

Ewaluacja wobec wyzwań
stojących przed sektorem
finansów publicznych

Ewaluacja wobec wyzwań stojących przed sektorem finansów publicznych

redakcja

Agnieszka Haber

Maciej Szałaj

Polska Agencja Rozwoju Przedsiębiorczości
Warszawa 2009

Ewaluacja wobec wyzwań stojących przed sektorem finansów publicznych
pod redakcją Agnieszki Haber i Macieja Szałaja

Poglądy autorów tekstów zawartych w publikacji są wyrazem ich własnych opinii i nie odzwierciedlają oficjalnego stanowiska PARP.

© Copyright by Polska Agencja Rozwoju Przedsiębiorczości 2009

ISBN 978-83-7633-064-8

Wydanie I

Nakład 650 egzemplarzy

Opracowanie wydawnicze:
Jacek Pacholec, Joanna Iwanowska

Przygotowanie do druku, druk i oprawa:



Wydawnictwo Naukowe Instytutu Technologii Eksploatacji – PIB
ul. K. Pułaskiego 6/10, 26-600 Radom, tel. centr. (48) 364-42-41, fax (48) 3644765
e-mail: instytut@itee.radom.pl <http://www.itee.radom.pl>

Spis treści

Wprowadzenie	7
---------------------------	----------

Część I. Ewaluacja w zarządzaniu finansami publicznymi

Jarosław Górniak

Ewaluacja jako instrument współczesnej polityki gospodarczej	13
---	-----------

Wojciech Misiąg

Ewaluacja wobec wyzwań stojących przed sektorem finansów publicznych	27
---	-----------

Marta Postuła

Rola ewaluacji w budżetowaniu zadaniowym	41
---	-----------

Część II. Ewaluacja a stanowienie prawa

Agnieszka Sternik

Ocena Skutków Regulacji – metody i podejścia na przykładzie systemu oceny skutków Komisji Europejskiej	61
---	-----------

Artur Nowak-Far

Ocena Skutków Regulacji (OSR) – regulacja instrumentem polityki	77
--	-----------

Część III. Metodologia badań ewaluacyjnych – wokół jakości i technologii

Jadwiga Przewłocka

CAWI – specyfika, wykorzystanie, perspektywy rozwoju	97
---	-----------

Jacek Pokorski

Zastosowanie techniki CAWI w badaniach ewaluacyjnych	109
---	------------

Jan M. Zajęc, Dominik Batorski

Metody zwiększania zwrotności w badaniach internetowych	129
--	------------

Franciszek Sztabiński

Ankieta internetowa (CAWI): czy rzeczywiście idealna technika?	149
---	------------

Agnieszka Haber, Mirosław Kalinowski

Metody kontroli w badaniach ewaluacyjnych	173
--	------------

Jarosław Górniak

Efekt przyczynowy w ewaluacji	189
--	------------

Szanowni Państwo,

Ewaluacja obecna jest w praktyce polskiej administracji publicznej już od blisko dekady. Jej początki lokują się więc w podobnym okresie, co powstanie Polskiej Agencji Rozwoju Przedsiębiorczości, która w następnym roku będzie obchodziła dziesięciolecie działalności. PARP była jedną z pierwszych instytucji, jakie podjęły się systematycznej realizacji badań ewaluacyjnych, tworząc w tym celu wyspecjalizowaną komórkę oraz konsekwentnie rozwijając stosowany warsztat badawczy.

Od momentu przystąpienia Polski do struktur Unii Europejskiej nastąpił niezwykle gwałtowny rozwój ewaluacji związany w dużej mierze formalnym wymogiem przeprowadzania oceny. Obecnie możemy mówić z całą pewnością o ukształtowanym systemie oceny, który tworzą, obok instytucji centralnych i regionalnych, różnorodni partnerzy społeczni (uczelnie, organizacje gospodarcze, instytucje społeczeństwa obywatelskiego) oraz poszerzający się stale krąg ekspertów. Na tle innych krajów, Polska wyróżnia się zarówno poziomem prowadzonych badań ewaluacyjnych, jak również dynamiką lokalnego środowiska oraz intensywnością działań na polu tworzenia potencjału ewaluacyjnego.

Doświadczenia związane z ewaluacją przedsięwzięć finansowanych ze środków europejskich są już na tyle bogate, że pożądanym jest ich przeniesienie na obszar krajowych środków publicznych, gdzie istnieje paląca potrzeba unowocześnienia stosowanych mechanizmów zarządzania publicznego oraz poprawy jakości wydatkowania. Należy podkreślić, iż podejmowane są w tym celu działania legislacyjne rozszerzające zakres obligatoryjnego prowadzenia ewaluacji na krajowe działania prorozwojowe. Działaniom tym powinna towarzyszyć debata angażująca społeczność tworzącą polski system ewaluacji. Chcielibyśmy, aby niniejsza publikacja stała się przyczynkiem do tego rodzaju debaty.

Pierwszą część tomu, nawiązującą bezpośrednio do tytułu, tworzą rozdziały poświęcone refleksji nad wykorzystaniem ewaluacji w celu poprawy najbardziej elementarnych funkcji państwa – gospodarowania publicznymi funduszami oraz stanowienia prawa. Jakość realizacji tych funkcji rzutuje na życie każdego z nas jako podatnika, oddającego istotną część swych dochodów z nadzieją na ich właściwe spożytkowanie oraz jako obywatela oczekującego, iż prawodawcy, którym powierza rządy, zapewnią mu warunki dla bezpiecznej i komfortowej realizacji prywatnych zamierzeń.

Druga część publikacji jest odpowiedzią na oczekiwania wydawnicze zgłaszane przy różnorodnych okazjach przez przedstawicieli polskiej społeczności ewaluacyjnej, którzy wskazują na potrzebę rozwoju swojego warsztatu badawczego. Podobnie więc, jak w przypadku ostatnich dwóch tomów publikowanych przez PARP w ramach serii ewaluacyjnej, jeden z rozdziałów poświęciliśmy tematyce metodologicznej. Tym razem koncentrujemy się na dwóch wiodących wątkach – budowaniu jakości w badaniach ewaluacyjnych oraz zastosowaniu bardzo dynamicznie się rozwijających badań internetowych.

Mamy nadzieję, iż książka, którą przekazujemy Państwu, spotka się z Państwa zainteresowaniem i dostarczy sposobności do inspirującej lektury.

Bożena Lublińska-Kasprzak

Prezes Polskiej Agencji Rozwoju Przedsiębiorczości

Część I

**Ewaluacja w zarządzaniu
finansami publicznymi**

Najważniejszą nadzieją związaną z ewaluacją od jej samego początku było i jest do dziś oczekiwanie, iż będzie pełniła ona rolę narzędzia zapewniającego racjonalne i oszczędne gospodarowanie funduszami publicznymi. Postulat ekonomizacji funkcjonowania sektora publicznego stanowił także fundament programów reform, będących odpowiedzią na kryzysy gospodarcze minionego wieku. Szczególną rolę odegrał on na fali kryzysu lat siedemdziesiątych. W toku toczącej się wówczas debaty ukształtowana została koncepcja nowego zarządzania publicznego (New Public Management), zakładająca, iż państwo, podobnie jak organizacje komercyjne, powinno kierować się logiką rynkową, której immanentnym elementem pozostaje dbałość o finansową opłacalność prowadzonych działań.

Konsekwencją przyjętych założeń stała się potrzeba zrewidowania całościowego paradygmatu zarządzania finansami publicznymi, w tym zasad alokowania i wydatkowania zasobów oraz rozliczania dysponentów środków. W opozycji do klasycznej formuły budżetowej opartej o sztywną strukturę wydatków, przedstawiona została koncepcja budżetowania zadaniowego. Jej fundamenty oparte zostały na powiązaniu wielkości alokowanych środków ze specyfiką i skalą działań przydzielanych w określonej perspektywie czasowej oraz odzwierciedlały aktualne priorytety strategiczne.

Próby wdrożenia budżetowania zadaniowego podejmowano dotychczas, z różnym powodzeniem, w wielu krajach. Od kilku lat, w ramach reformy finansów publicznych, podejmowane są one również w Polsce. Skuteczne osadzenie instytucjonalne procesów budżetowania zadaniowego wymaga przygotowania odpowiedniego zaplecza informacyjnego, dostarczającego podstaw do podejmowania trafnych decyzji alokacyjnych oraz do oceny sposobu wydatkowania zasobów. Potrzeba ta stanowi uzasadnienie zestawienia problematyki budżetu zadaniowego z ewaluacją, którą należy w tym kontekście postrzegać jako unikalne źródło informacji o efektach działań instytucji publicznych.

Ewaluacja i budżet zadaniowy stanowią atrybuty nowoczesnego państwa i jako takie są instrumentami wzajemnie się uzupełniającymi. Ich konsekwentne wdrażanie i odpowiednia koordynacja pozwoliłyby spełnić pokładane w nich nadzieje, co z kolei znalazłoby wymierne odzwierciedlenie w jakości funkcjonowania instytucji publicznych i ich oddziaływania na życie obywateli.

Szerokie pole dla zastosowania ewaluacji w zarządzaniu finansami publicznymi nie ogranicza się jednak wyłącznie do procesów budżetowania zadaniowego. Badania ewaluacyjne odgrywać mogą zróżnicowaną rolę w działaniach wychodzących naprzeciw głównym wyzwaniom budżetowym, z jakimi boryka się Polska od początku okresu transformacji, a które stanowią jedną z najpoważniejszych barier rozwoju. Wśród czynników najbardziej dotkliwie obciążających krajowe finanse publiczne wskazać należy nadmierny udział wydatków sztywnych (obsługa zadłużenia, zobowiązania socjalne), czy też niedofinansowanie obszarów kluczowych z punktu widzenia globalnej konkurencyjności polskiej gospodarki (edukacja, szkolnictwo wyższe, badania naukowe).

Ewaluacja, jako wspomniany atrybut nowoczesnego państwa, stanowi element zarządzania oparteego na wiedzy, a więc modelu alternatywnego wobec tzw. „Polski resortowej”, czyli nieskoordynowanej i podporządkowanej rozmaitym partykularyzmom organizacji państwa. Zarządzanie oparte na wiedzy jest koncepcją cieszącą się ostatnio dużą popularnością. Jego wartość pozostaje jednak wątpliwa, póki będzie ono jedynie katalogiem górnolotnych frazesów. Właściwa wartość tego podejścia może być potwierdzona jedynie jego praktycznym zastosowaniem. Niezbędny element tej praktyki stanowią badania ewaluacyjne, o ile tylko

stosowane są świadomie i zgodnie z profesjonalnymi kanonami. Szerokie zastosowanie zarządzania bazującego na wiedzy w sektorze publicznym byłoby wehikułem swoistej „rewolucji”, na fali której państwo podporządkowane zostałoby zupełnie nowemu kodowi funkcjonowania. Można powiedzieć, iż państwo zostałoby wówczas „przeprogramowane” z orientacji „wegetowanej” (możolne domykanie budżetu, balansowanie na granicy akceptowalnych pułapów zadłużenia) na „rozwojową” (planowanie wieloletnie, skoncentrowanie wydatków na działaniach priorytetowych w perspektywie strategicznej). Wkład, jaki ewaluacja może wnieść w tworzenie nowego kodu funkcjonowania państwa, określa jej rolę wobec współczesnych wyzwań stojących przed polskim sektorem finansów publicznych.

Ewaluacja jako instrument współczesnej polityki gospodarczej

„Współczesne rządy znajdują się pod presją wynikającą z konieczności oceniania swoich działań. Informacje o jakości działania sektora publicznego służące poinformowaniu opinii publicznej mogą być również wykorzystane dla wykazania, iż rząd poprzez swoje działania zapewnia odpowiedniej jakości usługi, wykorzystując efektywnie dostępne środki finansowe. Najistotniejsze jest w tym chyba to, że informacje o jakości działania mogą pomóc decydom publicznym konstruować lepszy budżet i podejmować lepsze decyzje”.

Globalny kryzys po długiej fali wzrostu gospodarczego rozbudził debatę na temat roli państwa w gospodarce. Dość zgodny chór głosów wskazuje na konieczność wzmocnienia funkcji regulacyjnych w stosunku do sektora finansowego, wykraczających poza granice poszczególnych państw. Rządy w wielu krajach, z USA na czele, zaangażowały się w konstruowane ad hoc przedsięwzięcia antykryzysowe, przeznaczając na ten cel ogromne środki budżetowe i pogłębiając deficyty finansów publicznych i tak już nadwyrężonych przez kurczące się na fali kryzysu przychody budżetu. W Polsce, mimo wykazywanej przez rząd wstrzemięźliwości wobec antycyklicznego alimentowania gospodarki i ostrożnej polityki budżetowej, deficyt pogłębia się na skutek spadku przychodów budżetu, ale i niemożności ograniczenia wydatków. Sytuacje nadzwyczajne odwracają uwagę od bieżących, „pozytywistycznych” działań zmierzających do racjonalizacji funkcjonowania sektora finansów publicznych i poddania go lepszej obywatelskiej kontroli, jego stopniowej, systemowej ewolucji. Pojawienie się dużych interwencji publicznych zorientowanych na sytuacje kryzysowe pociąga za sobą zainteresowanie społeczne oceną efektów takich działań, ale ma to charakter kampanijny, pozostający poza głównym nurtem bieżącej polityki budżetowej.

W tej atmosferze zdarzeń dla gospodarki nadzwyczajnych bez specjalnych komentarzy przeszły ostatnie zmiany *Ustawy o finansach publicznych*, tworzące warunki do aktywniejszego wprowadzenia zasad budżetu zadaniowego do jednostek sektora finansów publicznych. Jednak kurz wzbity przez globalny kryzys powoli opada. Warto zastanowić się nad tym, jakie są kluczowe korzyści, które funkcjonowanie finansów publicznych może wyciągnąć z dobrego wprowadzenia nowych zasad i gdzie kryją się potencjalne pułapki. Tezą tego artykułu jest twierdzenie, że kluczową korzyścią z wprowadzenia zasad budżetowania zadaniowego może się stać przede wszystkim silniejsze wprowadzenie do systemu finansów publicznych orientacji na cele oraz zasad ewaluacji programów publicznych.

Ewaluacja zdaje się być czymś oczywistym dla każdego działania rozpatrywanego z perspektywy racjonalnego podmiotu działającego. Z prakseologicznego punktu widzenia, działanie zmierza ku celowi – zamierzonego stanu rzeczy w przyszłości – który chce osiągnąć działający. Jest więc czymś oczywistym sprawdzenie stopnia osiągnięcia celu i w razie sukcesu utwierdzenie się w słuszności dokonanego wyboru środków, a w razie porażki namysł nad jej powodami. Tyle model. W rzeczywistości, zwłaszcza organi-

¹ *Performance Budgeting in OECD Countries*, OECD 2007.

zacyjnej, a już szczególnie w przypadku opartej na procedurach administracji, formalnie wyodrębniona ewaluacja nie jest tak oczywista.

W hierarchicznych, sformalizowanych, scentralizowanych organizacjach działania oparte są o procedury, do których przestrzegania zobowiązani są ich funkcjonariusze. Oczywiście, źródłem procedur jest namysł nad celami, przy czym na ogół chodzi tu o ogólne, „statutowe” cele organizacji, a także dążenie do zapewnienia ogólnej koordynacji i zabezpieczenia przed nieprawidłowościami. Kontrola prowadzonych działań koncentruje się na ich zgodności z procedurami. Na problemy w osiąganiu celów reaguje się korektą procedur. W stabilnym otoczeniu, gdzie stale, w sposób powtarzalny muszą być realizowane określone zadania, taka organizacja może okazać się sprawna. Mowa tu oczywiście o biurokracji. Max Weber uważał ją za ucieleśnienie procesu racjonalizacji życia społecznego. Choć wielu autorów konstatowało i analizowało niesprawności tego modelu², znany brytyjski badacz organizacji W. Bennis, nie bez racji, zauważał w latach sześćdziesiątych: „parafrazując ironiczną uwagę Churchilla na temat demokracji, można powiedzieć o biurokracji, że jest najgorszą z możliwych teorii organizacji, poza tymi, które dotychczas były wypróbowane”³.

Rosnąca dynamika zmian społecznych i gospodarczych, zmiany w technologii, presja na wzrost konkurencyjności państw i regionów stanowiących domenę odpowiedzialności władz publicznych, a także rozwój koncepcji zarządzania postawiły przed administracją publiczną wyzwania, którym próbowano sprostać poszukując nowych koncepcji zarządzania publicznego, sięgając między innymi do doświadczeń organizacji gospodarczych. W przeciwieństwie do przedsiębiorstw, administracja publiczna nie staje wobec naturalnego sprawdzianu, jakim dla podmiotów gospodarczych jest uzyskiwana pozycja w konkurencji rynkowej i wymiarność uzyskiwanych korzyści lub strat. Pojawia się w związku z tym problem bodźców do dostosowywania się do zmieniających się społecznych potrzeb i zwiększania efektywności wydatkowania środków publicznych. Reformatory administracji próbują się uporać z tymi problemami poprzez poszukiwanie sposobu na powiązanie wydatków publicznych z osiąganiem określonych celów. Z tego sposobu myślenia wyrósł zarówno zwrot w kierunku wykorzystania w zarządzaniu sferą publiczną koncepcji zarządzania programami składającymi się na określone polityki czy strategie, jak i poszukiwanie swoistego „kamienia filozoficznego” nowoczesnego rządzenia, którym miał być budżet zadaniowy – budżetowanie zorientowane na efekty (*performance budgeting*). Ewaluacja odgrywa swoją istotną rolę zarówno w cyklu programowania i implementacji polityk publicznych, jak i w przypadku budżetu zadaniowego, który nadaje programom i politykom publicznym całościowe ramy.

Budżet zadaniowy jest ideą dyskutowaną już od kilku dziesięcioleci, ale w zakresie prób implementacji przeżywająca rozkwit w rozwiniętych gospodarczo demokracjach od lat dziewięćdziesiątych. Podstawowa idea kryjąca się za hasłem budżet zadaniowy jest prosta: należy powiązać nakłady budżetowe z pomiarem uzyskiwanych efektów. Zwykle podkreśla się dobre strony tego rozwiązania, wynikające przede wszystkim z wprowadzenia definicji i pomiaru efektów, w tym zwłaszcza:

- przesunięcie uwagi z wydatków budżetowych i reguł ich wykorzystania na mierzalne efekty działań władz publicznych,
- ujęcie działań władz publicznych w struktury programów o jasno zdefiniowanych celach ujętych w kategoriach rezultatów – zmian wywołanych działaniami, a nie nakładów, czy nawet produktów – bezpośrednich wytworów działań,

² Por. np. przykładowe klasyczne teksty: R.K. Merton, *Teoria socjologiczna i struktura społeczna*, Rozdz. VIII „Struktura biurokratyczna i osobowość”, PWN, Warszawa 1982; M. Crozier, *Biurokracja: Anatomia zjawiska*, PWE, Warszawa 1967; J. Staniszkis, *Patologie struktur organizacyjnych*, ZN im. Ossolińskich, Wrocław – Warszawa – Kraków 1972.

³ W.G. Bennis, *Rozwój organizacji i losy biurokracji*, [w:] *Zachowanie człowieka w organizacji, wybór tekstów* pod red. W.E. Scotta Jr. i L.L. Cumminsa, t.1, PWN, Warszawa 1983, s. 440–456.

- wzrost skuteczności i efektywności realizowanych programów,
- alokacja zasobów na projekty charakteryzujące się większą efektywnością i wygaszanie wydatków nieprzynoszących widocznych efektów lub efektów uzasadniających poniesione koszty.

Duże znaczenie, warte odrębnego podkreślenia, ma zwiększenie przejrzystości polityk publicznych i wydatkowania środków budżetowych dla organów przedstawicielskich i dla szerokiej opinii publicznej, co zwiększa możliwości obywatelskiej kontroli decyzji w tym zakresie. Pomiar i analiza efektów uzyskiwanych przez programy publiczne sprzyja poprawie jakości debaty publicznej. Takie oczekiwania towarzyszyły utworzeniu w okresie administracji Georga W. Busha, w 2006 r., portalu ExpectMore.gov (<http://www.whitehouse.gov/omb/expectmore/>), którego celem jest prezentacja opinii publicznej wyników systematycznie prowadzonej oceny programów federalnych przy pomocy specjalnie opracowanej metodyki PART (*The Program Assessment Rating Tool*): wysoce wystandaryzowanej odmiany ewaluacji (czy, jak kto woli, audytu efektywnościowego). W PART zadaje się wobec każdego programu 25 podstawowych pytań, podzielonych na 4 sekcje. Pierwsza dotyczy konstrukcji programu i jego celów: czy są jasne, konkretne i uzasadnione. Druga część związana jest z planowaniem strategicznym i oceną poprawności formułowania rocznych i wieloletnich celów przez agencję. Trzecia sekcja odnosi się do zarządzania programem, w tym monitorowania finansowego i usprawnień. Część ostatnia dotyczy rezultatów programu, które można wykazać dokładnie i spójnie⁴. Wśród pytań zadawanych programom znajdujemy np. takie, jak:

- Czy program odnosi się do konkretnych i istniejących problemów, interesów lub potrzeb?
- Czy program ma ograniczoną liczbę konkretnych, długookresowych wskaźników, które koncentrują się na efektach i sensownie odzwierciedlają cele programu?
- Czy program (łącznie z partnerami programu) osiąga swoje roczne cele w zakresie efektów?
- Czy program wykazał stosowny postęp w osiąganiu swoich długofalowych celów w zakresie efektów?
- Czy program wykazuje poprawę efektywności kosztowej w osiąganiu celów programu w każdym roku?

Prezentacja oceny każdego programu dokonuje się w trzech kategoriach programów przynoszących efekty (najlepsze należą do kategorii skutecznych – *effective*) i dwóch kategoriach programów nieprzynoszących efektów (najgorsze należą do kategorii „rezultaty nie są widoczne”). Z funkcjonowaniem portalu wiąże się oczekiwania, że wzrośnie zainteresowanie opinii publicznej wynikami programów, zwiększy się presja na agencje w zakresie doskonalenia programów (w tym: jawność planów naprawczych i klarowność deklaracji dotyczących celów i efektów), wzrośnie zapotrzebowanie na aktualne i dobrej jakości dane dotyczące efektów⁵.

