0%	75,4%
	II	58,7%	70,9%	83,2%

Źródło: opracowanie własne na podstawie wyników badań ankietowych.

Tabela 23 zawiera te same informacje dla oszacowania wskaźnika wykorzystania EWD w II operacjonalizacji.

Tabela 23. Porównanie II wskaźnika korzystania z EWD z badań w latach 2013 i 2015 z przedziałami ufności w wariancie pesymistycznym (z uwzględnieniem systematycznych braków udziału).

	Fala	2,5 %	osz. pkt.	97,5 %
Razem	I	62,4%	76,4%	90,4%
	II	72,8%	82,9%	92,9%
Gimnazjum	I	66,2%	80,6%	95,0%
	II	70,7%	82,9%	95,2%
Liceum	I	62,2%	75,8%	89,3%
	II	70,2%	81,4%	92,6%
Technikum	I	49,9%	65,1%	80,3%
	II	66,9%	78,6%	90,3%

Źródło: opracowanie własne na podstawie wyników badań ankietowych.

6. Wnioski

Analizy prób zbadanych w roku 2013 i 2015 skłaniają do wniosku, że są one reprezentatywne i pozwalają ocenić odsetek szkół wykorzystujących wskaźniki EWD w sposób adekwatny z wykorzystaniem klasycznych technik statystycznych. Wyjątkiem są licea, dla których zachodzi wysokie ryzyko obciążenia wyników w badaniu z roku 2013. Z tego powodu, do oszacowań dla liceów z tego roku stosowane będzie bardziej ostrożne podejście. Fakt obciążenia wyników uzyskanych dla liceów w I fali badania oznacza również obciążenie wyników uzyskanych na połączonej próbie gimnazjów, liceów i techników.

Efekt programu *Rozwój metody edukacyjnej wartości dodanej na potrzeby wzmocnienia ewaluacyjnej funkcji egzaminów zewnętrznych* w postaci upowszechnienia wykorzystania wskaźników EWD w gimnazjach, liceach i technikach oszacowano w dwóch operacjonalizacjach. Wersja pierwsza powinna być traktowana jako podstawowa, zaś druga (poniżej wskazywana jako pomiar kontrolny) jako sugestia jakiej wielkości wskaźnik można uzyskać przy innych rozstrzygnięciach metodologicznych. Uzyskany rozmiar efektu to:

- 5,4% łącznie dla trzech rodzajów szkół (pomiar kontrolny – 6,8%, pomiar może być obciążony),
- 1,3% dla gimnazjów (pomiar kontrolny – 3,5%, różnice nieistotne statystycznie, efekt zatem nie nastąpił lub był bardzo mały),
- 8,9% dla liceów (pomiar kontrolny – 11,6%, pomiar może być obciążony),
- 15,7% dla techników (pomiar kontrolny – 12,5%).

Oszacowanie punktowe wskaźnika wykorzystania EWD w roku 2014 dla to:

- 90,5% łącznie dla trzech rodzajów szkół (pomiar kontrolny – 82,9%, pomiar może być obciążony),
- 92,2% dla gimnazjów (pomiar kontrolny – 84,4%),
- 89,6% dla liceów (pomiar kontrolny – 84,1%, analizy wskazują pomiar może być obciążony, dlatego należy brać pod uwagę, że wartości te mogą być nawet tak niskie jak odpowiednio 65,6% oraz 70,2%),
- 85,3% dla techników (pomiar kontrolny – 76,1%).

W raporcie w części 5.2 przedstawiono również dokładniejsze wartości uwzględniające przedziały ufności. Dodatkowo w części 5.3 zawarto wartości i oszacowania przedziałów ufności bardzo rygorystyczne pod względem statystycznym i uwzględniające najbardziej pesymistyczny scenariusz, mówiący, że szkoły w których nie udało się zrealizować badania nie wykorzystują wskaźników EWD do oceny efektywności nauczania w szkole.