Oparcie budżetowania na wiedzy o efektach i ich związku z konstrukcją i przebiegiem programów publicznych stawia przed administracją publiczną nowe wyzwania. Większość administracji cierpi na problemy ze sprostaniem trudnościom w dostarczeniu decydentom informacji o dobrej jakości, adekwatnej i godnej zaufania, a do tego na czas, tak by mogła być wykorzystana w podejmowaniu decyzji budżetowych⁶. Rządy nie rezygnują jednak z nacisku na postępy w tym zakresie. Wymaga to rozwiązania szeregu problemów w dziedzinie pomiaru efektów, rozwoju potencjału instytucjonalnego ministerstwa finansów i ministerstw sektorowych, przezwyciężenia oporu administracji i zmian w kulturze organizacyjnej oraz zmiany w sposobie działania polityków⁷. Ze względu na taką skalę wyzwań reformy są wprowadzane

⁴ Por. R.J. Shea, *Performance Budgeting in the United States*, OECD Journal on Budgeting, Vol. 8 – No. 1, OECD 2008.

⁵ Ibidem.

⁶ Summary of Key Findings from the 2006 Meeting of the Senior Budget Officials (SBO) Network on Performance and Results, GOV/PGC/RD(2006)5, OECD, Paris, 2006.

⁷ Por. *Performance Budgeting in OECD Countries*, OECD 2007, s. 68–70.

dzane ewolucyjnie, drogą stopniowych ulepszeń i innowacji, nierzadko sektorowo i poprzez programy pilotażowe.

Wiele jest pozytywnych głosów na temat budżetowania zadaniowego i niewątpliwie wniosło ono znaczące zmiany do sfery polityk publicznych. Nie można jednak wyciągać z tego optymistycznego wniosku, że alokacja środków budżetowych może w łatwy sposób zostać przypisana do algorytmów i wartości mierników. Doświadczenia międzynarodowe⁸ wskazują, że definicyjne dla budżetu zadaniowego (przynajmniej w niektórych koncepcjach) powiązanie decyzji budżetowych z uchwycenymi w pomiarze efektami działa tylko w wąskich zakresach, w stosunku do niektórych, wyodrębnionych części budżetu. Jak pisali w 2006 r. Gilmour i Lewis: „Jest mało systematycznych dowodów, że jak dotąd budżetowanie zadaniowe, tak jak zostało wdrożone w stanach i miastach, miało duży wpływ na decyzje budżetowe”⁹. Poza tym decyzje o alokacji budżetowej podejmowane są ze względu na oczekiwane efekty przyszłe, a nie tylko ze względu na uzyskane dotychczas wyniki. Wreszcie, słabe efekty dotychczasowych nakładów na dany program mogą wynikać z niedostatecznych nakładów lub błędów w zarządzaniu programem, a nie z jego małego społecznego znaczenia. Jednym z kluczowych, wymienianych przez decydentów publicznych i badaczy tego zagadnienia, powodów ograniczonego zakresu alokacji budżetu w oparciu o efekty są niedostatki w zakresie jakości pomiaru efektów¹⁰. Uzyskiwane w pomiarze wartości wskaźników muszą być interpretowane w kontekście danego programu, innych programów i polityk publicznych. Interpretacja informacji i decyzje o wdrożeniu takich, a nie innych programów mogą być łatwo upolitycznione, gdyż mają na nie wpływ preferencje, interesy i subiektywne interpretacje. „Jest prawdopodobne, że każda partia będzie postrzegała programy, które lubi i które popiera, jako kluczowe i nieprawdopodobne, że będzie postrzegała słabe wyniki, jako podstawę do zamknięcia programu”¹¹. To może prowadzić do selektywności w stosowaniu reguł budżetowania w oparciu o efekty w taki sposób, że chronione będą pewne faworyzowane, a niewykazujące jednoznacznie pozytywnych wyników programy zainicjowane przez inną partię będą eliminowane.

Dotychczasowe doświadczenia nakazują ostrożność co do szybkiego ujęcia sfery wydatków publicznych w ramy sprawnego mechanizmu budżetu zadaniowego w jego wersji najbardziej zaawansowanej, silnie wiążącej alokację wydatków z uzyskiwanymi ocenami efektów. Funkcjonowanie takich silnych, opartych na algorytmach mechanizmów budżetowych jest możliwe odcinkowo i wymaga w każdym przypadku oceny efektów samego mechanizmu dla celów programów czy polityk, do których ten „silny” wariant zostanie wprowadzony. Bardziej prawdopodobny jest słabszy wariant budżetowania, w którym nie oczekuje się, że wszystkie decyzje alokacyjne zostaną powiązane z wynikami, ale gdzie wprowadza się informację o efektach programów do prac nad budżetem, debaty budżetowej w organach przedstawicielskich i do dyskursu publicznego. Stopniowo zakres powiązań z decyzjami alokacyjnymi może być powiększany wraz ze wzrostem jakości i możliwości dokonywania prawidłowej analizy efektów i ich wyjaśniania.

Z powyższych rozważań wynika, że pierwszy etap wdrażania reform w sektorze finansów publicznych powinien polegać na wprowadzeniu logiki zarządzania projektowego i ewaluacji do jednostek sektora publicznego. Oznacza to wykonanie dużej pracy nad umiejętnością jasnego i precyzyjnego formułowania celów, ustanawiania właściwych wskaźników do pomiaru stopnia ich osiągnięcia oraz wprowadze-

⁸ Por. *Performance Budgeting in OECD Countries*, OECD 2007; M. Robinson, J. Brumby, *Does Performance Budgeting Work? An Analytical Review of the Empirical Literature*, IMF Working Paper WP/05/210, International Monetary Fund 2005; A. Aristovnik, J. Seljak, *Performance budgeting: selected international experiences and some lessons for Slovenia*, MPRA Paper No. 15499, Munich 2009.

⁹ J.B. Gilmour, D.E. Lewis, *Does Performance Budgeting Work? An Examination of the Office of Management and Budget's PART Scores*, *Public Administration Review*, September/October 2006, s. 743.

¹⁰ Ibidem.

¹¹ Ibidem.

nia ewaluacji, która w sektorze finansów bywa często określana mianem audytu efektywnościowego¹². Przygotowywane wraz budżetem w układzie klasycznym budżety zadaniowe są dobrym krokiem w tym kierunku. Analiza formułowanych w tych dokumentach celów oraz wskaźników prowadzi do przekonania, że w tym obszarze istnieje jeszcze duża przestrzeń dla możliwych ulepszeń¹³. W tym zakresie powinny być wykorzystane gromadzone stopniowo, pozytywne i negatywne doświadczenia w zakresie ewaluacji programów i projektów finansowanych ze środków funduszy strukturalnych. W polskiej administracji publicznej jest coraz więcej zespołów gromadzących doświadczenia w tym zakresie¹⁴.

Co może dać zaplanowane i konsekwentne wprowadzenie zasad ewaluacji? Powiedzmy przedtem kilka słów o samej ewaluacji. Ewaluacja to badanie programu, jego nakładów, efektów i sposobu realizacji pod kątem przyjętego zestawu kryteriów. Zestaw ten powinien być dopracowany pod kątem potrzeb podmiotu wykorzystującego wiedzę z ewaluacji do podejmowania decyzji w zakresie analizowanego programu, programów pokrewnych lub możliwych kontynuacji. Standardowo, Komisja Europejska zaleca stosowanie w ewaluacji następujących ogólnych kryteriów (ich operacjonalizacja zależy od konkretnych potrzeb informacyjnych)¹⁵:

1. Trafność. Wymiar ten dotyczy konstrukcji samego programu, jego celów i możliwości ich osiągnięcia za pomocą realizowanych działań. Po zrealizowaniu programu pozwala jeszcze raz spojrzeć na jego logikę, a zwłaszcza precyzję, wymierność, realność i osiągalność celów przy założonych metodach działania i zaangażowanych zasobach. Doświadczenie uczy, że powodzenie i porażki programów zależą w znaczącym stopniu właśnie od jakości i precyzji celu.
2. Skuteczność. Kryterium skuteczności wymaga sporządzenia bilansu zamierzonych i osiągniętych efektów programu. Dokonuje się tego ustalając wartość przyjętego zestawu wskaźników i porównanie z wartościami założonymi. W przypadku analizy skuteczności programu ważne jest wskazanie związków pomiędzy konstrukcją działań a powodzeniami i niepowodzeniami w zakresie planowanych efektów, w szczególności pomiar efektu netto programu, czyli tego zakresu uchwyconej wskaźnikowo zmiany, który da się przypisać samej interwencji finansowanej ze środków publicznych, a nie czynnikom zewnętrznym. Trafna analiza przyczynowa skuteczności programu jest jednym z kluczowych czynników decydujących o przydatności ewaluacji i jej wpływie na programowanie przyszłych interwencji.
3. Efektywność: ocena (także porównawcza) stosunku efektywności/nakłady i wskaźnikowe ustalenie produktywności poniesionych nakładów w kategoriach wyników i rezultatów. Analiza tego, jak efektywnie nakłady zostały przekształcone w poszczególne wyniki, rezultaty i wpływ oraz z czego wynikają różnice w tym zakresie pomiędzy kontrahentami, jest istotnym elementem ewaluacji.

¹² Analiza literatury poświęconej ewaluacji z jednej strony, a audytowi efektywnościowemu z drugiej, wskazuje na wyraźne pokrewieństwo tych koncepcji, a różnica nazwy jest wynikiem odmiennej genezy audytu, a nie różnicy w podejściu. Mówimy tu, oczywiście, o ewaluacji programów według standardów określanych np. przez Komisję Europejską dla programów finansowanych z funduszy strukturalnych, które trzeba odróżnić od rozmaitych interesujących pomysłów, związanych głównie z ewaluacją projektów w skali mikro, opartych na podejściu jakościowym i silnie partycypacyjnym; te ostatnie pozostają w tym miejscu poza zakresem naszych zainteresowań. Audyt efektywnościowy rozumiany jest tutaj jako systematyczny, celowy i zorganizowany proces rzetelnej, obiektywnej i niezależnej oceny planowych, realizowanych i zakończonych polityk, programów i projektów publicznych oraz oceny bieżącego funkcjonowania jednostek sektora finansów publicznych w oparciu o kryteria gospodarności (*economy*), efektywności (*efficiency*) i skuteczności (*effectiveness*) ich działań.

¹³ Szczegółowe wnioski w tym zakresie przedstawił Ministerstwu Finansów kierowany przeze mnie zespół na ponad 420 stronach w opracowaniu: J. Górniak – kierownik zespołu, M. Jelonek, S. Mazur, A. Strzebońska, A. Szczucka, *Warunki i zakres wprowadzenia audytu efektywnościowego (ewaluacji) w polskiej administracji*, Centrum Ewaluacji i Analiz Polityk Publicznych UJ, Kraków 2008.

¹⁴ Do zespołów o największym doświadczeniu należą bez wątpienia Krajowa Jednostka Oceny w Ministerstwie Rozwoju Regionalnego i Sekcja Ewaluacji w Polskiej Agencji Rozwoju Przedsiębiorczości.

¹⁵ Por. J. Górniak, *Ewaluacja w cyklu polityk publicznych*, [w:] *„Ewaluacja funduszy strukturalnych. Perspektywa regionalna”*, MSAP, Akademia Ekonomiczna w Krakowie, Kraków 2007.

4. Użyteczność. Ten wymiar skupia uwagę na ostatecznych beneficjentach i ich potrzebach. Ewaluator powinien ustalić, w jakim stopniu oddziaływanie programu odpowiada potrzebom grupy docelowej i rozwiązaniu zidentyfikowanych problemów. W tym wymiarze mamy do czynienia z podejściem określanym przez Michaela Scrivena mianem *goal-free evaluation*¹⁶. Ewaluator bada wszystkie konsekwencje programu, niezależnie od tego, czy zostały ujęte w celach, czy nie. Następnie odnosi je do tego, co było powodem przyjęcia danego programu, czyli do określonych potrzeb społecznych, odpowiadając na pytanie, w jakim stopniu programowi udało się te problemy rozwiązać. Jest to bardzo ważny aspekt ewaluacji, często dezinterpretowany, realizowany zbyt pobieżnie lub sprowadzany do sondowania opinii beneficjentów programu.
5. Trwałość. W przypadku wykorzystania środków publicznych trwałość jest bardzo pożądaną cechą uzyskanych efektów. Trwałość należy oceniać z perspektywy wszystkich celów programu, a więc nie tylko z punktu pozytywnych skutków dla ostatecznych beneficjentów, lecz także dla instytucji, które program miał wzmocnić: administracji publicznej różnych szczebli i instytucjonalnego zaplecza służącego np. rozwojowi zasobów ludzkich na potrzeby rynku pracy oraz aktywizacji zawodowej bezrobotnych.

Ten zestaw kryteriów jest silnie ukierunkowany na tę funkcję ewaluacji, którą jest dokonywanie udokumentowanej oceny, jakie efekty przyniósł program oraz czy uzyskane efekty są rzeczywiście skutkiem podjętych działań. Cele programów wskazują zwykle na pożądaną zmianę w pewnym obszarze rzeczywistości (lub niedopuszczenie do zmiany wywoływanej przez rozmaite czynniki). W spełnianiu tej funkcji kapitalną rolę odgrywają odpowiednio dobrane wskaźniki. Warunkiem koniecznym dobrego systemu wskaźników, ale też pomyślnego wdrożenia programu w ogóle, są dobrze ustanowione i zdefiniowane cele.

Na temat roli wskaźników i zasad związanych z ich wykorzystaniem w ewaluacji oraz ich związków z poprawnym formułowaniem celów sporo już zostało powiedziane, także w innych publikacjach wydanych przez PARP¹⁷, dlatego w tym miejscu pomijam szerszą dyskusję tego zagadnienia.

Ewaluacja odpowiada także na pytania dotyczące przyczyn uzyskania takich, a nie innych efektów, leżących zarówno wewnątrz programu, a więc adekwatności doboru instrumentów, sposobie organizacji i zarządzania projektem, jak i na zewnątrz, w okolicznościach towarzyszących programowi i wpływających na jego przebieg. Jest to ważne dla dobrego spełniania przez ewaluację funkcji gromadzenia wiedzy potrzebnej w procesie planowania nowych przedsięwzięć o podobnym charakterze lub modyfikacji programów.

Klasycznie wyróżnia się, z punktu widzenia etapu programu, na którym badanie jest wykonywane, trzy rodzaje ewaluacji:

1. *Ex ante*, która jest oceną oczekiwanych skutków programu pogłębioną o analizę poprawności i spójności jego konstrukcji oraz przyjętego sposobu formułowania celów, wskaźników oraz zasad zarządzania i monitoringu programu. Na tym etapie poszukujemy odpowiedzi między innymi na następujące pytania dotyczące strategicznej orientacji programu:
 - czy szczegółowe cele programu i jego konstrukcja zostały zdefiniowane w sposób odpowiadający kluczowym potrzebom w obszarze, którego dotyczy?

¹⁶ M. Scriven, *Pros and Cons About Goal-Free Evaluation*, [w:] *Philosophical Redirection of Educational Research: The Seventy-First Yearbook of the National Society for the Study of Education*, edited by L.G. Thomas, Chicago: University of Chicago Press 1972.

¹⁷ Por. J. Górniak, K. Keler, *Wskaźniki w ewaluacji ex-post programów publicznych*, [w:] *Ewaluacja ex-post w teorii i praktyce*, PARP, Warszawa 2007; J. Górniak, K. Keler, *Ewaluacja i wskaźniki w cyklu planowania i realizacji polityk publicznych*, „Zarządzanie Publiczne” Nr. 1(3)/2008; J. Górniak, K. Keler, *Rola systemu wskaźników w ewaluacji*, [w:] K. Olejniczak, M. Kozak, B. Lendzion (red.), *Teoria i praktyka ewaluacji interwencji publicznych*, Wydawnictwa Akademickie i Profesjonalne, Warszawa 2008; J. Górniak, *Diagnoza oraz strategia i jej kwantyfikacja (cele, priorytety, wskaźniki)*, [w:] *„Ewaluacja ex ante – podsumowanie doświadczeń administracji polskiej”*, Ministerstwo Rozwoju Regionalnego, Warszawa 2007.

- czy cele programu zostały sformułowane w sposób zapewniający ich precyzję, wymierność, realność i osiągalność przy założonych metodach działania i zaangażowanych zasobach?
- czy kwantyfikacja celów w postaci podstawowych wskaźników przyjętych w programie i zaplanowanych z ich pomocą poziomów osiągnięć oddaje istotę celów postawionych w programie i pozwala na skuteczne monitorowanie i ewaluację? Czy przyjęty zestaw wskaźników jest kompletny i proporcjonalny, a same wskaźniki właściwie ujmują założone efekty programu? Czy wskaźniki mają jasną interpretację normatywną?
- w jakim stopniu cele poszczególnych działań uzupełniają, a w jakim pokrywają się z podobnymi działaniami finansowanymi z innych środków publicznych?¹⁸
- czy docelowe wartości wskaźników zostały zaplanowane w sposób realistyczny i osiągalny, a jednocześnie ambitny.

Pytania o cele są bardzo ważne, ale nie wyczerpują zakresu ewaluacji *ex ante* i podane zostały dla zilustrowania tego etapu i wzmocnienia znaczenia pracy nad celami i wskaźnikami programów.

2. Ewaluacja śródkresowa w toku programu (*on-going*). Głównym zadaniem tego rodzaju ewaluacji jest dostarczanie udokumentowanej wiedzy na temat etapowych efektów programu i jego sposobu funkcjonowania. Służyć może zarówno korekcie osi strategicznej programu, a więc także jego celów, jak i sposobu zarządzania. Ma silnie formatywny charakter i powinna kończyć się praktycznymi, wdrażalnymi rekomendacjami.
3. Ewaluacja *ex post* jest najbliższa koncepcji pomiaru efektów spotykanego w konceptualizacjach budżetu zadaniowego. Ma sumatywny charakter i służy zbadaniu zakończonego programu pod kątem ustanowionych kryteriów, w tym skuteczności i efektywności, ale, jak już wiemy – nie tylko. Ewaluacja nie służy jedynie rozliczeniu programu, lecz także gromadzeniu wiedzy o tym, co działa, a co nie; dlaczego się tak dzieje, które koncepcje – „teorie” – programu okazały się trafne, a które oparte były na niewłaściwych założeniach o związkach przyczynowych między interwencją a planowanymi efektami; czy dobrze została wybrana koncepcja wdrożenia i jak została zrealizowana, jakie wnioski należy wyprowadzić dla innych programów?

Ewaluacja służy budowaniu bazy wiedzy przydatnej w decyzjach o uruchamianiu programów publicznych, budżetowaniu zadaniowym czy programowaniu strategicznym wtedy, gdy spogląda na każdą interwencję publiczną jak na swoistą „teorię” mówiącą, że w danych warunkach podjęcie działań wynikających z projektu interwencji doprowadzi do osiągnięcia założonych skutków. Polityki publiczne opierają się na takich przyczynowych konstrukcjach. Nie zawsze są one wyrażane *explicite*, wręcz przeciwnie, często ukryte są pod powierzchnią formułowanych celów i przyjmowanych sposobów działania. Te ostatnie bywają zresztą ułomną operacjonalizacją istoty „teorii programu”. „Teoria programu” to zbiór założeń co do związków przyczynowych pomiędzy podejmowanymi działaniami a pożądanymi efektami, towarzyszy im na ogół zbiór przekonań dotyczących najbardziej właściwych sposobów wdrożenia interwencji¹⁹. Obok „teorii programu” każdemu programowi towarzyszy „teoria wdrożenia”, która jest zestawem przyjmowanych założeń dotyczących najbardziej skutecznych sposobów implementacji przyjętych rozwiązań. Efektem takiej „teorii wdrożenia” są przyjmowane rozwiązania w zakresie struktur wdrożenia (np. centralizowane vs. zdecentralizowane), sposobu zarządzania, organizacji monitoringu, kontroli finansowej itp.

¹⁸ Por. J. Górniak, *Ewaluacja w cyklu polityk publicznych...*, op. cit.

¹⁹ H.-T. Chen, *The Roots of Theory-Driven Evaluation*, [w:] M.C. Alkin (ed.), *Evaluation Roots: Tracing Theorists' Views and Influences*, Sage, Thousand Oaks, London, New Delhi, 2004, s. 132–152.

Koncepcje działania, które pomyślnie przejdą ewaluację, tworzą korpus wiedzy podnoszącej skuteczność polityk publicznych. Faktycznie, spontaniczny proces weryfikacji występuje zawsze tak czy owak w postaci społecznej oceny skutków prowadzonych polityk. Ewaluacja jest metodologicznie kontrolowanym i usystematyzowanym sposobem takiej weryfikacji prowadzonym przez ekspertów. Z natury rzeczy zatem przyczynia się do legitymizacji lub delegitymizacji efektywnościowego sposobu rządzenia. Nie ma tu miejsca na szersze omawianie zagadnień funkcji ewaluacji w prowadzeniu polityk publicznych. Warto jedynie skonstatować znaczenie tego źródła wiedzy, do którego decydenci sięgają jednak częściej ze względu na wymagania Komisji Europejskiej niż z powodu uznania rzeczywistej potrzeby. Dlatego mamy do czynienia z taką sytuacją, że środki pochodzące z funduszy europejskich podlegają ocenie pod względem celowości wydatkowania, a środki krajowe ciągle na to czekają. Jest to przejaw ogólniejszej słabości procesu rządzenia.

W tradycyjnym systemie budżetowym wprowadzenie ewaluacji jest możliwe, choć w ograniczonym zakresie. Ograniczenie to wynika z:

- braku dobrze zdefiniowanych celów działań finansowanych w ramach budżetu (problem ten ma być przewyciężony przez konstrukcje budżetu zadaniowego);
- braku ustalonego sposobu pomiaru efektów – systemu wskaźników dla efektów poszczególnych działań;
- problemów w zakresie dostępności, jakości i aktualności danych, zwłaszcza danych instytucjonalnych oraz przydatnych do oceny rezultatów i oddziaływania, a nie tylko nakładów i produktów działań.

Rzeczywiste i trwale usunięcie tych przeszkód prowadzi do zmiany systemowej w kierunku budżetowania zadaniowego, co najmniej w wersji budżetowania informowanego przez ocenę efektów (*performance informed budgeting*). Po to, by ewaluacja mogła być prowadzona w sposób sensowny, to znaczy dostarczała odpowiedniej wartości, muszą być spełnione następujące warunki:

1. Wydatki publiczne powinny być dokonywane w oparciu o budżety związane z programami. W Polsce nastąpił już pewien proces upowszechniania się standardów programowania związanych z wydatkowaniem środków unijnych, dlatego najlepiej skorzystać z tych standardów, upraszczając je w zakresie wynikającym z dotychczasowych doświadczeń i przeprowadzonych ewaluacji (te nadmierne regulacje dotyczą przede wszystkim oprzyrządowania procesu wdrożenia i kontroli).
2. Należy wzmocnić rolę Strategii Rozwoju Kraju (lub podobnego dokumentu o charakterze strategicznym) i związanych z nią programów rządowych o charakterze wieloletnim (programy operacyjne) w procesie budowania budżetu państwa (pewne zmiany w tym kierunku zawiera ostatnia nowelizacja *Ustawy o zamówieniach publicznych*). Dobrą praktyką jest przygotowywanie rocznych planów działania, wynikających z programów rządowych, które oparte są na precyzyjnie zaplanowanych zadaniach i oczekiwanych efektach. Dokumenty strategiczne powinny podlegać corocznemu przeglądowi w oparciu o dane z monitoringu i ewaluacji.
3. Cechą istotną takich dokumentów jak Strategia Rozwoju Kraju czy programy operacyjne, jest wyrażenie zdefiniowania hierarchii celów i określenie narzędzi, za pomocą których zamierza się je osiągnąć. Ewaluacja staje się nieodzownym elementem takiego systemu, gdyż dostarcza istotnych informacji na wszystkich etapach procesu programowania i wdrożenia: ewaluacja *ex ante* na etapie opracowania projektu polityki/programu; ewaluacja bieżąca w trakcie realizacji programu, jako zasilenie dorocznej procedury przygotowania planów działania i budżetu; ewaluacja *ex post* po zakończeniu wdrożenia zarówno po to, by umożliwić rozliczalność struktur publicznych, jak i po to, by dokonywać opartych o udokumentowaną wiedzę rewizji strategii i programów.

Wprowadzenie ewaluacji do procesu planowania i sprawozdawczości z wydatków publicznych z natury rzeczy stawia przed podmiotami zaangażowanymi w ten proces zagadnienie „ewaluowalności” programów. Zdolność programu do poddania ewaluacji jest swoistym sprawdzianem jakości *ex ante*, tak jak w nauce falsyfikowalność hipotez stanowi o ich „naukowości”. Warunkiem koniecznym ewaluowalności programów i działań jest prawidłowe zdefiniowanie celów działania i związanych z tym mierników oraz, w dalszej kolejności, jasne przedstawienie „teorii programu”, czyli założeń o relacjach między celami a przyjętą strukturą interwencji. Zarówno analiza obecnych zapisów budżetu zadaniowego, jak i doświadczenia z programowaniem związanym z funduszami unijnymi dają podstawę do stwierdzenia, że z określaniem celów, a w konsekwencji z doбором wskaźników ich wykonania mamy w Polsce problemy. Właściwe ujęcie celów działania jest jednym z najtrudniejszych elementów procesu programowania polityk publicznych. Cele często są mylone ze środkami działania i samymi działaniami. Często także formułowane są w sposób niejasny i nieprecyzyjny, z użyciem terminów wieloznacznych i niezdefiniowanych. Cele powinny być określone w kategoriach stanu rzeczy, który chce się osiągnąć.

Prawidłowo ujęte cele interwencji powinny dawać się łatwo wyrazić w kategoriach wskaźników, czyli zmiennych, których wartości (a dokładniej: zmiana wartości) pozwolą nam określić, czy cel został osiągnięty lub w jakim stopniu się do niego zbliżyliśmy. Wartości docelowe wskaźników nie są tożsame z celami, ale konkretyzują go w tym sensie, że od razu określimy, podczas planowania, jaka wartość docelowa wskaźnika będzie dla nas satysfakcjonująca.

Wprowadzenie ewaluacji/audytu efektywnościowego jest nierozdzielnie związane z pomiarem efektów programów/działań, czyli właściwym doбором wskaźników i zapewnieniem możliwości ich zasilenia wiarygodnymi i aktualnymi danymi. Funkcjonujące ramy sprawozdawczości finansowej dostosowane są do formy budżetu tradycyjnego, koncentrując uwagę na wydatkach nie zaś na celach czy zadaniach zaplanowanych do osiągnięcia przez dysponentów środków budżetowych. Po to, by do finansów publicznych wprowadzić ewaluację, musimy też zwrócić uwagę na szereg elementów zwiększających efektywność takiego podejścia, takich jak: budowa spójnego systemu monitoringu jako systemu wspomagania decyzji w państwie, konieczność dostosowania sprawozdawczości budżetowej do potrzeb oceny osiągania przez państwo założonych celów, konieczność zapewnienia komplementarności i standaryzacji zbieranych danych, kontrole ich jakości, konieczność uwzględniania wymogu szacowania efektu netto już na etapie zbierania danych, czy też usprawnienie i uelastycznienie pracy Głównego Urzędu Statystycznego.

Obecnie występuje szereg barier dotyczących systemu programowania planowania budżetowego i sprawozdawczości, które przekładają się na problemy z ewaluacją *ex ante* i *ex post*.

1. Planowanie skoncentrowane jest na wydatkach. Planowanie dotyczy wydatków, a nie opiera się na zdefiniowanych celach i ich miernikach (pewne zmiany w tym zakresie, trudne do wprowadzenia przy dotychczasowym stopniu przygotowania jednostek, wprowadza nowy projekt *Ustawy o finansach publicznych*).
2. Sprawozdawczość skoncentrowana jest również na wydatkach. W zakresie sprawozdawczości kluczowe znaczenie, podobnie jak przy planowaniu, ma sprawozdanie z poniesionych wydatków. Procedury kontrolne i audytorskie nastawione są na zgodność wydatków z przepisami i ograniczenie ich do działań, które tymi przepisami zostały przewidziane (nowe przepisy dotyczące sprawozdawczości budżetowej i kontroli zarządczej prowadzą do zmian w tym zakresie, o ile zostaną wypełnione treścią).
3. Występują problemy w zakresie definiowania celów działań finansowanych w ramach budżetu (problem ten ma być przezwyciężony przez konstrukcje budżetu zadaniowego, ale ciągle jest to obszar problemowy).

4. Występują poważne problemy w zakresie sposobu pomiaru efektów – systemu wskaźników dla efektów poszczególnych działań.
5. Ciągłe dotkliwe są problemy w zakresie dostępności, jakości i aktualności danych, zwłaszcza danych instytucjonalnych oraz przydatnych do oceny rezultatów i oddziaływania, a nie tylko nakładów i produktów działań.

Problemy te są przejawem szerszej gamy problemów związanych z programowaniem polityk publicznych w administracji (poza ogólnymi problemami systemowymi ze strategicznymi podstawami prowadzenia polityk publicznych wskazywanymi wcześniej), takich jak:

1. Słabość metodologiczna i merytoryczna diagnoz rozwojowych i systemu analizy polityk publicznych. Należy zbudować system wsparcia analitycznego dla władz publicznych i administracji.
2. Brak w ośrodkach naukowych tradycji prowadzenia badań diagnostycznych o charakterze problemowo-przyczynowym i badań ewaluacyjnych.
3. Problemy w zakresie opracowania przetestowanych systemów wskaźników. Wskaźniki okazały się istną „piętą achillesową” procesu programowania.
4. Braki w zakresie zasilania systemu w adekwatne i aktualne dane. Samo ustanowienie systemu wskaźników nic nie da, jeśli nie będzie możliwości zasilenia go adekwatnymi i aktualnymi danymi pochodzącymi z systemu statystyki publicznej i źródeł instytucjonalnych, tylko wyjątkowo wspieranych badaniami *ad hoc*.
5. Nieumiejętność właściwego budowania koncepcji interwencji, ujawniająca się m.in. w błędach w formułowaniu jej celów. Problemy te mają podstawy metodologiczne. Sam wybór priorytetów rozwojowych jest kwestią polityczną, mającą swoje fundamenty w dyskursie społecznym, wizjach strategicznych i negocjowaniu interesów rozmaitych stron. Profesjonalnego doprecyzowania wymaga określenie:
 - a) co ma być dostarczone przez interwencję, na co zostaną zużyte środki w toku zaplanowanych działań, czyli zdefiniowanie produktów;
 - b) jakie zmiany w sytuacji lub cechach beneficjentów interwencji mają zostać osiągnięte bezpośrednio w wyniku podjętych działań, czyli zdefiniowanie rezultatów;
 - c) jak taka zmiana sytuacji lub cech beneficjentów bezpośrednich odbije się na ich długotrwałej kondycji;
 - d) jak dana interwencja wpłynie na rozwiązanie określonego, szerszego problemu społecznego, który adresuje²⁰.

Mimo występowania całego katalogu problemów zdecydowanie należy wprowadzić ewaluację efektów wydatków jednostek sektora budżetowego. Ewaluację można zaplanować w tej sytuacji koncentrując się na badaniu faktycznych konsekwencji (efektów) wydatkowania funduszy publicznych i pod tym kontem angażując pozostałe kryteria ewaluacyjne, w tym analizując wynikową efektywność wydatków. Jest to podejście nawiązujące do wspomnianego już, lansowanego przez Michaela Scrivena podejścia *goal-free evaluation* i unijnego kryterium użyteczności zdefiniowanego wyżej. Cele można odtworzyć z logiki działań podejmowanych przez administrację i poddać refleksji, dodefiniowaniu i uspojnieniu w obrębie drzewa celów poszczególnych polityk publicznych²¹. Wprowadzenie ewaluacji jest wręcz ko-

²⁰ Szersze rozważania na ten temat oraz dokumentacje formułowanych wniosków zawiera ekspertyza, której dysponentem jest Ministerstwo Finansów: J. Górniak – kierownik zespołu, M. Jelonek, S. Mazur, A. Strzebońska, A. Szczucka, *Warunki i zakres wprowadzenia audytu efektywnościowego (ewaluacji) w polskiej administracji*, Centrum Ewaluacji i Analiz Polityk Publicznych UJ, Kraków 2008.

²¹ Zespół Centrum Ewaluacji i Analiz Polityk Publicznych UJ ma już dobre doświadczenia w zakresie stosowania tego rodzaju podejścia, które stanowi swego rodzaju *action research* i przyczynia się do doskonalenia (zwłaszcza samodoskonalenia) zespołów zarządzających programami.

nieczne dla właściwego opracowania konkretnych, mierzalnych, realistycznych i osiągalnych drzew celów dla poszczególnych obszarów wydatków publicznych. Takie też powinno być jedno z kluczowych zadań stawianych ewaluatorom.

Nowelizowana *Ustawa o finansach publicznych* zawiera szereg przesłanek do wprowadzenia na szeroką skalę ewaluacji do systemu finansów publicznych i w tym sensie jest to nowelizacja nowoczesna, idąca z trendem innowacyjnych, ale już sprawdzonych rozwiązań międzynarodowych. Te przesłanki zawarte są w następujących rozwiązaniach:

- wprowadzenie do jednostek sektora finansów publicznych kontroli zarządczej, a z nią oceny efektywności wydatków, czyli relacji zmierzonych efektów do nakładów, co logicznie (definitywnie) zakłada odnoszenie nakładów do uzyskanych efektów;
- wprowadzenie obowiązku sporządzenia planu działalności zawierającego określenie celów w obrębie zadań i podzadań, mierników ich wykonania i ich planowanych wartości;
- wprowadzenie wieloletniego planowania finansowego, a w jego ramach zasady prognozowania wydatków budżetu państwa w układzie funkcji państwa, zadań i podzadań z wyraźnym wskazaniem na rolę celów w strukturyzowaniu wydatków publicznych;
- formułowania celów i mierników ich wykonania, zadań i okresu realizacji w przypadku programów wieloletnich;
- wprowadzenie do informacji o wykonaniu budżetu omówienia realizacji założonych celów oraz zestawienia planowanych i osiągniętych wartości mierników;
- zasada przedstawiania przez Radę Ministrów Sejmowi i NIK informacji o wykonaniu zadań, co oznacza konieczność dokonania pomiaru efektów²².

Wszystkie jednostki sektora finansów publicznych powinny być zainteresowane ustaleniem standardów oraz dokonaniem wstępnych ewaluacji (we wskazanej formule otwartej, czyli w obliczu braku jasno zdefiniowanych celów) swoich obszarów działania.

Podsumowując: Ewaluacja powinna być integralnym elementem procesu zarządzania programami publicznymi. Wprowadzenie zasady ewaluacji programów nie tylko służy zasadzie rozliczalności władz publicznych, lecz także doskonaleniu procesu programowania i implementacji programów finansowanych ze środków publicznych. Budżet zadaniowy jest faktycznie silnym wprowadzeniem ewaluacji do procesu planowania i sprawozdawczości w obszarze budżetu państwa. Bezpośrednio budżetowi zadaniowemu ewaluacja dostarcza narzędzi pracy nad celami i ich wskaźnikami – ich konkretnością, realizmem i adekwatnością, a także metodologii pomiaru efektów programów, ich efektywności, a także trwałości i użyteczności. Wszystko to służyć może doskonaleniu procesu decyzji o alokacji środków budżetowych na poszczególne programy. Lepsze zrozumienie procesu wdrożenia programu i tego, co decyduje o jego powodzeniu lub niepowodzeniu, a także właściwe oszacowanie efektu netto sprzyja podniesieniu jakości decyzji o jego kontynuacji i zwiększeniu lub redukcji albo wręcz zaniechaniu dalszego finansowania. To, w jakim stopniu informacja o efektach będzie wykorzystywana faktycznie w procesie budżetowania, jest zawsze decyzją polityczną. Dlatego warto podkreślić znaczenie systematycznego komunikowania wyników ewaluacji, czego pewnym przykładem są przywoływane praktyki w Stanach Zjednoczonych. W Polsce mamy już pewne doświadczenia w zakresie ewaluacji zgromadzone przy okazji wdrażania programów finansowanych ze środków funduszy strukturalnych UE. Trzeba te doświadczenia analizować, sumować i wykorzystywać w kształtowaniu metodyki i praktyki budżetu zadaniowego. Odbędzie się to na pewno z pożytkiem dla sprawności funkcjonowania państwa.

²² Por. Ustawa z dnia 16 lipca 2009 r. o finansach publicznych, Tekst ustawy przekazany do Senatu zgodnie z art. 52 regulaminu Sejmu.

Jarosław Górniak, dr hab., prof. Uniwersytetu Jagiellońskiego i Wyższej Szkoły Europejskiej im. ks. J. Tichnera, twórca i kierownik Centrum Ewaluacji i Analiz Polityk Publicznych UJ, kierownik Zakładu Socjologii Gospodarki, Edukacji i Metod Badań Społecznych Instytutu Socjologii UJ, prodziekan Wydziału Filozoficznego UJ. Socjolog i ekonomista, specjalista w zakresie analizy polityk publicznych i ewaluacji, socjologii gospodarki i organizacji i metod badań społecznych. Ekspert w zakresie statystycznej analizy danych. Doświadczony badacz rynku i opinii, przed dziesięcią laty przesunął ciężar swojej aktywności naukowo-badawczej na zagadnienia analizy polityk publicznych i ewaluacji. Był m.in. konsultantem międzynarodowym UNDP, ekspertem Zespołu koordynacyjnego ds. ewaluacji ex-ante dokumentów unijnych na lata 2007–2013 w Ministerstwie Rozwoju Regionalnego. Kierował projektami badawczymi dotyczącymi ewaluacji i wskaźnikowego pomiaru efektów programów publicznych (m.in. w zakresie rynku pracy, wykluczenia społecznego, rozwoju kapitału ludzkiego i rozwoju regionalnego). Konsultant w zakresie programowania strategicznego i ewaluacji, a także w zakresie metodologii badań i analizy danych w zarządzaniu i administracji publicznej. Członek Rady CBOS. Członek komitetu sterującego projektu METRIS (*Monitoring European Trends in Social Sciences and Humanities*) przy DG Research Komisji Europejskiej.

Bibliografia

- Aristovnik A., Seljak J. (2009), *Performance budgeting: selected international experiences and some lessons for Slovenia*, MPRA Paper No. 15499, Munich.
- Bennis W.G. (1983), *Rozwój organizacji i losy biurokracji*, [w:] *Zachowanie człowieka w organizacji, wybór tekstów* pod red. W. E. Scotta Jr. i L. L. Cumminhsa, t.1, PWN, Warszawa.
- Chen H.-T., *The Roots of Theory-Driven Evaluation*, [w:] Alkin M. C. (ed.) (2004), "Evaluation Roots: Tracing Theorists' Views and Influences", Sage, Thousand Oaks, London, New Delhi.
- Crozier M. (1967), *Biurokracja: Anatomia zjawiska*, PWE, Warszawa.
- Gilmour J.B., Lewis D.E. (2006), *Does Performance Budgeting Work? An Examination of the Office of Management and Budget's PART Scores*, Public Administration Review, September/October.
- Górniak J. (2007), *Diagnoza oraz strategia i jej kwantyfikacja (cele, priorytety, wskaźniki)*, [w:] „Ewaluacja ex ante – podsumowanie doświadczeń administracji polskiej”, Ministerstwo Rozwoju Regionalnego, Warszawa.
- Górniak J. (2007), *Ewaluacja w cyklu polityk publicznych*, [w:] „Ewaluacja funduszy strukturalnych. Perspektywa regionalna”, MSAP, Akademia Ekonomiczna w Krakowie, Kraków.
- Górniak J., Keler K. (2007), *Wskaźniki w ewaluacji ex-post programów publicznych*, [w:] *Ewaluacja ex-post w teorii i praktyce*, PARP, Warszawa.
- Górniak J., Keler K. (2008), *Ewaluacja i wskaźniki w cyklu planowania i realizacji polityk publicznych*, „Zarządzanie Publiczne” Nr 1(3)/2008.
- Górniak J., Keler K. (2008), *Rola systemu wskaźników w ewaluacji*, [w:] K. Olejniczak, M. Kozak, B. Lendzion (red.), *Teoria i praktyka ewaluacji interwencji publicznych*, Wydawnictwa Akademickie i Profesjonalne, Warszawa.
- Górniak J. – kierownik zespołu, Jelonek M., Mazur S., Strzebońska A., Szczucka A. (2008), *Warunki i zakres wprowadzenia audytu efektywnościowego (ewaluacji) w polskiej administracji*, Centrum Ewaluacji i Analiz Polityk Publicznych UJ, Kraków.
- Merton R.K. (1982), *Teoria socjologiczna i struktura społeczna*, Rozdz. VIII „Struktura biurokratyczna i osobowość”, PWN, Warszawa.

Performance Budgeting in OECD Countries (2007), OECD.

Robinson M., Brumby J. (2005), *Does Performance Budgeting Work? An Analytical Review of the Empirical Literature*, IMF Working Paper WP/05/210, International Monetary Fund.

Scriven M. (1972), *Pros and Cons About Goal-Free Evaluation*, [w:] *Philosophical Redirection of Educational Research: The Seventy-First Yearbook of the National Society for the Study of Education*, edited by L.G. Thomas, Chicago: University of Chicago Press.

Shea R.J. (2008), *Performance Budgeting in the United States*, OECD Journal on Budgeting, Vol. 8 – No. 1, OECD.

Staniszkis J. (1972), *Patologie struktur organizacyjnych*, ZN im. Ossolińskich, Wrocław – Warszawa – Kraków.

Summary of Key Findings from the 2006 Meeting of the Senior Budget Officials (SBO) Network on Performance and Results (2006), GOV/PGC/RD(2006)5, OECD, Paris.

Ustawa z dnia 16 lipca 2009 r. o finansach publicznych, Tekst ustawy przekazany do Senatu zgodnie z art. 52 regulaminu Sejmu.

Ewaluacja wobec wyzwań stojących przed sektorem finansów publicznych

Termin „ewaluacja” robi ostatnio karierę. Znaczenie ewaluacji dla sprawnego zarządzania środkami publicznymi jest tak duże, że można się z tego tylko cieszyć. Warto jednak uświadomić sobie, że z ewaluacją (procesem, a nie terminem) mamy do czynienia od dawna, choć wiele osób dopiero niedawno dowiedziało się – jak molierowski pan Jourdain, który odkrył, że mówi prozą – że od dawna zajmuje się ewaluacją.

Nie ma jednej, powszechnie przyjętej definicji ewaluacji. Udokumentowana np. w artykule K. Olejniczaka¹ dyskusja na temat definicji ewaluacji ma mocno teoretyczny charakter i dla omawianych dalej zagadnień nie ma kluczowego znaczenia. Posługiwał się więc będę krótką definicją zawartą w dokumentach PARP², stanowiącą, że ewaluacja to „ocena wartości programu lub projektu w oparciu o określone kryteria”, zastosowana do „programów sektora publicznego, których wartość nie jest bezpośrednio weryfikowana przez rynek”, za integralną część tej definicji przyjmując również wymóg, by ewaluacja stanowiła proces towarzyszący całemu programowi, od etapu jego planowania, aż do oceny bezpośrednich i pośrednich efektów jego wykonania.

Nie jestem zwolennikiem podkreślania różnic między ewaluacją (w ukształtowanym przez unijną praktykę znaczeniu tego słowa), a szeroko rozumianą oceną programów i projektów finansowanych środkami publicznymi. Nie jestem też zwolennikiem dość ostatnio częstego nadawania różnych znaczeń terminom polskim i ich obcojęzycznym odpowiednikom. Uznajmy zatem, że od terminologicznych sporów ważniejsze jest to, iż w procedurach zarządzania środkami publicznymi pojawia się, jako niezbędny składnik tych procedur nasza tytułowa ewaluacja, wymuszająca obowiązek ciągłego, rozpoczynającego się już w fazie projektowania, badania przedsięwzięć absorbujących środki publiczne (niestety jeszcze nie wszystkich) z punktu widzenia efektów, jakie realizacja tych przedsięwzięć winna przynieść.

Jakie znaczenie dla polskich finansów publicznych miałyby upowszechnienie procedur i praktyk ewaluacyjnych? Aby na to pytanie odpowiedzieć, spróbujmy najpierw krótko scharakteryzować obecny stan naszych finansów publicznych i wskazać najważniejsze problemy, które przyjdzie rozwiązywać w najbliższym czasie.

Obecną sytuację sektora finansów publicznych w Polsce można zwięźle scharakteryzować w następujący sposób:

- po pierwsze – od 1991 roku utrzymuje się nieprzerwanie stan nierównowagi dochodów i wydatków publicznych, którego skutkiem jest rosnące zadłużenie państwa i rosnące koszty obsługi tego zadłużenia, pochłaniające już około 3% PKB rocznie,

¹ K. Olejniczak, *Teoretyczne podstawy ewaluacji ex-post*, [w:] A. Haber (red.), *Ewaluacja ex-post. Teoria i praktyka badawcza*, PARP, Warszawa 2007.

² Por. np. <http://www.parp.gov.pl/index/index/1077>

- po drugie – bardzo wysoki jest udział w wydatkach publicznych tzw. wydatków sztywnych, w tym – wspomnianych już kosztów obsługi długu publicznego oraz licznych wydatków socjalnych; stosunkowo mały jest natomiast udział wydatków wspierających bezpośrednio szeroko rozumiane procesy rozwojowe – zbyt mało środków przeznaczamy na szkolnictwo wyższe i badania, katastrofalnie niedofinansowane są szkoły wiejskie, zbyt mała część nakładów na rolnictwo przeznaczana jest na przedsięwzięcia służące unowocześnieniu produkcji rolnej itd.;
- po trzecie – nieprzerwanie rośnie realna wartość wydatków publicznych, nie posiadamy natomiast żadnych rezerw, które mogłyby zostać wykorzystane w okresach pogarszania się koniunktury gospodarczej i zahamowania wzrostu dochodów publicznych,
- po czwarte – w zbyt wolnym tempie wykorzystujemy oddane do naszej dyspozycji środki budżetowe UE, choć napływ tych środków i tak w istotny sposób przyczynił się do rozwiązywania takich palących problemów, jak wysokie bezrobocie czy drastyczne braki w infrastrukturze technicznej, komunalnej i społecznej,
- po piąte wreszcie – nie wypracowano dotychczas skutecznych mechanizmów podporządkowujących decyzje dotyczące alokacji środków publicznych i ocenę ich wykorzystania konieczności osiągnięcia jasno zdefiniowanych celów gospodarczych i społecznych, trudności nastręcza też samo zdefiniowanie tych celów.

Z jednej strony narasta więc sygnalizowana od wielu lat potrzeba podjęcia radykalnych działań na rzecz obniżenia wydatków publicznych, z drugiej zaś – decyzje alokacyjne podejmowane są bez rzetelnej oceny skutków. Co więcej – oceny wykonania budżetu i wykorzystania publicznych środków pozabudżetowych dokonywane są przede wszystkim na podstawie kryteriów formalnej poprawności poniesionych wydatków (zgodności z planem, przestrzegania formalnych procedur), marginalne znaczenie ma natomiast ocena wydatków bazująca na ocenie osiągniętych efektów. Nawet tam, gdzie oceny takie są dokonywane, nie wychodzimy zazwyczaj poza badanie mierników produktu, bez uwzględnienia mierników efektu i oddziaływania.

Rzetelnej analizy skutków wydatków budżetowych brakuje zarówno na etapie planowania, jak i w procedurach rozliczenia wykorzystanych środków publicznych. W dokumentach „uzasadniających” potrzebę wydatkowania środków publicznych na ogół starannie unika się deklarowania oczekiwanych skutków – co oczywiście powoduje niemożność ich rzetelnego uzasadnienia. Analizując wydatki budżetowe *ex post* raczej szukamy usprawiedliwienia dla naszych działań, niż próbujemy odpowiedzieć sobie na pytanie o efektywność i skuteczność naszych działań.

Warto przypomnieć, że obowiązująca obecnie ustawa o finansach publicznych zawiera przepis (art. 35 ust. 2), zobowiązujący do tego, by wydatki dokonywane były:

- 1) w sposób celowy i oszczędny, z zachowaniem zasady uzyskiwania najlepszych efektów z danych nakładów;
 - 2) w sposób umożliwiający terminową realizację zadań;
 - 3) w wysokości i terminach wynikających z wcześniej zaciągniętych zobowiązań,
- co wymaga – przynajmniej z formalnego punktu widzenia – przeprowadzenia przed dokonaniem jakiegokolwiek wydatku publicznego analizy pozwalającej na ustalenie takiego sposobu wykonania zadania.

Praktyka odbiega jednak daleko zarówno od wzorców ustawowych, jak i od dobrych doświadczeń krajów lepiej zarządzających środkami publicznymi. Sektor finansów publicznych wymaga więc zasadniczej reformy, której głównymi celami powinno być wzmocnienie strategicznego zarządzania finansami publicznymi, czyli podporządkowanie bieżących decyzji jasno zarysowanym celom długookresowym, a także zasadnicze wzmocnienie dyscypliny finansowej, rozumianej jednak nie jako konieczność ścisłego

przestrzegania rozlicznych przepisów o charakterze formalnym, lecz przede wszystkim jako konieczność bezwzględnej egzekwowania celowości i efektywności wydatków publicznych. Wdrożenie takiej reformy okazuje się bardzo trudne – ani politycy, ani urzędnicy nie tęsknią za poddaniem ich działalności ostrym rygorom wynikającym ze sformułowanych powyżej zasad, społeczeństwo jest nadal zbyt słabo zorganizowane, by przez organizacje obywatelskie wymusić konieczne zmiany. Dochodzi do tego jeszcze jeden problem – techniczny, ale ważny. Polega on na tym, że w tradycji polskiego prawa finansowego brakuje instrumentów zarządzania finansowego adekwatnych dla opisanej powyżej filozofii zarządzania środkami publicznymi.

Przystąpienie Polski do Unii Europejskiej zdynamizowało nasze kontakty z innymi krajami europejskimi, tworząc jednocześnie – dzięki zasileniom z budżetu UE – warunki do zasadniczego przyspieszenia rozwoju gospodarczego. Dodatkowe środki na poprawę infrastruktury, rozwój obszarów wiejskich czy inwestycje w kapitał ludzki są najbardziej spektakularną korzyścią z naszej obecności w UE. Korzyści tych jest jednak znacznie więcej. Jednym z ważnych, a jak się wydaje niedocenianych, efektów naszego członkostwa w UE jest upowszechnienie w Polsce nowych, efektywnych metod i instrumentów gospodarowania środkami publicznymi³. Przyjęte w UE procedury wymusiły konieczność wdrożenia planowania wieloletniego, sformalizowanych metod doboru zadań wspieranych środkami publicznymi, czy wreszcie – upowszechnienia procedur ewaluacyjnych. Na ile z procedur tych wzięto „literę”, a na ile „ducha” – trudno ocenić. Tak oceniają to również M. Ferry i K. Olejniczak: „trudno jest jednak stwierdzić, na ile ewaluacja wyszła poza mechaniczne wypełnianie unijnych przepisów i stała się aktywnym narzędziem uczenia się instytucji publicznych oraz podnoszenia jakości polityki publicznej”⁴.

Nowe narzędzia nie tylko były (bo musiały być) stosowane w odniesieniu do programów i projektów wspieranych środkami UE, lecz niewątpliwie wywarły również wpływ na sposób zarządzania środkami publicznymi wykorzystywanymi na inne cele. Pod wpływem procedur unijnych zaczęto również do zasad gospodarowania środkami krajowymi wprowadzać nowe elementy:

- rozpoczęły się eksperymenty (na razie dość nieśmiało i niezbyt udane) z zastosowaniem do konstrukcji budżetu metod budżetowania zadaniowego,
- ustanowiono obowiązek powoływania w większych instytucjach sektora finansów publicznych audytorów wewnętrznych,
- wprowadzono obowiązek sporządzania dla każdego projektu ustawy dokumentu oceny skutków regulacji (OSR), stanowiących obligatoryjnie część uzasadnienia projektu ustawy.

Wspólnym celem tych przedsięwzięć (a przynajmniej celem deklarowanym) było wzmocnienie mechanizmów kontroli wydatków publicznych i silniejsze powiązanie procedur planistycznych oraz procedur oceny wykorzystania środków publicznych⁵ z oceną efektów osiągniętych dzięki dokonany wydatkom. Cel ten – jak dotychczas – osiągnięty został w bardzo niewielkim stopniu, a zadecydowały o tym – moim zdaniem – istotne błędy popełnione przy wdrażaniu wymienionych powyżej nowych mechanizmów. Warto jednak prześledzić przyczyny niepowodzeń przy wdrażaniu metod, które miały zrewolucjonizować sposób zarządzania środkami publicznymi. Analiza taka będzie przydatna, gdy podejmowane będą próby wdrażania kolejnych instrumentów służących tym samym celom.

Niepowodzenia przy wdrażaniu audytu wewnętrznego, budżetu zadaniowego oraz OSR, miały – moim zdaniem – trzy zasadnicze przyczyny:

- wprowadzanie nowych instrumentów bez należytego przygotowania,

³ Dobrą szkołą takich metod były już procedury gospodarowania środkami przedakcesyjnymi.

⁴ M. Ferry, K. Olejniczak, *Wykorzystanie ewaluacji w zarządzaniu programami unijnymi w Polsce*, Ernst & Young, Warszawa 2008, s. 7.

⁵ Bardziej precyzyjne byłoby określenie „środków budżetu państwa”, gdyż zasięg reformatorskich przedsięwzięć nie objął – niestety – bardzo w Polsce rozbudowanej sfery pozabudżetowych instytucji państwowych.

- nadmierne sformalizowanie nowych instrumentów,
- powierzenie ustalenia kryteriów oceny (albo dokonanie samej oceny) podmiotom ocenianym.

Prace nad wdrożeniem (w pierwszej kolejności tylko do budżetu państwa) zasad budżetowania zadaniowego trwają już od pięciu lat. Zgodzić się jednak trzeba z opinią rządowych ekspertów, iż „przejście od administracji PROCEDUR do administracji REZULTATÓW”⁶ jest dopiero przed nami. Podjęcie tych prac było bardzo ważne, lecz dotychczasowe wyniki nie są znaczące. Przyczyniły się do tego zarówno liczne błędy popełnione w toku wdrażania budżetu zadaniowego. Z punktu widzenia potencjalnych problemów związanych z wdrożeniem do zarządzania środkami publicznymi technik ewaluacyjnych, za najważniejsze z popełnionych błędów uznać należy:

- powierzenie opracowania zestawów wskaźników (mierników) budżetu zadaniowego w poszczególnych dziedzinach sfery publicznej organom odpowiedzialnym za funkcjonowanie tych sfer – każdy z ministrów mógł zatem samodzielnie ustalić te wskaźniki (a przez to ustalić kryteria oceny swojej działalności); nie można się więc dziwić temu, że większość wskaźników skonstruowana została tak, by uciec przed koniecznością rozliczania się z niewykonania tych zadań (przedsięwzięć), które nie mają charakteru wewnętrznego;
- przystąpienie do wdrażania procedur budżetowania zadaniowego bez uprzedniego uruchomienia sprawnie i rzetelnie działających systemów informacyjnych, umożliwiających obliczenie i ocenę wartości mierników przypisanych do poszczególnych zadań;
- przystąpienie do wdrażania procedur budżetowania zadaniowego bez uprzedniego uruchomienia jednolitego dla całej sfery objętej budżetowaniem zadaniowym systemu ewidencji wydatków w układzie funkcji państwa, programów i działań;
- przystąpienie do wdrażania procedur budżetowania zadaniowego bez uprzedniego rozeznania potencjalnych zagrożeń i bez przygotowania działań pozwalających na minimalizację ryzyka wystąpienia tych zagrożeń oraz na minimalizację ewentualnych skutków „zmaterializowania się” tych zagrożeń; jako przykład takiego zignorowanego zagrożenia wymienić można np. zignorowanie problemów, jakie może tworzyć niechętny stosunek znacznej części administracji do wdrażania nowych metod zarządzania środkami publicznymi;
- formalistyczne podejście do kwestii rozliczenia wykonania budżetu zadaniowego – zadania te są już od dwóch lat publikowane, lecz nie mają one istotnego wpływu na prace nad kolejnymi budżetami;
- niedostateczne przygotowanie merytoryczne wielu pracowników, którym powierzono analizę zakresu działań poszczególnych instytucji, opracowanie wykazu zadań (działań), zestawu mierników odpowiadających tym zadaniom oraz docelowych wartości tych mierników;
- nieuwzględnianie kwestii wykonania założeń budżetu zadaniowego w procedurach ocen funkcjonowania poszczególnych agend rządowych oraz w procedurach ocen pracowników tych agend, w tym w szczególności – w procedurach ocen kadry kierowniczej; brak takich powiązań powoduje zazwyczaj powstanie sytuacji, w której obowiązują jednocześnie dwa na ogół sprzeczne ze sobą układy celów, z których jeden ma swe odzwierciedlenie w budżecie zadaniowym, a drugi – w rzeczywistych „lokalnych” preferencjach jednostki wykonującej zadania publiczne;
- zaakceptowanie zestawu (w skali całego budżetu państwa) mierników wykonania zadań publicznych, niespełniającego teoretyczne założenia budżetowania zadaniowego.

⁶ *Polska 2030. Wyzwania rozwojowe*, Kancelaria Prezesa Rady Ministrów, Warszawa 2009, s. 300.

Opisane powyżej błędy i zaniechania w żadnej mierze nie zaprzeczają celowości szybkiego wdrożenia metod budżetowania zadaniowego, wskazują jednak na konieczność przestrzegania pewnych dość oczywistych zasad i zachowania przy wdrażaniu logicznej kolejności przedsięwzięć⁷.

Nie przyniosło też zadowalających efektów obligatoryjne utworzenie we wszystkich większych jednostkach sektora finansów publicznych stanowisk audytora wewnętrznego⁸. Przyczyną było – moim zdaniem – nadmierne sformalizowanie procedur audytowych. Od kandydatów na audytorów wymagano zdania państwowego egzaminu, mocno sformalizowano kwestie planowania zadań audytowych i sprawozdawczości z wykonania tych zadań, wreszcie – ustanawiając zewnętrzny nadzór nad działalnością audytorów doprowadzono do tego, że audyt wewnętrzny zatracił swe zasadnicze funkcje – wspomaganie kierujących jednostkami publicznymi, stając się – w pewnym sensie – częścią systemu kontroli zewnętrznej. Ze zbyt daleko idących rozwiązań zaczęto się później wycofywać, ale istota problemu pozostała – w większości instytucji audytorzy wewnętrzni działają dziś dlatego, że tak nakazuje prawo (i po to, by prawne wymagania wypełnić), a nie dlatego, że w każdej instytucji jakaś forma wewnętrznego audytu jest niezbędna.

Praktycznie żadnych efektów nie przyniosło również wprowadzenie obowiązku sporządzania dla każdej projektowanej ustawy oceny skutków regulacji. Sporządzenie OSR, stanowiącej w istocie rzeczy rodzaj ewaluacji *ex ante* dla specyficznego projektu, jakim jest wdrożenie nowych rozwiązań prawnych, powierzono projektodawcom, co oczywiście spowodowało, że w OSR uwzględniano tylko te fakty (a raczej hipotezy), które mogły pomóc w doprowadzeniu do uchwalenia wnoszonego projektu ustawy. Nie zadbano też o właściwy poziom merytoryczny i metodologiczną poprawność OSR.

W rezultacie z wszechstronnego studium przewidywanych skutków nowej regulacji, OSR stała się potrzebnym ze względów formalnych, lecz faktycznie mało istotnym dodatkiem do uzasadnienia projektu ustawy. Powszechną praktyką stało się w OSR zaniżanie kosztów wprowadzenia nowych regulacji, nieliczne są też przypadki solidnej metodologicznej oceny wpływu projektowanej ustawy na funkcjonowanie różnych sfer gospodarki. W ostatecznym efekcie wprowadzenie OSR nie spowodowało – wbrew oczekiwaniom – korzystnych zmian w procesie tworzenia prawa.

Wróćmy teraz do tytułowych wyzwań stojących przed sektorem finansów publicznych. W największym skrócie można je zdefiniować tak:

**W MOŻLIWIE KRÓTKIM CZASIE MUSIMY DOPROWADZIĆ DO ZRÓWNOWAŻENIA
FINANSÓW PUBLICZNYCH, DOKONUJĄC JEDNOCZEŚNIE TAKICH ZMIAN W STRUKTURZE
WYDATKÓW, BY ŚRODKI PUBLICZNE W WIĘKSZYM STOPNIU WSPIERAŁY PROCESY ROZWOJOWE**

Bez zasadniczej poprawy efektywności wykorzystania środków publicznych, bez koncentracji wydatków na sprawach naprawdę niezbędnych z punktu widzenia naszych celów rozwojowych, jednoczesne osiągnięcie obu celów jest niemożliwe. Reforma finansów publicznych, na którą wciąż czekamy, musi nie tylko doprowadzić do eliminacji wydatków ewidentnie zbędnych, ale musi również wzmocnić proefektywnościowe mechanizmy w zarządzaniu finansami publicznymi.

To „wzmocnienie mechanizmów” nie jest jednak sprawą prostą. Za każdym, nawet najbardziej nieracjonalnym (i czasem sprzecznym z poczuciem społecznej sprawiedliwości) wydatkiem stoi grupa bene-

⁷ Szczegółową analizę procesu wdrażania budżetowania zadaniowego znaleźć można, [w:] W. Misiąg, *Budżet zadaniowy na rok 2008 – i co dalej*, „Biuletyn Finansów Publicznych” 4(24)/2007, s. 23–30.

⁸ Zgodnie z art. 49 ustawy o finansach publicznych utworzenie audytu wewnętrznego jest obowiązkowe w wymienionych w ustawie instytucjach rządowych oraz we wszystkich jednostkach sektora finansów publicznych, które gromadzą znaczne środki publiczne lub dokonują znacznych wydatków publicznych. Graniczną wielkość znacznych środków publicznych określa w rozporządzeniu Minister Finansów.

ficjentów. W administracji bardzo często czysto oportunistyczna chęć utrzymania *status quo* wygrywa z ambicjami reformatorskimi. Dążąc do poprawy efektywności funkcjonowania finansów publicznych (co naprawdę oznacza dążenie do bardziej skutecznego i efektywnego funkcjonowania państwa), musimy mieć świadomość trudności i starannie dobierać środki służące wdrożeniu nowych mechanizmów.

Wiele dobrych pomysłów upadło u nas przez nadmierną wiarę w skuteczność prawa. Nie jest niestety prawdą, że aby wzmocnić mechanizmy proefektywnościowe, wystarczy je zadekretować i – na wszelki wypadek – ustanowić system kontroli i sprawozdawczości. W interesującym nas przypadku znacznie większe znaczenie miałoby wymuszenie zmian poprzez konsekwentne przestrzeganie zasady „trudnego pieniądza” i rozliczanie dysponentów środków publicznych na wszystkich szczeblach budżetowej piramidy nie z formalnej poprawności operacji finansowych, lecz z osiągniętych, realnych efektów.

Nowych metod trzeba się uczyć. Znajomość nowoczesnych metod zarządzania sprawami publicznymi, w tym – zarządzania środkami finansowymi, nie jest najmocniejszą stroną polskiej administracji. Interesujące jest przy tym, że więcej udanych prób wdrażania nowych koncepcji zarządczych obserwujemy w jednostkach samorządu terytorialnego niż w administracji rządowej. Przypomnieć też warto, że samorządowcy – oczywiście nie wszyscy – uczyli się (głównie od zagranicznych ekspertów) nowoczesnego zarządzania sprawami publicznymi i wdrażali nowe rozwiązania na długo przed tym, gdy na przykład w prawie finansowym pojawiły się pierwsze próby wykorzystywania budżetowania zadaniowego.

Od tworzenia norm prawnych wymuszających formalne stosowanie procedur budżetowania zadaniowego, audytu wewnętrznego, czy wprowadzających obowiązek ewaluacji wszystkiego i bez przerwy, ważniejsza jest edukacja – rządowi zdecydowanie brakuje dziś silnego ośrodka, w którym mogłyby rozwijać się nowe metody zarządzania, który mógłby uczyć nowych metod, promować najlepsze praktyki krajowe i zagraniczne, pomagać we wdrażaniu nowych metod. Uważam przy tym, że uczyć należy nie tylko pracowników administracji, ale również tych, którzy tę administrację chcą kontrolować. Przekonany jestem, że jakość pracy administracji, a przez to – efektywność gospodarowania środkami publicznymi – zależy w dużej mierze od aktywnej postawy organizacji pozarządowych, które w wielu rozwiniętych krajach są najlepszym kontrolerem i recenzentem działań administracji publicznej. To, że w Polsce trzeci sektor nie jest jeszcze (przynajmniej w sferze tzw. *watchdogs*) dostatecznie mocny, jest jednym z ważnych powodów trudności w reformowaniu finansów publicznych.

I w ten sposób doszliśmy do określenia miejsca, jakie ewaluacja winna zajmować w procedurach zarządzania środkami publicznymi. Z formalnie wymaganej procedury zarządzania środkami pochodzącymi z budżetu Unii Europejskiej, powinna stać się powszechnie stosowanym narzędziem planowania finansowego i podstawową metodą analizy wykorzystania środków publicznych. Ewaluacja – rozumiana jako metodologicznie uporządkowany proces towarzyszący każdemu przedsięwzięciu podejmowanemu przez władze publiczne od etapu jego koncepcji do ostatecznego rozliczenia realizacji w ujęciu rzeczowym (efektów) i finansowym – jest dziś pilnie potrzebna.

Osiągnięcie stanu, w którym metody ewaluacyjne staną się standardem zarządzania środkami publicznymi, nie będzie jednak proste. Można przy tym wskazać co najmniej kilka źródeł trudności, na jakie napotkamy wdrażając szeroko zasady i metody ewaluacji. Za najważniejsze z nich uznać należy:

- słabe przygotowanie administracji publicznej do przyjęcia, zaakceptowania i stosowania nowych metod zarządzania,
- obecną strukturę formalną budżetu państwa, w której myślenie kategoriami instytucji publicznych przeważa nad myśleniem w kategoriach celów i programów służących realizacji tych celów,
- brak jasno określonej hierarchii celów realizowanych przez administrację publiczną, wynikające zarówno z nieistnienia w Polsce sformalizowanego planowania wieloletniego (poza tym, które zostało

wymuszone przez procedury unijne), jak i z nadmiaru i niespójności dokumentów określających strategię sektorowe,

- brak doświadczeń w analizie *ex ante* dużych programów rządowych.

Podobną diagnozę stanu finansów publicznych zaleźć można w dokumentach rządowych⁹. Niestety, jak dotąd świadomość braków i przekonanie o konieczności zmian nie przekłada się na bardziej energiczne działania.

Jeśli ma być skuteczne, upowszechnianie zasad ewaluacji w sektorze finansów publicznych musi więc być traktowane jako fragment większego przedsięwzięcia – głębokiej reformy systemu finansów publicznych, w której zasadniczą rolę odegrać powinno realne, a nie tylko formalne, wdrożenie zasad budżetowania zadaniowego i wprowadzenia planowania wieloletniego. Z żalem trzeba stwierdzić, że uchwalona w sierpniu 2009 roku nowa ustawa o finansach publicznych nie może być uznana za początek takiej reformy, co więcej – prawdopodobnie drogę do takiej reformy wydłuża. Próba powszechnego wdrożenia metod ewaluacyjnych w nieprzygotowanym do tego otoczeniu prawdopodobnie skończy się niepowodzeniem.

Najlepszą drogą do szybszego przygotowania sektora finansów publicznych do przyjęcia nowych metod zarządzania, powinny być próbne wdrożenia, prowadzone na dobrze wybranych „poligonach”. Poligonami takimi mogą być na przykład:

- proponowane przez rząd programy wieloletnie; przeanalizowałem wszystkie uruchomione w ostatnich latach programy wieloletnie i na tej podstawie z przekonaniem mogę stwierdzić, że w przeważającej większości przypadków uzasadnienia ustaw stanowiących te programy (będące szczególnym rodzajem ewaluacji *ex ante*) były dramatycznie słabe merytorycznie;
- niektóre agencje rządowe – instytucje o relatywnie dobrze sprecyzowanych celach i sporej samodzielności działania, oceniane dziś głównie na podstawie kryteriów formalnych, bez głębszej analizy wpływu dokonywanych wydatków na stopień realizacji celów działania¹⁰.

Programując szersze wdrażanie metod ewaluacyjnych do praktyki zarządzania środkami publicznymi pamiętać należy o tym, by nie powtarzać już raz popełnionych błędów. Opisane wcześniej przypadki wdrażania audytu wewnętrznego i oceny skutków regulacji wyraźnie wskazują na to, że przekonywanie o celowości stosowania nowych metod, propagowanie dobrych praktyk, jest znacznie ważniejsze niż „załatwienie sprawy” przez uchwalenie odpowiednich przepisów.

Wdrożenie procedur ewaluacyjnych jest jednym z podstawowych warunków pomyślnego wprowadzenia zasad budżetowania zadaniowego. Istotą reformy budżetu, prowadzącej do podporządkowania zasad planowania i wykonywania budżetu (a również planowania i realizacji zadań w pozabudżetowych instytucjach publicznych) nie jest to, by wymyślić wskaźniki realizacji zadań i potem obserwować ich wykonanie, lecz to, by dobór sposobów wykorzystania środków publicznych podporządkować osiągnięciu założonych celów. Podejmując nowe przedsięwzięcia, musimy wiedzieć, czy służą one naprawdę realizacji założonych celów (tu swoje miejsce ma rzetelna ewaluacja *ex ante*), potem musimy śledzić wykonanie programu i oceniać na bieżąco jego efekty (podejmując, jeśli trzeba, decyzje korygujące), wreszcie – rozliczając wykonanie zadania musimy skonfrontować poniesione nakłady ze stopniem realizacji celów. Bez rzetelnej ewaluacji budżet zadaniowy stanie się tylko nową dekoracją dla starej niegospodarności i niskiej skuteczności.

⁹ Por. np. *Założenia systemu zarządzania rozwojem Polski*, Ministerstwo Rozwoju Regionalnego, Warszawa 2009.

¹⁰ Poleciałbym na przykład przestudiowanie przypadku Agencji Nieruchomości Rolnych, następczyni Agencji Własności Rolnej Skarbu Państwa. Postawione jej prawie 20 lat temu zadania pozostają nadal niewykonane, nie ma to jednak istotnego wpływu na sposób oceny ANR w kolejnych latach budżetowych.

W ostatnich latach kilkakrotnie popełniliśmy poważny błąd, nakazując instytucjom publicznym, by same ustaliły kryteria swej oceny, same oceniały zadania, które zamierzają podjąć i w dodatku same ustalały, czy nałożone na nie zadania zostały wykonane. Tego błędu nie można powtórzyć wdrażając metody ewaluacyjne. Istotne jest w szczególności to, by:

- kryteria ewaluacji ustalane były poza instytucją, która ewaluowany program będzie realizować,
- powierzyć sporządzanie dokumentów OSR wyspecjalizowanej agencji rządowej.

Oddzielną uwagę poświęcić należy związkowi pomiędzy ewaluacją i przejrzystością finansów publicznych. Znaczenie tej kwestii wynika z dwóch podstawowych przesłanek:

- jawność (dostęp do danych) oraz przejrzystość (czytelne reguły działania i efekty) finansów publicznych są nie tylko niezbywalnym prawem obywateli-podatkników, ale i jednym z podstawowych warunków uzyskania zadowalającej efektywności wykorzystania środków publicznych,
- bez jasnego określenia zakładanych celów i bez poddania publicznej ocenie stopnia realizacji tych celów nie ma prawdziwej jawności i przejrzystości.

Dzisiejszy stan jawności i przejrzystości finansów publicznych w Polsce scharakteryzować można stwierdzeniem, iż w zasadzie wiemy **kto** wydaje środki publiczne i **ile** ich wydaje, natomiast znacznie mniej wiemy o tym **po co** wydatki te są ponoszone i **co z nich wynika**.

Sam fakt wydatkowania konkretnej kwoty przez pewien organ administracji publicznej jest na ogół trudny do oceny. O poprawności („słuszności”) danego wydatku decyduje bowiem jego efekt (rozumiany jako ilość pewnego dobra dostarczanego obywatelom), a także relacja wydatkowanej kwoty do ilości tego dobra. Dla oceny funkcjonowania instytucji zarządzającej środkami publicznymi istotna jest przy tym nie tylko osiągnięta relacja nakładów (wydatków) do efektu (ilości dostarczonego dobra), ale i to, jak osiągnięta efektywność wydatków publicznych ma się do efektywności zakładanej. Gdy planowana efektywność nie została osiągnięta, obywatele mają prawo wiedzieć dlaczego tak się stało. Czy celowo zawyżono planowane efekty, by zwiększyć szanse na uzyskanie środków (tworzenie budżetu jest – w pewnym stopniu – „konkurem efektywności”), czy też warunki, w których realizowano zadanie były gorsze od zakładanych, czy może dlatego, że realnej efektywności nie udało się zrealizować ze względu na niesprawność instytucji wykonującej zadanie publiczne i w związku z tym gospodarującej przydzieloną jej kwotą środków publicznych? Na to pytanie trudno jest odpowiedzieć, gdy realizacji zadania publicznego nie towarzyszy proces ewaluacji tego zadania.

Za udowodniony empirycznie fakt należy uznać istnienie ścisłej korelacji pomiędzy stanem jawności i przejrzystości finansów publicznych i tempem wzrostu gospodarczego¹¹.

Dla szerokiego i efektywnego wykorzystania metod ewaluacji w praktyce sektora finansów publicznych niezbędne są zmiany kryteriów oceny instytucji wykonujących zadania publiczne oraz osób kierujących tymi instytucjami. Dopóki podstawowym kryterium oceny nie stanie się skuteczność i efektywność wykonywania powierzonych instytucji zadań, wszelkie nakazy zmierzające do obligatoryjnego wdrożenia nowoczesnych technik zarządzania wymuszających – w założeniu – wyższą efektywność wykorzystania środków publicznych – traktowane będą jak biurokratyczny obowiązek, a nie jak narzędzie pomagające w osiągnięciu zakładanych celów i – przy okazji – jako sposób na uzyskanie wysokiej oceny prowadzonej działalności. *Casus* Ministerstwa Infrastruktury pokazuje, że doprowadzenie do takiego stanu będzie dość skomplikowane.

Dużą szansą na unowocześnienie procesów zarządzania środkami publicznymi stworzyło wprowadzenie w nowej ustawie o finansach publicznych¹² przepisów dotyczących kontroli zarządczej. Zgodnie

¹¹ Por. np. H. Farhan, *Fiscal Transparency and Economic Outcomes*, IMF Working Paper 05-225, International Monetary Fund, New York 2005.

¹² USTAWA z dnia 27 sierpnia 2009 r. o finansach publicznych, Dz. U. Nr 157, poz. 1240.

z art. 70 tej ustawy minister kierujący działem sporządza corocznie plan działalności dla kierowanych przez niego działów administracji rządowej. Plan ten ma zawierać:

- określenie celów w ramach poszczególnych zadań budżetowych,
- wskazanie podzadań służących osiągnięciu celów,
- mierniki określające stopień realizacji celu i ich planowane wartości.

Minister ma również sporządzić sprawozdanie z wykonania planu działań. Termin na opracowanie tego sprawozdania (koniec kwietnia roku następującego po roku budżetowym) wskazuje, że sprawozdania ministrów zamierza się wykorzystać przy opracowywaniu rządowego sprawozdania z wykonania budżetu państwa.

Wprowadzenie takiej zasady jest oczywiście słuszne. „Filozofia” wykonywania budżetu państwa staje się jasna: minister otrzymuje prawo dysponowania środkami budżetowymi nie po to, by utrzymać podległe mu instytucje, ale po to, by sfinansować przedsięwzięcia ujęte w planie działań i osiągnąć zamierzone, precyzyjnie skwantyfikowane cele.

Na to, by wprowadzenie planów działania faktycznie stanowiło postęp w metodach zarządzania środkami publicznymi, konieczne jest jednak spełnienie kilku ważnych warunków. Sformułować je można tak:

- po pierwsze – wstępne wersje planów działania powinny zostać sporządzone na tyle wcześniej, by mogły być wykorzystane podczas prac nad projektem ustawy budżetowej;
- po drugie – sporządzony przez ministra plan działania powinien być poddany ocenie zewnętrznego ewaluatora, który powinien przeanalizować:
 - a) prawidłowość doboru celów,
 - b) adekwatność wyboru podzadań do przyjętych celów,
 - c) metodologiczną poprawność zaproponowanych mierników,
 - d) zaproponowane do osiągnięcia wartości tych mierników, przy czym analiza ta powinna dotyczyć przede wszystkim realności zaproponowanych wartości, ocenianej zarówno z punktu widzenia wielkości środków wyasygnowanych na realizację planu działania, jak i z punktu widzenia potencjalnych skutków pełnej realizacji planu działania;
- po trzecie – w toku wykonywania planu działania powinny być na bieżąco analizowane dane o dokonywanych wydatkach, jak i o potencjalnych zagrożeniach wykonania planu „rzeczonego”;
- po czwarte – w toku prac nad sprawozdaniem Rady Ministrów wykonanie planów działania powinno zostać poddane zewnętrznej ewaluacji, obejmujące, obok analizy przedsięwzięć składających się na plan działania, również ocenę funkcjonowania systemu gromadzenia, weryfikacji i przetwarzania danych wykorzystywanych do ustalenia wartości mierników określających stopień realizacji celów założonych w planie działania;
- po piąte wreszcie – zbiorcze wyniki ewaluacji planów działania winny być przeanalizowane przez Radę Ministrów przed rozpoczęciem kolejnego cyklu prac budżetowych.

Jeśli plany działania nie staną się przedmiotem prowadzonych według wszelkich reguł sztuki procedur ewaluacyjnych, będą one tylko jeszcze jednym wypracowaniem pisanim „ku pokrzepieniu serc”, by móc wykazać się stosowaniem nowoczesnych metod planowania. Właśnie dlatego konieczne jest, by resortowe plany działania podlegały ewaluacji przez podmiot zewnętrzny.

Przepisy o kontroli zarządczej są więc szansą, ale stanowią też – gdyby zrealizował się opisany powyżej pesymistyczny scenariusz, poważne zagrożenie, gdyż pozorowane stosowanie metod mających swe źródło w koncepcjach *new public management* jest gorsze od solidnego stosowania metod tradycyjnych.

Sformułowane powyżej refleksje i uwagi podsumować można następującymi siedmioma wnioskami:

- po pierwsze – procedury i metody ewaluacji powinny jak najszybciej zostać wdrożone do praktyki zarządzania możliwie dużą częścią środków publicznych;
- po drugie – efektywne wdrożenie procedur i metod ewaluacji wymagać będzie uprzednich zmian w funkcjonowaniu sektora finansów publicznych;
- po trzecie – unikać należy nadmiernej formalizacji procedur ewaluacyjnych, nie można też wierzyć w to, że ustawowe nakazy stosowania ewaluacji doprowadzą nas do pożądaných rezultatów;
- po czwarte – szczególnie istotne jest powszechne stosowanie ewaluacji nowych programów i przedsięwzięć podejmowanych przez administrację rządową, a kluczowe znaczenie będzie miała zasadnicza poprawa – w relacji do stanu obecnego – jakości ewaluacji *ex ante*;
- po piąte – konieczne wydaje się utworzenie rządowego ośrodka wspierającego wdrażanie nowych technik zarządzania publicznego, przy czym ośrodek ten powinien być w strukturze administracji tak usytuowany i wyposażony w takie kompetencje, by mógł nie ulegać branżowym naciskom – jest to szczególnie istotne przy ustalaniu kryteriów ewaluacyjnych;
- po szóste – dążyć trzeba do tego, by ewaluacja nie tylko była procesem towarzyszącym całemu przedsięwzięciu, lecz również wpływała na funkcjonowanie władz publicznych po zakończeniu ewaluowanego przedsięwzięcia,
- po siódme – korzyści z szerszego wykorzystywania metod ewaluacyjnych objawia się w pełni przy ewaluacji programów wieloletnich,
- po ósme wreszcie – nowym metodom oceny planowanych i realizowanych przedsięwzięć towarzyszyć muszą zmiany kryteriów osób i instytucji gospodarujących środkami publicznymi.

Zastosowanie się do tych wniosków nie daje gwarancji powodzenia, lecz ich zignorowanie redukuje niemal do zera szanse istotnej poprawy funkcjonowania finansów państwa.

Wdrożenie wniosków wyliczonych w poprzednim punkcie wymaga nie tylko zmian o charakterze instytucjonalnym (a wcześniej – politycznej woli, by to uczynić) i przestrzegania elementarnych zasad „kultury ewaluacji”, lecz również poważnych prac metodologicznych.

Słabym punktem dotychczasowej praktyki stosowania metod ewaluacyjnych wydaje się przede wszystkim technika wykonywania ewaluacji *ex ante* programów o złożonej strukturze. Jest to zdecydowanie najtrudniejsze, niepoddające się próbom wtłoczenia w jednolite, formalne algorytmy. Podstawowym problemem przy prowadzeniu takich ewaluacji wydaje się przede wszystkim brak teoretycznych podstaw dla określenia potencjalnych skutków interwencji publicznych podejmowanych w nadal niestabilnej i szybko zmieniającej się gospodarce, przechodzącej w przyspieszonym tempie drogę od realnego socjalizmu do gospodarki rynkowej w demokratycznym państwie. Jest to wyzwanie dla badaczy procesów społeczno-ekonomicznych, nie może być jednak uzasadnieniem zaniechania prób ewaluacji *ex ante* dużych programów rządowych. Dość łatwo wskazać można w praktyce ostatnich lat przypadki podejmowania takich działań, które nie przyniosły zakładanych efektów i które prawdopodobnie nie byłyby podjęte, gdyby przed ich wdrożeniem przeprowadzono – przy istniejących niedostatkach wiedzy teoretycznej – rzetelne postępowanie ewaluacyjne. Wskazać też można przypadki, gdy decydowano o wydatkowaniu znacznych kwot nawet nie próbując ocenić potencjalnych efektów.

Stwierdzenie to warto poprzeć kilkoma, dość różnorodnymi przykładami.

- A. **Oddłużanie szpitali.** Od 1999 roku, czyli od momentu przekształcenia ZOZ – jednostek budżetowych w samodzielne publiczne zakłady opieki zdrowotnej (SPZOZ), rząd uruchomił cztery programy oddłużeniowe. Polegały one na udzielaniu szpitalom pożyczek na uregulowanie zobowiązań oraz

na całkowitym lub częściowym umorzeniu długów. Mimo poniesienia wielomiliardowych nakładów nie udało się osiągnąć oczekiwanej redukcji zadłużenia. Jedynym widocznym efektem było zastąpienie w SPZOZ wymagalnych zobowiązań z tytułu dokonanych wydatków przez zobowiązania z tytułu pożyczek udzielonych ze Skarbu Państwa.

Przyczyn niepowodzenia było tu kilka:

- po pierwsze – dla żadnego z programów oddłużenia nie przeprowadzono ewaluacji *ex ante*, która – być może – wskazałaby na nieskuteczność i nieefektywność uruchamianego programu,
- po drugie – w czasie realizacji programów nie analizowano na bieżąco wypełniania przez oddłużane szpitale warunków, pod którymi oddłużanie miało być przeprowadzone,
- po trzecie – nie przeprowadzono (po żadnym z programów) solidnej ewaluacji *ex post* i w związku z tym nie wykorzystywano w następnych programach wniosków wynikających z przeprowadzonej ewaluacji¹³.

- B. **Subwencje oświatowe.** Obowiązujący obecnie system subwencjonowania jednostek samorządu terytorialnego skonstruowany jest w ten sposób, że subwencje oświatowe przyznawane na utrzymanie szkół w małych miejscowościach są większe od subwencji na szkoły w większych miastach. Miało to wyrównać szanse młodzieży z małych miejscowości, zapewniając im dostęp do dobrze wyposażonych (w kadre i inne zasoby) szkół. W 2004 r. na zlecenie MEN Przemysław Śleszyński dokonał obliczeń, z których wynikało, że korelacja między nakładami na oświatę w gminach i wynikami uzyskiwanymi przez uczniów na sprawdzianie po szkole podstawowej i wynikami egzaminu gimnazjalnego jest ujemna¹⁴. Nie spowodowało to zmian w systemie. Późniejsze badania IBnGR wykazały, że preferencje dla szkół w małych miejscowościach nie pokrywają znacznej części obiektywnie uzasadnionej różnicy kosztów utrzymania szkół wiejskich i miejskich¹⁵. W efekcie w szkołach wiejskich udział nauczycieli na wyższych szczeblach awansu zawodowego jest wyraźnie niższy niż w miastach, a wyniki sprawdzianu (po klasie szóstej) są w szkołach wiejskich dramatycznie gorsze od wyników w miastach. W przykładzie tym bulwersuje fakt, iż administracja rządowa nie tylko nie przeprowadziła samodzielnej ewaluacji mechanizmu subwencjonowania szkół, ale i nie zareagowała na sygnały z zewnątrz.
- C. **Fundusz Pracy.** Od wielu lat w Funduszu Pracy obserwowane jest zjawisko polegające na przesuwaniu (w trakcie roku budżetowego) środków z bardziej efektywnych form przeciwdziałania bezrobociu na formy mniej efektywne¹⁶. Fakt ten jest dobrze znany, lecz mimo to problem nieefektywnej alokacji nie stał się przedmiotem analitycznych prac w agendach rządowych.
- D. **Narodowy program zwalczania chorób nowotworowych.** W 2005 r. Sejm uchwalił ustawę o ustanowieniu programu wieloletniego „Narodowy program zwalczania chorób nowotworowych”¹⁷. Program miał trwać 12 lat, na jego realizację sejm przeznaczył kwotę 3 mld zł (250 mln zł rocznie). Ani w ustawie, ani w żadnym dokumencie rządowym nie określono żadnych parametrów ilościowych dla tego programu, nie ustanowiono też systemu gromadzenia danych o osiągniętych efektach. Wydaje się oczywiste, że przy tak ważnym programie, długotrwałym i angażującym znaczące środki, należało

¹³ Szerzej na ten temat: W. Misiąg, *Zadłużenie szpitali publicznych*, „Biuletyn Finansów Publicznych” 2(27)/2009, s. 17–33.

¹⁴ P. Śleszyński, *Ekonomiczne uwarunkowania wyników sprawdzianu szóstoklasistów i egzaminu gimnazjalnego prowadzonych w latach 2002–2004*, PAN – Instytut Geografii i Przestrzennego Zagospodarowania, Warszawa 2004.

¹⁵ E. Malinowska-Misiąg, W. Misiąg, M. Tomalak, *Centralne finansowanie ochrony zdrowia i edukacji w Polsce*, Instytut Badań nad Gospodarką Rynkową, Warszawa 2008.

¹⁶ Najprostszym miernikiem efektywności wykorzystania środków Funduszu Pracy jest wielkość kosztów zastosowania pewnego instrumentu aktywnego przeciwdziałania bezrobociu (np. robót interwencyjnych, dotacji na tworzenie miejsc pracy), przypadająca na jedną osobę, która w wyniku stosowania tego instrumentu uzyskała trwałe zatrudnienie.

¹⁷ Ustawa z dnia 1 lipca 2005 r. o ustanowieniu programu wieloletniego „Narodowy program zwalczania chorób nowotworowych”, Dz. U. Nr 143, poz. 1200, z późn. zm.

nie tylko określić (i kontrolować) wskaźniki „produktywne”, ale i postawić wymagania dotyczące efektów realizacji programu.

Żaden z opisanych powyżej przypadków nie był przedmiotem jakichkolwiek analiz ewaluacyjnych prowadzonych poza resortem odpowiedzialnym za realizację danego problemu. Uważam to za kolejny argument za stworzeniem stałych, ponadresortowych „służb ewaluacyjnych” – powierzenie obowiązku ewaluacji programów rządowych ministrom odpowiedzialnym za wykonanie programu utrwala resortowe (branżowe) podziały, odbiera analizom ewaluacyjnym wiarygodność i utrudnia prowadzenie rozsądnej polityki finansowej.

Zwrócić też należy uwagę na fakt, że rozszerzenie zakresu stosowania procedur ewaluacyjnych na programy rządowe finansowane bez wykorzystania środków UE oznaczać będzie konieczność ewaluacji programów w obszarach, których „unijne” ewaluacje nie obejmowały – to również stanowi poważne wyzwanie dla badaczy i praktyków ewaluacji.

Wojciech Misiąg, dr nauk ekonomicznych (SGPiS 1984), od 1994 r. kierownik Zakładu Finansów Publicznych w Instytucie Badań nad Gospodarką Rynkową, od 2003 r. doradca prezesa Najwyższej Izby Kontroli, członek Kolegium NIK. W latach 1974–1987 pracownik naukowy Szkoły Głównej Planowania i Statystyki, od 1987 do 1994 r. w administracji rządowej, w tym w latach 1989–1994 – podsekretarz stanu w Ministerstwie Finansów, odpowiedzialny za finanse publiczne i politykę finansową. Po roku 1994 pracował również w Wyższej Szkole Przedsiębiorczości i Zarządzania oraz był profesorem Wyższej Szkoły Informatyki i Zarządzania w Rzeszowie. Prowadził zajęcia na studiach podyplomowych w Uniwersytecie Warszawskim i Akademii Leona Koźmińskiego. Autor ponad 100 opracowań (w tym ponad 30 publikacji książkowych) z zakresu programowania matematycznego, metod symulacyjnych oraz finansów publicznych. Specjalizuje się w zagadnieniach finansów publicznych, ze szczególnym uwzględnieniem problematyki organizacji i zasad funkcjonowania instytucji publicznych, finansowania sfery budżetowej oraz finansów samorządowych. Ekspert Sejmu RP i Banku Światowego. W latach 2004–2008 był redaktorem naczelnym publikowanego przez IBnGR kwartalnika „Biuletyn Finansów Publicznych”.

Bibliografia

- Farhan H. (2005), *Fiscal Transparency and Economic Outcomes*, IMF Working Paper 05-225, International Monetary Fund, New York.
- Ferry M., Olejniczak K. (2008), *Wykorzystanie ewaluacji w zarządzaniu programami unijnymi w Polsce*, Ernst & Young, Warszawa.
- Kancelaria Prezesa Rady Ministrów (2009), *Polska 2030. Wyzwania rozwojowe*, Kancelaria Prezesa Rady Ministrów, Warszawa.
- Malinowska-Misiąg E., Misiąg W., Tomalak M. (2008), *Centralne finansowanie ochrony zdrowia i edukacji w Polsce*, Instytut Badań nad Gospodarką Rynkową, Warszawa.
- Misiąg W. (2007), Budżet zadaniowy na rok 2008 – i co dalej, *Biuletyn Finansów Publicznych* 4(24).
- Misiąg W. (2009), Zadłużenie szpitali publicznych, *Biuletyn Finansów Publicznych* 2(27).
- Ministerstwo Rozwoju Regionalnego (2005), *Założenia systemu zarządzania rozwojem Polski*, Ministerstwo Rozwoju Regionalnego, Warszawa.

- Olejniczak K. (2007), *Teoretyczne podstawy ewaluacji ex-post*, [w:] Haber A. (red.), *Ewaluacja ex-post. Teoria i praktyka badawcza*, PARP, Warszawa.
- Śleszyński P. (2004), *Ekonomiczne uwarunkowania wyników sprawdzianu szóstoklasistów i egzaminu gimnazjalnego prowadzonych w latach 2002–2004*, PAN – Instytut Geografii i Przestrzennego Zagospodarowania, Warszawa.
- Ustawa z dnia 1 lipca 2005 r. o ustanowieniu programu wieloletniego „Narodowy program zwalczania chorób nowotworowych”, Dz. U. Nr 143, poz. 1200, z późn. zm.
- Ustawa z dnia 27 sierpnia 2009 r. o finansach publicznych, Dz. U. Nr 157, poz. 1240.

Marta Postuła

Rola ewaluacji w budżetowaniu zadaniowym

Wstęp

W ramach realizowanych w Polsce zmian w zakresie finansów publicznych prowadzone są intensywne prace zmierzające do wprowadzenia i zastosowania nowoczesnej metody budżetowania i zarządzania finansami publicznymi – budżetu zadaniowego. Podjęcie decyzji dotyczącej wdrażania budżetu zadaniowego podyktowane zostało potrzebą poprawy skuteczności, efektywności oraz przejrzystości wydatkowania środków publicznych w kontekście wyzwań rozwojowych stojących przed Polską. Ważną zaletą budżetu zadaniowego w tym kontekście, jako narzędzia zarządzania finansami publicznymi, jest lepsza przejrzystość i zrozumiałość takiego budżetu. Budżet ten będzie stanowić ważne źródło informacji dla Państwa jako Członków Rządu, dla Ministra Finansów, urzędników i pracowników administracji rządowej biorących udział w planowaniu i wykonywaniu budżetu, polityków uchwalających ustawę budżetową (parlament), a ponadto dla instytucji monitorujących, ewaluujących oraz kontrolujących wykonanie wydatków publicznych. W związku z powyższym ważne jest przeanalizowanie zagadnień związanych z budżetem zadaniowym w kontekście globalnego spojrzenia na proces ewaluacji.

Budżet zadaniowy jest więc nową metodą zarządzania finansami publicznymi stosowaną i wprowadzaną w systemach finansów publicznych wielu krajów członkowskich OECD, a także zalecaną (m.in.) przez Komisję Europejską. Budżet w układzie zadaniowym w sposób naturalny wpisuje się w szeroki proces ogólnosiękatowego trendu głębokiej transformacji administracji publicznej oraz wprowadzania na szeroką skalę podejścia określanego mianem Nowego Zarządzania Publicznego (*New Public Management* – NPM)¹.

Ten sposób ujmowania w terminologii budżetu zadaniowego jest bliski definicji, która wykorzystywana jest w niniejszej publikacji, gdzie budżet zadaniowy traktowany jest jako metoda zarządzania środkami publicznymi, ujmowana w postaci skonsolidowanego planu wydatków jednostek sektora finansów publicznych – sporządzanego w czasie dłuższym niż rok, w układzie funkcji państwa, zadań i podzadań budżetowych wraz z miernikami określającymi stopień realizacji celu².

Mając na uwadze tak zdefiniowane pojęcie budżetu zadaniowego, w pierwszej części niniejszego rozdziału przedstawione zostaną podstawowe pojęcia z nim związane, następnie omówione zostaną polskie doświadczenia we wdrażaniu budżetu zadaniowego, a następnie zaprezentowane zostanie po-
wiązanie budżetu zadaniowego z procesem ewaluacji, któremu poświęcona jest cała publikacja.

¹ Nowe podejście do administracji publicznej zwane Nowym Zarządzaniem Publicznym (*New Public Management* – NPM) pojawiło się najpierw w Wielkiej Brytanii, Australii i Nowej Zelandii w latach osiemdziesiątych XX w., a od początku lat dziewięćdziesiątych rozprzeczniło się w Stanach Zjednoczonych (na temat NPM w Nowej Zelandii zob. M. Swora, *Państwo kontraktujące. Reforma sektora państwowego w Nowej Zelandii*, „Ruch Prawniczy, Ekonomiczny i Socjologiczny”, z. 4/2005, s. 149 in.). W mniejszym lub większym zakresie jego recepcja ma dzisiaj miejsce praktycznie w każdym kraju europejskim, także w Polsce.

² M. Postuła, P. Perczyński (red. naukowa), *Budżet zadaniowy w administracji publicznej*, Ministerstwo Finansów, EIPA, Warszawa 2008.

Podstawowe pojęcia związane z budżetem zadaniowym

Jak wynika z wyżej sformułowanej definicji budżetu zadaniowego, pojęciami bezpośrednio z nim powiązanymi są zagadnienia związane z terminami celu i miernika w kontekście osiągania danych rezultatów. W związku z powyższym w tej części tego rozdziału zajmujemy się tymi dwoma zagadnieniami, począwszy od miernika.

Nie można mówić o miernikach w ramach budżetu zadaniowego bez uzmysłowienia sobie, że są one częścią określonego systemu – zwanego systemem pomiaru wykonania celów/zadań (*Performance Measurement System*). W niniejszym rozdziale odnosimy się będziemy do kategorii zadań, tym niemniej przedstawiane rozważania mają zastosowanie do kategorii podzadań i działań wyodrębnianych w ramach polskiej metodologii budżetu zadaniowego.

Pomiar wykonania zadań jest ciągłym procesem monitorowania i raportowania w zakresie osiągnięć danego zadania lub programu³, dotyczących w szczególności postępu w realizacji określonych wcześniej celów. Mierniki wykonania zadań dotyczyć mogą rodzaju lub poziomu działalności (procesów), bezpośrednich produktów i usług tworzonych przez program (wyników) i/lub rezultatów związanych z tymi produktami oraz usługami (skutków). Przez program należy rozumieć działalność, projekt, program wieloletni w rozumieniu ustawy o finansach publicznych⁴, funkcję lub politykę, która ma zdefiniowany cel lub zbiór celów. Inaczej ujmując, pomiar wykonania celów/zadań jest instrumentem służącym ocenie postępu w stosunku do założonych celów danego programu lub zadania. Obejmuje on:

- gromadzenie danych dotyczących „procesu produkcji”, który obejmuje działania zmierzające do przekształcenia nakładów w wyniki, tj. dobra lub usługi świadczone zazwyczaj na zewnątrz jednostki sektora finansów publicznych na rzecz obywateli,
- ocenę skutków – ekonomicznych lub społecznych zmian wynikających z danego zadania lub programu oraz porównywanie ich z założonymi celami.

Celem systemu pomiaru wykonania zadań jest: poprawa zarządzania, zwiększenie rozliczalności oraz poprawa procesu podejmowania decyzji – zarówno w czasie teraźniejszym, jak i przyszłym⁵.

Mierniki, gdy są we właściwy sposób sformułowane, dostarczają informacji o stopniu realizacji celów/zadań (*Performance Information*), która wykorzystywana jest w procesie zarządzania zadaniowego jednostkami sektora finansów publicznych (*Performance Management*). Niniejsza informacja zadaniowa wykorzystywana jest do:

- pomiaru postępu w osiąganiu celów strategicznych, taktycznych i operacyjnych,
- promowania rozliczalności jednostek sektora finansów publicznych,
- porównywania kosztów wykonania zadań w celu szukania okazji do poprawy,
- uczenia się jednostek od innych jednostek.

Proces pomiaru stopnia realizacji celów składa się z następujących elementów: celów wykonania zadań, czyli oczekiwanych (pożądanych) skutków działań jednostek sektora finansów publicznych; mierników wykonania celów/zadań, tj. ilościowych lub jakościowych cech wykonania lub oczekiwanego wykonania zadań, tj.:

- pożądanego poziomu wykonania zadań w odniesieniu do każdego z wykorzystywanych mierników,
- rezultatów wykonania zadań, czyli bieżącego poziomu wykonania zadań w odniesieniu do każdego z wykorzystywanych mierników.

³ Dany program składać się może z kilku zadań, podzadań bądź, w polskiej metodologii, może być równoznaczny z danym zadaniem państwa.

⁴ Ustawa z dnia 30 czerwca 2005 r. o finansach publicznych (Dz.U. Nr 249, poz. 2104, z późn. zm.)

⁵ Podobne cele i zadania mają również monitoring oraz ewaluacja programów.

Każda jednostka sektora finansów publicznych przystępując do pomiaru wykonywanego przez siebie stopnia realizacji celów musi sobie odpowiedzieć na dwa podstawowe pytania:

- jakich obszarów jej działalności powinny dotyczyć mierniki stopnia realizacji celów?
- jakie elementy tej działalności powinny być mierzone?

Mierniki stopnia realizacji celów powinny być skoncentrowane na tych obszarach działań jednostki, które są dla niej najważniejsze. Dlatego też obszary te muszą mieć jasno zdefiniowane cele oraz sposoby ich osiągania. Jednostka powinna również posiadać odpowiedni system informujący ją o tym, czy i na ile cele te są osiągnięte.

Mierniki stopnia realizacji celów powinny koncentrować się na wszystkich poziomach inicjatyw podejmowanych przez jednostki, które są odpowiedzialne za osiąganie zamierzonych celów. Mierniki wysokiego poziomu powinny dotyczyć ogólnych aspektów realizacji założonych celów. Natomiast mierniki na niższych poziomach np. prowadzonych działań powinny dotyczyć obszarów działalności taktycznej, operacyjnej (bieżącej).

Jednostka sektora finansów publicznych powinna również przestrzegać zasady, aby starać się zmierzyć to, co jest w jej działalności najważniejsze i co powinno być mierzone, zamiast koncentrować się na tym, co jest łatwe do zmierzenia.

Czym zatem jest cel, w kontekście budżetu zadaniowego, skoro wszystkie rozważania prowadzone powyżej na temat pojęcia miernika pozwalają na konkluzję, że cały proces rozpoczyna się od właściwie zdefiniowanego celu.

Według najbardziej ogólnej definicji, cel to pożądaný przyszły stan rzeczy, który pragnie osiągnąć osoba lub organizacja. Innymi słowy, jest to pożądaný stan docelowy podejmowanych działań.

Za najważniejsze generalne zasady, których powinno się trzymać przy poprawnym formułowaniu celów – niezależnie od tego, czy mają dotyczyć one funkcjonowania organizacji biznesowej czy instytucji sektora publicznego – uważa się powszechnie⁶:

1. **Konkretność.** Cel powinien być jasno, precyzyjnie, jednoznacznie zdefiniowany, nie powinno być w nim mglistości, ogólnikowości i niejasności. Innymi słowy, powinien być przejrzysty i czytelny dla wszystkich interesariuszy.
2. **Mierzalność.** Dobrze sformułowany cel powinien zawierać metody pomiaru (ilościowego lub jakościowego) stopnia realizacji. Cel, w przypadku którego nie istnieje możliwość pomiaru stopnia realizacji, nie ma z definicji szans spełnić swojej roli – w rzeczywistości nie może odgrywać żadnej roli motywacyjnej, a co najwyżej może stanowić ogólne określenie kierunku działań.
3. **Akceptowalność.** Wszyscy najważniejsi interesariusze powinni zgodzić się z wyznaczonym celem (w drodze odpowiedniej procedury), a przynajmniej go zaakceptować, uznając jednocześnie za możliwy do osiągnięcia. W przeciwnym razie realizacja celu może być utrudniona lub uniemożliwiona przez prawdopodobny sabotaż ze strony niektórych ważnych uczestników realizacji.
4. **Realistyczność.** Cel musi być sformułowany w sposób realistyczny i dający szansę realizacji, a jednocześnie musi istnieć przekonanie, że realizacja celu ma sens z punktu widzenia potrzeb.
5. **Ograniczenie czasowe.** Osiągnięcie celu musi być zaplanowane w rozsądnym horyzoncie czasowym. Brak określenia horyzontu czasowego (istnienia jasnego terminarza osiągnięcia celu) oznacza, że cel najprawdopodobniej nigdy nie zostanie zrealizowany⁷.

⁶ por. E. M. Goldratt, J. Cox, *The Goal by Eliyashu*, Paperback, Revised, 1992.

⁷ W. Orlowski, *Cele formułowane w budżecie zadaniowym jako niezbędny element zarządzania środkami publicznymi przez efekty*, M. Postuła, P. Perczyński (red. nauk.), *Budżet zadaniowy w administracji publicznej*, Ministerstwo Finansów, EIPIA, Warszawa 2008.

Rozwój prac nad budżetem zadaniowym

Decydując się na wdrożenie budżetu zadaniowego, dokonuje się nie tylko wyboru strategii implementacji tego narzędzia. Należy odpowiedzieć sobie również na pytanie do czego chcemy je wykorzystywać oraz w jaki sposób. Znane są bowiem różne sposoby wykorzystywania budżetu zadaniowego w zależności od tego, na jakim etapie jego wdrożenia dany kraj się znajduje (a często również – na jakim pragnie poprzestać). Zgodnie z metodologią przyjętą i rozpowszechnianą przez OECD w procesie implementacji budżetu zadaniowego można wyróżnić trzy zasadnicze stadia użyteczności tego narzędzia⁸.

Prezentacyjny budżet zadaniowy (Presentational Budget) – w tym układzie informacje uzyskane na podstawie wykonania budżetu w układzie zadaniowym (np. stopnia realizacji celu itd.) nie przekładają się na rozstrzygnięcia dotyczące alokacji środków na rok następny. Podejście to służy jedynie poprawie przejrzystości prezentacji wydatków w celu poprawy jakości dialogu społecznego pomiędzy społeczeństwem, a politykami i urzędnikami, tworzącymi projekt budżetu w sposób bezpośredni, nie zwiększa jednak efektywności wydatkowania środków publicznych.

Budżet zadaniowy informujący o efektywności (Performance-informed Budget) – ten etap implementacji/wykorzystania budżetu zadaniowego charakteryzuje się tym, iż nie ma on wymiaru jedynie prezentowania wydatków publicznych w podziale na zadania wraz z opisem celów i mierników, ale odnosi się również do zakładanych wyników i rezultatów wykonania w przyszłości. W tym modelu obserwuje się już powiązanie informacji płynących z wyników ewaluacji z decyzjami alokacyjnymi⁹. Jednakże nie ma w nim wciąż mechanizmów automatycznych lub technicznych powiązań pomiędzy celami, osiągniętymi efektami ich realizacji a finansowaniem (alokacją)¹⁰.

Budżet zadaniowy efektywnościowy (Performance Budget) – to docelowe, najwyższe stadium rozwoju modelu budżetu zadaniowego. W ramach takiego stadium rozwoju istnieje bezpośrednie przełożenie pomiędzy alokacją środków na zadania w odniesieniu do osiągniętych efektów mierzonych stopniem realizacji celów, przez poszczególnych dysponentów.

Polskie doświadczenia w zakresie budżetu zadaniowego

W najbliższych latach Polska, tak jak i wiele innych krajów nowo przyjętych do Unii Europejskiej, stanie w obliczu przełomowego wydarzenia – przyjęcia wspólnej waluty krajów europejskich. Pomyślne sprostanie przez Polskę niełatwym wymogom, które wiążą się z tym wydarzeniem, będzie oznaczać dla naszego kraju wejście na ścieżkę zupełnie nowych, nieosiągalnych dotąd możliwości rozwoju gospodarki polskiej.

Priorytetowym obecnie zadaniem jest zapewnienie warunków sprzyjających dążeniu Polski do wypełnienia wymogów konwergencji (tj. obniżenie udziału wydatków publicznych w PKB, a także systematyczne zmniejszanie deficytu sektora finansów publicznych oraz długu publicznego w relacji do PKB), a ponadto poprawa warunków działalności gospodarczej w Polsce.

Wspomniane cele będą realizowane przez resort finansów poprzez opracowywaną obecnie reformę finansów publicznych, a także dodatkowo – poprzez liczne działania na rzecz poprawy konkurencyjności polskiej gospodarki – zapewnienia warunków dla stabilnego wzrostu gospodarczego.

⁸ *Performance Budgeting in OECD Countries*; OECD; Paris 2007.

⁹ W Polsce w tym roku ta faza budżetu zadaniowego ma zastosowanie do przygotowania projektu ustawy budżetowej na 2010 r. w odniesieniu do kategorii programów wieloletnich.

¹⁰ Używany jest głównie jako narzędzie do planowania budżetu.

Jednymi z ważnych narzędzi, mających w wydatny sposób wspomóc poprawę wyników polskiego sektora finansów publicznych, a pośrednio również wpłynąć na warunki prowadzenia działalności gospodarczej w Polsce, m.in. – wdrażana obecnie przez Ministerstwo Finansów w ramach reformy finansów publicznych – nowoczesna metoda zarządzania finansami publicznymi, a mianowicie budżet zadaniowy.

W Polsce obecnie znajdujemy się w pierwszym stadium wdrażania budżetu zadaniowego, zgodnie bowiem z art. 124 pkt 9 oraz 158 ust. 3 pkt 9 ustawy z dnia 30 czerwca 2005 r. o finansach publicznych (Dz.U. Nr 249, poz. 2104, z późn. zm.), w ramach konstruowania uzasadnienia do ustawy budżetowej na dany rok, a także rocznego sprawozdania z wykonania ustawy budżetowej, przedstawianego przez Radę Ministrów Sejmowi i Najwyższej Izbie Kontroli, do dnia 31 maja roku następującego po upływie roku budżetowego istnieje obowiązek sporządzania informacji o wykonaniu zadaniowego planu wydatków za dany rok budżetowy.

Sformułowanie w 2007 r. pierwszego dokumentu zawierającego plan części wydatków budżetu państwa w ujęciu zadaniowym (na rok 2008) było niewątpliwie wydarzeniem przełomowym dla rozwoju metodologii budżetowania zadaniowego w Polsce. Wysoka wartość tej próby, mimo trudnych do uniknięcia na tym etapie niedoskonałości, polegała na możliwości praktycznego przetestowania na gruncie krajowym rozwiązań metodologicznych zbliżonych do tych, które stosowane są w wielu krajach. Ze względu na fakt, iż nie ma możliwości bezpośredniego przeniesienia modeli zagranicznych, istniała oczywista potrzeba, aby sprawdzić, jak rozwiązania takie mogłyby funkcjonować w kontekście krajowych uwarunkowań społeczno-gospodarczych, a także specyfiki systemu polskich finansów publicznych.

W wyniku takiej próby opracowano pierwszy zarys metodologii budżetu zadaniowego, który – choć uległ nieuchronnym, typowym dla tego typu reformy zmianom o charakterze systemowym – to w wielu założeniach zachował niezmienną aktualność, stanowiąc istotny oraz trwały wkład w proces wdrażania budżetu zadaniowego w Polsce. W szczególności na aktualności nie straciły, a wręcz zostały wzmocnione takie rozwiązania jak implementacja ogólnych standardów w zakresie stanowienia celów i mierników, dążenie do objęcia budżetem zadaniowym jak największej części wydatków budżetu państwa, ustalenie dwóch podstawowych poziomów struktury – zadań i podzadań – a także dodatkowego poziomu – działań.

W zadaniowym planie wydatków na 2008 r. dysponenti poszczególnych części wykazywali swoje wydatki w ujęciu instytucjonalnym, tzn. oddzielnie – według części, za które byli odpowiedzialni. Ponadto, zadaniowy plan wydatków na rok 2008 został sformułowany, mimo początkowych zamierzeń, jedynie w układzie jednorocznym i dla wybranej grupy dysponentów. Takie rozwiązanie było właściwe dla początkowego etapu rozwoju tego systemu – wtedy, kiedy proponowane rozwiązania metodologiczne były niezmiernie trudne dla praktycznego zastosowania przez dysponentów nawet w przypadku przyjęcia planowania zadaniowego jedynie na rok najbliższy.

Taki układ powodował, iż struktura zadaniowego układu wydatków była odzwierciedleniem (przełożeniem) części wydatków (44%), zaprezentowanych w ustawie budżetowej, na ujęcie o charakterze programowym. Jest to więc pierwsza faza wdrożenia budżetu zadaniowego, która pozwala na przejście z tej formy budżetowania w kierunku takiego dokumentu, który w większym stopniu będzie pełnił funkcję narzędzia umożliwiającego ewaluację oraz wspomagającego planowanie wydatków publicznych poprzez dostarczanie przejrzystych informacji o skuteczności i efektywności wykonania poszczególnych zadań państwa, czyli drugiego stadium rozwoju.

Z uwagi na fakt, iż 2008 r. był dla znacznej ilości resortów pierwszym rokiem wdrażania budżetowania zadaniowego, większość dysponentów nie posiadała jeszcze wystarczających umiejętności w zakresie formułowania mierników w ramach zadań własnych, a także wyznaczania planowanych wartości dla oce-

ny ich realizacji. Pierwszy rok wdrażania budżetu zadaniowego był również początkowym rokiem rozwoju ogólnej metodologii mierzenia wyników wykonania zadań publicznych w zakresie wydatków budżetu państwa (realizacja tych prac została zapoczątkowana w Kancelarii Prezesa Rady Ministrów).

Z dotychczasowych doświadczeń krajów OECD, które wdrażały budżet zadaniowy, wynika, że rozwój oraz doskonalenie metod i procedur pomiaru wykonania poszczególnych elementów struktury programowej budżetu zadaniowego jest jednym z najtrudniejszych i najbardziej złożonych elementów tej reformy. Opracowanie skutecznych zasad metodologicznych w tym zakresie, a następnie ich stopniowe doskonalenie, jest procesem o charakterze długotrwałym oraz wieloetapowym. Taki stan rzeczy wynika m.in. z faktu, iż korekty wstępnie przyjętych ogólnych założeń oraz szczegółowych rozwiązań metodologicznych mogą być dokonane dopiero na podstawie wyników uzyskanych z wieloletnich obserwacji (a następnie weryfikacji) sposobu funkcjonowania przyjętych mierników w zakresie poszczególnych elementów struktury budżetu zadaniowego.

Zagadnienia związane z definiowaniem mierników są również szczególnie istotne w kontekście dokonywania ewaluacji zadań budżetowych i kontroli stopnia ich realizacji.

Mimo przedstawionych powyżej zastrzeżeń, pierwsze lata doświadczeń Polski w zakresie mierników wykazują, iż większość dysponentów skutecznie wykorzystała możliwość pełnego uczestnictwa na początkowym etapie wdrażania budżetu zadaniowego, poprzez sformułowanie mierników dla niemal wszystkich celów zadań opracowanej przez siebie struktury. Dzięki temu uzyskany został materiał, który posłużyć może do dalszych studiów nad doskonaleniem metod mierzenia wyników zadań publicznych we wszystkich funkcjach państwa. Ważnym elementem tych analiz będzie przede wszystkim kontynuacja dążeń do zapewnienia wszechstronnej spójności mierników z celami budżetowymi, których poziom osiągnięcia mają oceniać. W tym zakresie nadal potrzebna jest weryfikacja istotności przedmiotu mierzonego zjawiska dla procesu osiągania celu strategicznego. Ważnym działaniem na rzecz zapewnienia jak najściślejszego związku logicznego pomiędzy celami a miernikami powinno być dokonanie oceny, czy wybrany przez dysponenta miernik w sposób najbardziej reprezentatywny oraz wszechstronny (ze wszystkich mierników możliwych do sformułowania dla tego elementu struktury) opisuje końcowy wynik procesu realizacji celu.

Oprócz podania nazwy miernika większość dysponentów starała się w zadaniowym planie wydatków na rok 2008 umieścić również planowaną wartość dla przyjętych/skonstruowanych przez siebie mierników. Było to niezbędne nie tylko dla możliwości wykonania oceny poprawności przyjętych wariantów, ale też dla poznania możliwości oraz specyfiki stosowania narzędzi budżetowania zadaniowego w ramach poszczególnych obszarów działalności państwa. Zróżnicowana precyzja oraz szczegółowość w definiowaniu wskaźników na wszystkich poziomach budżetu zadaniowego, a także możliwość ich późniejszego pomiaru, wynikały często ze specyfiki przedmiotu zadań znajdujących się w ramach poszczególnych zadań państwa, stopnia zaawansowania systemów monitorowania tych zadań w oparciu o statystykę resortową jeszcze przed wprowadzeniem budżetu zadaniowego u poszczególnych dysponentów, czy też trudności w opisywaniu efektów niektórych zadań, w szczególności tych, które związane były z występowaniem tajemnicy służbowej.

Potencjalne możliwości rozwoju systemów mierzenia zadań publicznych sygnalizowały liczne, umieszczone w zadaniowym planie, deklaracje dysponentów, iż opracowany przez nich miernik nowego typu (mający nieraz pierwszorzędne znaczenie dla oceny skuteczności wykonywania zadań przez daną instytucję) *będzie monitorowany*. Na wstępnym etapie rozwoju budżetu zadaniowego taka inicjatywa miała istotne znaczenie z punktu widzenia rozwoju potencjału mierzenia zadań publicznych w zakresie zadaniowego planu wydatków. Warto jednak zaznaczyć, iż przy przyjęciu takiego rozwiązania brakowa-

to wymogu sporządzenia dodatkowej informacji o terminie, sposobie i możliwościach wprowadzenia takich pomiarów, a również o dodatkowych kosztach takiego przedsięwzięcia, co utrudniało ocenę celowości przyjęcia takiego rozwiązania. Przyczyna niemożności monitorowania zadeklarowanego miernika (np. z powodu braku istnienia odpowiednich procedur, kompetencji, struktur organizacyjnych, środków lub przepisów prawnych, umożliwiających zbieranie takich informacji), powinna być wykazana przez dysponenta w sprawozdaniu z wykonania zadaniowego planu wydatków na rok 2008.

Istniały też przypadki, gdzie poprawnie sformułowany cel nie posiadał w ogóle zdefiniowanej jednostki pomiaru miernika, ponieważ zjawisko objęte procesem pomiaru było opisane poprzez mierniki sporządzone dla celów niższych, składających się wspólnie na realizację tego celu jako celu nadrzędnego (np. zadanie „Nauka, kultura i ochrona dziedzictwa narodowego”). Dąży się do tego, aby rozwiązania takie (niejednokrotnie poprawne metodologicznie) były stopniowo ograniczane na rzecz ujednolicenia układu i wewnętrznej spójności dokumentu. Brak jednostki pomiaru mógł wynikać również z braku rozwinięcia nazwy miernika.

Istotną trudność sprawiał zbyt często (i sprawia niestety nadal) obowiązek ustalenia prawidłowych (lub ustalenia jakichkolwiek) wartości bazowych dla poszczególnych mierników, co powoduje trudność formułowania oceny co do wysokości poziomu założonego planu. Niemożliwość oceny stopnia postępu w realizacji zadań pozbawia budżet zadaniowy jednego z najważniejszych walorów o charakterze informacyjnym – nie można w ten sposób ocenić, w jakim stopniu przyjęty poziom wykonania jest realistyczny lub – zgodnie z terminologią noty – ambitny.

Ponadto analiza zestawienia wartości planowanych, przypisywanych poszczególnym miernikom, z poziomem wykonania tych mierników w 2008 r., co jest szczególnie istotne w kontekście ewaluacji, wykazuje często rozbieżności o różnej skali wielkości (od nieznacznych – 1–3% do sięgających nawet 50%), których przyczyny mogły wynikać z różnych czynników, takich jak np.:

- wadliwe metodologicznie sformułowanie miernika lub sposobu jego obliczania,
- nieadekwatność danego miernika do charakteru/specyfiki mierzonego przedmiotu zadania/podzadania w zadaniowym planie wydatków (włącznie z uwarunkowaniami społeczno-gospodarczymi jego realizacji),
- brak możliwości pozyskania precyzyjnych danych dla monitoringu wskaźnika wynikający z niedoskonałości/braku systemu zbierania informacji o poziomie wykonania miernika,
- przyjęcie nieodpowiedniego do posiadanych środków planu wykonania zadań/podzadań (wynik zbyt ambitny lub zakładający możliwość pewnego zrealizowania celu – przy przyjęciu poziomu na granicy dolnego progu możliwości jednostki),
- wpływ nadzwyczajnych niezależnych czynników zewnętrznych (niemożliwych do przewidzenia w momencie planowania), które w sposób istotny wpłynęły na proces realizacji (np. zbyt późne podpisanie umów z beneficjentami, niezależne od realizującego zadanie),
- brak umiejętności w planowaniu wielkości rzeczowych (brak znajomości specyfiki przedmiotu działań i zewnętrznych uwarunkowań ich realizacji), mogący wynikać z braku odpowiedniej partycypacji merytorycznych komórek organizacyjnych dysponenta, odpowiedzialnych za realizację poszczególnych zadań, z komórkami finansowo-budżetowymi,
- obiektywny brak możliwości precyzyjnego przewidywania poziomu wykonania jednego z aspektów mierzonego zadania, z przyczyn metodologicznych (występowanie stałego czynnika zewnętrznego o charakterze decydującym dla wyniku). Jednak również w tym wypadku wymierne wahania poziomu miernika, w odniesieniu do wartości zaplanowanej, nie dyskwalifikują sensowności tego typu pomiaru (np. nie jest możliwe dokładne zaplanowanie liczby uczniów-uczestników olimpiad w da-

nym roku, ale w dłuższym czasie można określić rząd wielkości takiej partycypacji niezależnie od tych wahań, w tym również jej tendencję wzrostową¹¹.

Właściwej interpretacji poprawności przyjętych rozwiązań mierników oraz sposobu planowania ich wartości, a także odpowiedniej ich ewaluacji sprzyjać może wyczerpujący oraz przejrzysty komentarz metodologiczny umieszczony w części opisowej z wykonania zadaniowego planu wydatków.

Zadaniowy plan wydatków na lata 2009–2011 został ułożony już w układzie innego typu, według 22 podstawowych funkcji państwa, które grupują wydatki wszystkich dysponentów części budżetowych według podstawowych działalności – polityk prowadzonych przez państwo.

Dzięki takiemu zagregowanemu układowi wydatków istnieje możliwość uzyskania przejrzystej informacji, jaki jest całościowy poziom finansowania poszczególnych głównych rodzajów działalności państwa. W celu uniknięcia nadmiernej szczegółowości zadań i osiągnięcia odpowiedniej przejrzystości budżetu, dla każdej z 22 funkcji wyodrębniono zadania, które w sposób syntetyczny grupują działania służące osiągnięciu jednego celu. Przyjęto, że subwencja ogólna dla jednostek samorządu terytorialnego, rezerwy celowe¹², rezerwa ogólna, składka do budżetu UE ujęte są w Funkcji 4. Zarządzanie finansami państwa tak, aby cały budżet państwa ujęty został w postaci zadaniowej. Założono, że zarówno funkcja państwa, jak i zadanie może być realizowane przez wielu dysponentów. Trzeci prezentowany poziom stanowią podzadania, które dysponenti części budżetowych sami zdefiniowali w ramach realizowanych przez siebie zadań.

Taki sposób prezentacji zwiększa przejrzystość budżetu. Korzyści wynikające z zastosowania nowoczesnego narzędzia zarządzania finansami publicznymi będą możliwe do uzyskania jedynie w przypadku, kiedy prawidłowo działać będzie mechanizm zapewniający uzyskiwanie rzetelnych i użytecznych informacji dotyczących skuteczności i efektywności poszczególnych polityk, a więc danych mówiących o stopniu osiągnięcia celów wyznaczonych dla poszczególnych obszarów działalności państwa.

Praktyczna realizacja polityki ewaluacji wydatków publicznych wymaga zastosowania zobiektywizowanego pomiaru wykonania zadań realizowanych za pośrednictwem wydatków publicznych oraz odniesienia tego pomiaru do założonych wcześniej celów – dla uzyskania oceny skuteczności i efektywności realizacji tych zadań. Ocena taka powinna być wszechstronna i precyzyjna, a w szczególności powinna odnosić się do kosztowych aspektów osiągnięcia rezultatów zadań. Kluczowym przedsięwzięciem w procesie reformy budżetowania jest stworzenie sprawnego i miarodajnego systemu mierzenia wyników zadań. Ważne doświadczenia w mierzeniu efektów wydatkowanych środków publicznych polska administracja publiczna uzyskała poprzez realizację (i ewaluację) projektów unijnych.

Dotychczasowe doświadczenia w zakresie ewaluacji pozwoliły na ukształtowanie się systemu oceny polityki strukturalnej. Jednak pomimo dynamicznego rozwoju polskiego systemu ewaluacji wspomnieć należy także o jego słabych stronach, takich jak niepełna instytucjonalizacja oraz ograniczona uwaga poświęcana w badaniach finansowym aspektom przedsięwzięć publicznych. Ostatnie z wymienionych zagadnień wskazuje na wyzwanie, jakie dla ewaluacji niesie reforma finansów oparta o wdrażania budżetowania zadaniowego. Wkomponowanie działań ewaluacyjnych w proces tworzenia i oceny budżetu zadaniowego stwarza bowiem z jednej strony przestrzeń do wykorzystania formatywnym funkcji ewaluacji w zapominanym dotąd wymiarze finansowym. Z drugiej strony integracja budżetowania i ewaluacji kładzie podstawy do trwałej modernizacji sektora publicznego poprzez rozwój kultury administracyjnej bazującej na transparentnych i opartych na obiektywniej informacji procesach decyzyjnych. W tym kontekście dodatkowego znaczenia nabiera wspomniana już kwestia mierników i sprawozdawczości.

¹¹ Informacja z realizacji wydatków w układzie zadaniowym, Rada Ministrów, Warszawa, (maj) 2009 r.

¹² Jedynie niewielka część rezerw celowych przypisana została przez dysponentów do zadań na tym etapie prac.

Poprawność metod stosowanych w tych obszarach stanowi warunek wyjściowy dla wykorzystania potencjału ewaluacji w zarządzaniu finansami publicznymi.

Przy tworzeniu zasad dla opracowywania zadaniowego planu wydatków na rok 2009 uznano, iż stosowanie mierników na najwyższym poziomie funkcji (oddziaływania), których pomiar obarczony jest wysokim ryzykiem błędu (z uwagi na realną możliwość wpływu niezależnych czynników zewnętrznych na rezultat mierzonych zadań) na wstępnym etapie rozwoju metodologii budżetu zadaniowego jest niecelowy.

W ramach prowadzonych prac nad rozwojem metodologii budżetu zadaniowego dokonano pewnego uszczegółowienia odnoszącego się do formułowania zasad w zakresie konstruowania mierników na potrzeby zadaniowego planu wydatków na 2009 r., przyjęto bowiem, iż cele i odpowiadające im wartości dysponenci określają na poziomie zadań i podzadań (dla każdego zadania oraz podzadania nie więcej, niż dwa cele i trzy odpowiadające im mierniki, dla podzadania zaś – odpowiednio – dwa cele i jeden miernik). Podkreślono przy tym, iż wszystkie mierniki powinny być konkretne, mierzalne i określone w czasie.

Zgodnie z tymi wymogami mierniki w zadaniowym planie wydatków powinny:

- być zdefiniowane w sposób umożliwiający ciągłość ich pomiaru w wieloletniej perspektywie,
- mierzyć tylko to, na co wykonawca zadania/podzadania ma realny wpływ,
- posiadać wiarygodne i szybko dostępne źródło danych,
- posiadać wartość bazową, co umożliwia ocenę postępu rzeczowego/efektywnościowego osiągniętego w wyniku realizacji zadania/podzadania¹³.

Dla zapewnienia spójności dokumentu z wytycznymi rządu o charakterze strategicznym ustalono zasadę, iż przy formułowaniu każdego z celów i mierników realizacji zadań/podzadań w budżecie zadaniowym zapewnia się ich spójność z poziomami określonymi w dokumentach strategicznych, właściwych dla polityki danego dysponenta.

Pomimo doświadczeń nabytych w latach poprzednich, a także określenia dyspozycji o charakterze metodologicznym sformułowanie mierników dla struktury zadaniowego planu wydatków na rok 2009 nadal było dla wielu dysponentów zadaniem trudnym i złożonym, dlatego też podlega ono procesowi systematycznego doskonalenia.

Przy formułowaniu mierników na potrzeby zadaniowego planu wydatków na lata 2009–2011 nadal występowały bowiem przypadki, kiedy problemem dla dysponentów było zreorientowanie sposobu myślenia o mierzeniu wyników realizacji zadań ze skuteczności wydatkowania zaplanowanych środków – na myślenie o stopniu osiągnięcia założonych celów. Dlatego też zdarzało się, iż stosowane były mierniki obrazujące jedynie sprawność wydatkowania zaplanowanych wcześniej środków.

Miały miejsce też sytuacje, że dobre zastosowanie mierników łączy się z brakiem ogólnego komentarza natury metodologicznej omawiającego sposób pomiaru badanego zjawiska. Dotyczy to w szczególności potrzeby sprecyzowania w części opisowej, na jakich zasadach konstruowane są mierniki procesu – tj. w jaki sposób oceniany będzie stopień (%) wykonania zadań/podzadań w poszczególnych latach realizacji.

Często dla opisu sprawności działania danej instytucji państwa wystarczające są proste mierniki produktowe – szczególnie na poziomie podzadania/działania, zdarzają się jednak sytuacje, gdzie istnieją – niewykorzystane i oczywiste możliwości stworzenia bardziej złożonych wskaźników rezultatu na poziomie zadań.

¹³ Rozporządzenie Ministra Finansów z dnia 9 maja 2008 r. w sprawie szczegółowego sposobu, trybu i terminów opracowania materiałów do ustawy budżetowej na rok 2009.

Konieczne jest też każdorazowe podawanie pełnej treści mierników (a także celów), gdyż brak ich rozwinięcia utrudnia ich powszechne zrozumienie lub wprowadza swobodę interpretacyjną. W tym kontekście należy zwrócić uwagę, iż mierniki dotyczące zadań o charakterze wysokospecjalistycznym powinny być zrozumiałe dla odbiorcy dokumentu budżetowego. Z tego też względu warto, aby w komentarzu do takich zadań zamieszczano zawsze syntetyczny opis metodologiczny tego typu wskaźników.

Zbyt szeroko stosowane są również mierniki opierające się na wartościach logicznych, które powinny być używane przez wąską grupę instytucji państwa o bardzo specyficznym charakterze działalności. Zdarzało się też, iż mierniki nie oddają skali postępu procesu wykonania w poszczególnych latach: wartość zakładana na końcu realizacji zadania jest prezentowana we wszystkich latach realizacji zadań/podzadań.

Chociaż w 2009 r. dążono do ograniczenia w sposób istotny stosowania celów i mierzących ich mierników, na których osiągnięcie dysponenti mają jedynie ograniczony wpływ z uwagi na to, iż na ostateczny efekt zbyt silny wpływ mają czynniki zewnętrzne, nadal warto podkreślać potrzebę unikania typowych tzw. mierników oddziaływania, szczególnie na niższych poziomach struktury zadaniowego planu wydatków.

Szczególną wartość miały mierniki o charakterze kosztowym odnoszące się do wykonania konkretnych podzadań – pozwalające precyzyjnie wycenić rezultat wydatkowanych środków i takie mierniki warto promować podczas opracowywania budżetu zadaniowego na rok następny.

Najciekawsze, warte wyróżnienia przykłady sprawnego zastosowania mierników do poszczególnych zadań/podzadań można znaleźć w budżetach zadaniowych większości resortów (np. Ministerstwa Gospodarki, Ministerstwa Pracy i Polityki Społecznej, Ministerstwa Sprawiedliwości, Ministerstwa Obrony, Narodowej, Ministerstwa Nauki i Szkolnictwa Wyższego), a nawet wielu instytucji, które w tym roku po raz pierwszy podjęły trudne wyzwanie opracowania struktury zadaniowej dla swoich części budżetowych (np. Rządowe Centrum Legislacji, Kancelaria Premiera Rzeczypospolitej Polskiej, Instytut Pamięci Narodowej).

Ponadto już w samej ustawie budżetowej na 2009 rok po raz pierwszy zostały ujęte elementy budżetowania zadaniowego poprzez zaprezentowanie programów wieloletnich w układzie zadaniowym. Dysponenti wykazali, w ramach obowiązującej struktury budżetu zadaniowego (odpowiednich funkcji państwa i zadań), realizowane przez siebie programy wieloletnie na poziomie podzadania lub działania i określili cele oraz mierniki służące ocenie efektywności danego programu wieloletniego. Zgodnie z ustawą budżetową na 2009 r. na realizację 69 programów wieloletnich w tym roku budżetowym zaplanowano wydatki budżetu państwa w wysokości około 12, 6 mld zł¹⁴. Nowy sposób prezentacji programów wieloletnich, dokonany na tym etapie wdrażania budżetu zadaniowego, ma przyczynić się do większej przejrzystości informacji dotyczących realizacji tych programów, do skuteczniejszej ich kontroli, a w efekcie do bardziej celowego, skutecznego i efektywnego funkcjonowania sektora finansów publicznych. Jest to również etap niezbędny do opracowania całościowej koncepcji ewaluacji wykonywania zadań publicznych.

Dlatego też mając na względzie potrzebę uzyskania jednolicie opracowanych danych o stanie realizacji programów wieloletnich, podjęto inicjatywę wprowadzenia kwartalnej sprawozdawczości ich wyników. W I kwartale 2009 r. wydatkowano środki budżetowe na realizację 18 programów wieloletnich w wysokości 3 124,4 mln zł (tj. 8,41% planu rocznego)¹⁵.

Z analizy materiałów nadesłanych przez dysponentów, obrazujących wykonanie planu wydatków przeznaczonych na realizację programów wieloletnich w układzie zadaniowym (za I kwartał 2009 roku)

¹⁴ Załącznik 13; Programy wieloletnie w układzie zadaniowym, Nowelizacja ustawy budżetowej na 2009 r.

¹⁵ www.budzetzadaniowy.gov.pl

wynika, iż mimo krótkiego okresu przygotowań, znaczna część dysponentów potrafiła w sposób poprawny metodologicznie sformułować cele wskazując ostateczny efekt działania. Niestety należy zauważyć, że nadal dużo problemów sprawia tworzenie i stosowanie mierników. Przykładem jest prezentowany przez dysponentów miernik *stopień realizacji programu (...)*, który nie dostarcza informacji o stopniu realizacji celów. Niestety przekazane materiały na tym etapie wskazały, że dysponenci pomimo funkcjonowania w ramach finansów publicznych od wielu lat instytucji programów wieloletnich nie opracowali metodologii procesu ich ewaluacji.

W przekazanych sprawozdaniach wystąpiły przypadki, kiedy stosowano charakterystykę opisową zamiast podania wartości liczbowej miernika. Z uwagi na specyfikę realizowanych zadań w ramach programów wieloletnich, przy dodatkowym uwzględnieniu istniejącego systemu monitorowania tych programów, w większości przypadków efekty wykonania tych programów będą mierzalne dopiero w skali rocznej – nie wszystkie dane są bowiem obecnie dostępne i z tego powodu nie zostały zaprezentowane przez dysponentów w sprawozdaniu. W związku z powyższym trudno na podstawie przekazanych informacji dokonać analizy i ewaluacji osiągniętych rezultatów w stosunku do zaangażowanych środków. Sporządzone sprawozdania dostarczyły materiały, który posłużyć może do dalszych prac nad doskonaleniem metod mierzenia wyników zadań publicznych w ramach wszystkich funkcji państwa, a także doskonaleniem systemu zbierania danych oraz wykonywania monitoringu. Materiały zbierane w tym roku, odnoszące się do programów wieloletnich i analizy dokonane na ich bazie, stanowią doskonałą bazę do opracowania systemu ewaluacji wykorzystywanego w planowaniu zadaniowym.

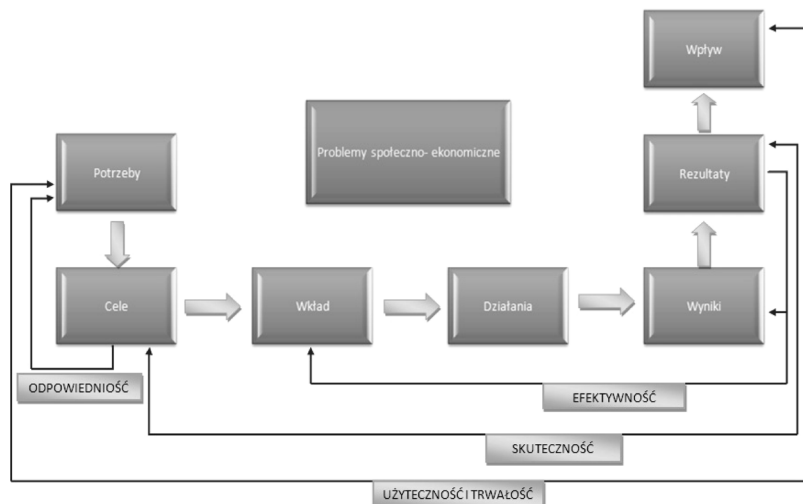
Budżet zadaniowy a kwestia ewaluacji

Podejście do pojęcia ewaluacji nie jest jednolite, co wynika między innymi z rozmaicie określanych funkcji, jakie ma odgrywać w danym procesie. Ze schematu zaprezentowanego poniżej wynika, że w swojej istocie ewaluacja wpisuje się w proces podejmowania decyzji dokonywanych przez rząd, odnosi się do spojrzenia na rozwiązanie problemów społeczno-ekonomicznych, a także opiera się na tych samych podstawowych elementach, tzn. efektywności, sprawności, przejrzystości.

Wspólnym elementem definicji ewaluacji stosowanych w ramach różnych paradygmatów jest obiektywizm dostarczanych informacji, wynikający zazwyczaj z wykorzystania warsztatu naukowego oraz oceniający charakter wniosków i rekomendacji. Ewaluacja wpisuje się tym samym w nurt działań służących racjonalizacji działań sektora publicznego oraz zarządzania publicznego opartego na wiedzy.

Na potrzeby niniejszych rozważań zaprezentowane zostanie podejście do ewaluacji jako przedsięwzięcia zmierzającego do określenia w oparciu o właściwe zgromadzone i przetworzone informacje, w jakim stopniu dane rozwiązanie (w kontekście metodologii budżetu zadaniowego, np. funkcja, zadanie, podzadanie) spełnia ustalone kryteria, w szczególności, w jakim stopniu osiągnięty został zakładany cel, dla osiągnięcia którego zostało podjęte oraz jakie są relacje pomiędzy nakładami, narzędziami jego realizacji, a także wynikami tego rozwiązania. Tak rozumiana ewaluacja wpisuje się w ciąg działań związanych z podejmowaniem decyzji o metodach realizacji danych funkcji, zadań i podzadań, jakie wykonuje państwo w stosunku do obywateli.

W cyklu tworzenia i wykonywania budżetu w ujęciu zadaniowym ewaluacja jest nie tyle zakończeniem łańcucha działań, co spoiwem pozwalającym na stałe doskonalenie prowadzonych działań zorientowanych na rozwiązanie problemów społecznych, poczynwszy od momentu podejmowania decyzji o ich rozpoczęciu. Gdyby przyjęte zostało inne założenie, mielibyśmy do czynienia wyłącznie z ewaluacją



Rys. 1. Schemat procesu ewaluacji

Źródło: Górniak J. (kierownik zespołu), *Warunki i zakres wprowadzenia audytu efektywnościowego (ewaluacji) w polskiej administracji*, Centrum Ewaluacji i Analiz Polityk Publicznych Uniwersytetu Jagiellońskiego; materiał wewnętrzny wykonany na zlecenie Ministerstwa Finansów, 2008.

ex post. Tak też jest ona lokowana na ogół w prezentacjach cyklu polityk publicznych¹⁶. W kontekście budżetu zdaniewego zasadnym jest, aby ewaluacja pojawiała się kilkakrotnie na różnych poziomach¹⁷:

- w fazie podejmowania decyzji, kiedy dokonywana jest analiza przyjętych rozwiązań, które mają służyć realizacji zadań państwa pod kątem potencjalnej zdolności osiągnięcia efektów, przy maksymalnie efektywnym wykorzystaniu środków publicznych i poprawności konstrukcji wszystkich elementów projektu – ewaluacja *ex ante*;
- w trakcie realizacji, zwłaszcza w przypadku przedsięwzięć złożonych, o długim okresie wdrożenia, gdy możliwe i celowe są korekty – ewaluacja śródkresowa lub w toku programu (*on-going*), np. w kontekście programów wieloletnich;
- po zakończeniu realizacji danego narzędzia służącego realizacji polityki państwa – ewaluacja *ex post*; jest to ewaluacja *sensu stricto* w klasycznym cyklu polityk publicznych.

Co zatem oznacza ewaluacja na każdym z wyżej przytoczonych etapów realizacji danych przedsięwzięć/programów?

Oznacza ona badanie i ocenę programu pod kątem stosownego zestawu kryteriów i wyprowadzanie wniosków dla decydentów. W przypadku dokonywania ewaluacji na różnych etapach cyklu inaczej rozkłada się akcenty, ale ewaluacja w swej istocie jest zawsze testowaniem pewnego projektu rozwiązań¹⁸. Ewaluacja niewątpliwie, jak już zostało wspomniane wcześniej, ma zastosowanie do badania polityk, programów, projektów i przedsięwzięć publicznych oraz bieżącej działalności jednostek sektora finansów publicznych podobnie jak audyt efektywnościowy. Posługują się również podobnymi instrumentami i procedurami badawczymi. Podstawowa różnica między nimi polega na tym, iż audyt efektywnościowy

¹⁶ Np. P. Bridgman, G. Davis, *The Australian Policy Handbook* (3rd ed), Allen & Unwin, Sydney 2004 [za:] J. Górniak *Ewaluacja w cyklu polityk publicznych*, [w:] S. Mazur (red.), *Ewaluacja funduszy strukturalnych – perspektywa regionalna*; Kraków 2007, s. 11.

¹⁷ J. Górniak, *Ewaluacja w cyklu polityk publicznych*, s. 11, [w:] S. Mazur (red.), *Ewaluacja funduszy strukturalnych – perspektywa regionalna*, Kraków 2007.

¹⁸ Ibidem, s. 12..

jest bardzo mocno powiązany z normatywnie pojmowaną koncepcją rozliczalności decydentów publicznych oraz jest wyrażeniem nakierowany na badanie jakości zarządzania zasobami publicznymi w kontekście budżetu zadaniowego środkami publicznymi. Ewaluacja natomiast w większej mierze zorientowana powinna być na procesy organizacyjnego uczenia się niż na kontrolę i rozliczalność, dodatkowo także szerszy jest również na ogół zakres przedmiotowy badań prowadzonych w jej ramach¹⁹. Częściej też audyt wykonywany jest przez jednostki organizacyjne (zwykle o znacznym stopniu autonomii) administracji publicznej, podczas gdy ewaluacja częściej zlecana jest na zewnątrz, niezależnym podmiotom, choć w obu obszarach występuje zarówno forma badania zewnętrznego, jak i wewnętrznego.

Tym niemniej zarówno najnowsza literatura przedmiotu, jak i faktyczny przebieg działań określanych jednym lub drugim mianem wskazują na bardzo daleko posuniętą konwergencję tych podejść zarówno w warstwie celów, jak i metod. Dlatego spotyka się już zamienne ich stosowanie. Różna jest geneza tych podejść, a zwłaszcza instytucjonalne osadzenie: audyt efektywnościowy jest w praktyce ewaluacją prowadzoną przez instytucje audytorskie, które rozszerzyły zakres przedmiotowy i metodologiczny swoich analiz w związku z wzrastającym naciskiem na ocenę wyników uzyskiwanych w zamian za wydatkowane środki i pojawiające się kolejne pytania dotyczące samych programów, w ramach których środki są wydatkowane²⁰. W przypadku polskich rozwiązań stosowanych w administracji publicznej dopiero nowy projekt ustawy o finansach publicznych w szerokim stopniu do instytucji publicznych wprowadza rozszerzoną definicję audytu skoncentrowanego w głównej mierze na spojrzeniu na cel, jaki postawiła sobie organizacja oraz sposób i efekt jego realizacji. Może to być jeden z bardziej istotnych kroków na przód w szerszym spojrzeniu na problem kontroli oraz ewaluacji dokonywanej w jednostkach sektora finansów publicznych.

Ewaluacja w swojej wielości nurtów obejmuje również takie, które nie mieszczą się w ramach audytu efektywnościowego: mowa o tzw. podejściu konstruktywistycznym (lub inaczej: ewaluacji IV generacji), które bywa jednak kwestionowane na gruncie samej debaty o ewaluacji²¹.

Każde zadanie państwa wykonywane w ramach gospodarki wolnorynkowej, niezależnie od jego genezy, jest swoistą „teorią”: opiera się na hipotezie, że w danych warunkach podjęcie działań wynikających z przyjętych założeń i przewidywań doprowadzi do osiągnięcia założonych skutków. Polityka prowadzona przez rząd opiera się na takich przyczynowych konstrukcjach. Nie zawsze są one wyrażane *explicite*, wręcz przeciwnie, często „ukryte” są pod powierzchnią formułowanych celów i przyjmowanych sposobów działania. (...) Po to, by przeprowadzić dobrze ewaluację, należy właściwie uchwycić logikę analizowanej interwencji, czyli jej „istotę”. „Istota danego zadania”, to założenie co do związków przyczynowych pomiędzy podejmowanymi działaniami a pożądanymi efektami, której towarzyszy na ogół zbiór przekonań dotyczących najbardziej właściwych sposobów wykonania danego zadania. Efektem takiej „teorii wdrożenia” są przyjmowane rozwiązania w zakresie struktur wdrożenia (przez władze centralne czy też jednostki samorządu terytorialnego), sposobu zarządzania, kontroli zarządczej i finansowej itp. Rozpoznanie „teorii wdrożenia” (realizacji danego zadania) następuje wśród osób projektujących ją i decydentów, co pozwala na lepsze zrozumienie polityki czy programu i jej pogłębioną ewaluację, a także przyczynia się do najefektywniejszej alokacji środków publicznych (a to jest istotą budżetu zadaniowego). Ewaluacja, w szczególności końcowa, sprawdza, czy w świetle uzyskanych efektów teoria programu daje się utrzymać, czy też należy ją odrzucić. Koncepcje działania, które pomyślnie przejdą ewaluację,

¹⁹ J. Górniak (kierownik zespołu), *Warunki i zakres wprowadzenia audytu efektywnościowego (ewaluacji) w polskiej administracji*, Centrum Ewaluacji i Analiz Polityk Publicznych Uniwersytetu Jagiellońskiego; materiał wewnętrzny wykonany na zlecenie Ministerstwa Finansów, 2008.

²⁰ Ibidem.

²¹ Ibidem.

tworzą korpus wiedzy podnoszącej skuteczność polityk publicznych. Faktycznie, spontaniczny proces weryfikacji występuje zawsze, tak czy owak, w postaci społecznej oceny skutków prowadzonych polityk. Ewaluacja jest metodologicznie kontrolowanym i usystematyzowanym sposobem takiej weryfikacji, prowadzonym przez ekspertów. Z natury rzeczy zatem przyczynia się do legitymizacji lub delegitymizacji efektywnościowej sposobu rządzenia²².

Decyzje w zakresie polityki społeczno-gospodarczej powstają w wyniku splotu inicjatywy i twórczych idei polityków ze zracjonalizowanym, opartym o wiedzę ekspercką działaniem administracji i ograniczeniami narzucanymi przez warunki działania i posiadane zasoby. Cele polityk publicznych nie wynikają automatycznie z diagnoz prowadzonych przez surowych, obiektywnych analityków. Cele są wynikiem wizji politycznych, ideologii rozwojowych, gry interesów i konkurencji politycznej. Eksperska diagnoza społeczna i analiza polityk publicznych powinna sprawdzać prawidłowość diagnoz i definicji sytuacji przyjmowanych przez polityków, ustalać zależności przyczynowe, bilansować zasoby niezbędne do działania, rysować scenariusze wynikające z proponowanych działań. Stanowią jednak zawsze tylko wsparcie dla inicjatywy polityków²³. W tym kontekście doświadczenia międzynarodowe i polskie jasno wskazują, iż podczas opracowywania budżetu zadaniowego warto kierować się następującymi założeniami: każde nowe zadanie powinno być poprzedzone ewaluacją *ex ante* (*white papers*), budżet powinien realizować swój podstawowy cel: uzasadnienie dla alokacji i kontrola jej wykonania, efektywność musi być podnoszona przy wykorzystaniu dodatkowych instrumentów (sam budżet zadaniowy nie zapewnia podniesienia efektywności).

Ponadto przed rozpoczęciem formułowania struktury budżetu zadaniowego należy zapewnić, aby w ramach ewaluacji *ex ante* warto rozważyć następujące aspekty finansowania poszczególnych polityk:

- jakie przeszkody w realizacji celów polityk można zidentyfikować?
- jakie są przyczyny tych przeszkód?
- dlaczego rząd uważa się za odpowiedzialny za rozwiązanie tego problemu (usunięcie tych przeszkód)?
- jakie cele mogą być sformułowane?
- jakie narzędzia mogą być użyte dla realizacji tych celów?
- jakie są przewidywane koszty realizacji celów?

Jedną z najsilniejszych zalet chociażby niderlandzkiego budżetowania zadaniowego jest doskonale zorganizowany system ewaluacji *ex post* i *ex ante*, w którym to systemie kluczową rolę pełni The Netherlands Court of Audit (Niderlandzki Trybunał Obrachunkowy – instytucja kluczowa w procesie ewaluacji holenderskiego budżetu zadaniowego).

Misją Trybunału Obrachunkowego jest „wykonywać audyt i poprawiać wydajność państwa i państwowych osób prawnych”. Trybunał dostarcza sprawozdań zawierających wyniki audytu Parlamentowi, a także osobom odpowiedzialnym za kierowanie audytowanymi instytucjami. Sprawozdania takie zawierają spostrzeżenia, będące rezultatem audytu, wnioski i rekomendacje. Trybunał tylko jedną trzecią całości swoich działań poświęca tradycyjnym audytom finansowym, a dwie trzecie swojej aktywności przeznaczają na audyty efektywnościowe (*value for money audits*). To, iż relatywnie mało czasu zajmuje Trybunałowi działalność w zakresie tradycyjnych audytów finansowych, uzasadnione jest tym, iż każdy resort posiada wewnętrzną komórkę audytu.

²² Por. J. Górniak *Ewaluacja w cyklu polityk publicznych*, s. 12, [w:] S. Mazur (red.), *Ewaluacja funduszy strukturalnych – perspektywa regionalna*, Kraków 2007.

²³ Ibidem, s. 13.

Ocena wykonania zadań państwa jest oceniana z użyciem danych, które są zbierane i analizowane każdego roku przez główną instytucję audytowo-kontrolną (*Court of Audit*) pod kątem trzech następujących zasad:

- ścisłego powiązania pomiędzy polityką danego obszaru a jego finansowaniem (budżetem),
- silniejszej zależności pomiędzy budżetem a sprawozdawczością,
- bardziej wnikliwej analizy relacji pomiędzy celami, środkami i wykonaniem²⁴.

Nowy system budżetowania zadaniowego oparty będzie o zestaw pytań ewaluacyjnych, który pomaga weryfikować poprawność konstrukcji i wykonania danego przedsięwzięcia budżetowego.

Aby ocenić, czy dana interwencja (działanie) w ramach poszczególnych polityk ma sens, należy przeprowadzać rzetelne analizy *ex ante*. Przewidywane wyniki interwencji należy porównywać z wynikami przewidywanymi w danym obszarze, kiedy takie wsparcie nie zostałoby wykonane. Wtedy uzyskuje się obiektywną prognozę, czy dane działanie odniesie oczekiwany rezultat.

Niezwykle ważne z punktu widzenia ewaluacji budżetu zadaniowego jest to, aby kształtować realistyczne oczekiwania dotyczące możliwości dostarczania informacji o produktach i rezultatach. Ponadto informacje te powinny być „technicznie aktualne”, a także odpowiednie dla zastosowania z punktu widzenia ich sprawozdawczości przez urzędników/menadżerów. Ewaluacje powinny zaopatrywać decydentów w informacje wiarygodne i pożyteczne, dlatego tak ważne jest opracowanie wysokiej jakości metodologii i standardów ewaluacyjnych. Tylko wysoka jakość ewaluacji może zapewnić podstawowy cel budżetu zadaniowego – wyciąganie wniosków – tj. uczenie się przez dysponentów korygowania błędów lub niedomagań w planowaniu realizacji działań wchodzących w skład polityk państwa.

Podsumowanie

Ze względu na długotrwałość i złożoność procesu wdrażania budżetu zadaniowego, również opracowanie docelowych rozwiązań systemu jego kontroli oraz ewaluacji będzie musiało być prowadzone etapowo i będzie procesem wieloletnim. Rozpoczęło się już kształtowanie założeń systemu kontroli w kontekście nowych rozwiązań w ustawie o finansach publicznych, definiujących pojęcie kontroli zarządczej. Posiadamy już pierwsze doświadczenia odnoszące do spojrzenia na sporządzony układ zadaniowy z perspektywy jego realizacji (sporządzono pierwsze pilotażowe *Sprawozdanie za rok 2007 z wykonania zadaniowego planu wydatków dla części budżetowych – Nauka i Szkolnictwo Wyższe*). W roku bieżącym opracowane zostało sprawozdanie z wykonania zadaniowego planu wydatków za rok 2008. Dalszy rozwój systemu kontroli budżetu zadaniowego umożliwi natomiast m.in. przyjęcie wspomnianego projektu ustawy, która przekaze Ministrowi Finansów kompetencje ustawowe do określania – w drodze rozporządzenia – rodzajów, form, terminów i sposobów sporządzania sprawozdań z wykonania budżetu zadaniowego przez wskazane jednostki sektora finansów publicznych.

W tym kontekście stworzony zostanie system ewaluacji w budżecie zadaniowym, który odnosił się będzie zarówno do etapu planowania, wykonywania i sprawozdawania wydatków sektora finansów publicznych w układzie zadaniowym.

Jednym z kluczowych zadań, jakie postawione zostanie ewaluacji w budżetowaniu zadaniowym jest ustalenie związku przyczynowego między podjętą interwencją a obserwowanymi wynikami. Aby to osiągnąć, niezbędne jest dokonanie pomiaru efektu całkowitego danego zadania jako ogólnej zmiany w zakresie zdefiniowanym przez cel (zmiany na poziomie przyjętych wskaźników), jak również odseparo-

²⁴ www.budzet Zadaniowy.gov.pl

wanie zmiany niezależnej od podjętego zadania, od zmiany, którą można przypisać wpływowi realizacji danego zadania (efekt netto).

Marta Postuła – doktor nauk ekonomicznych; Dyrektor Departamentu Reformy Finansów Publicznych w Ministerstwie Finansów, uprawniony audytor wewnętrzny jednostek sektora finansów publicznych, członek Grupy Roboczej ds. Jakości Finansów Publicznych, działającej przy Komitecie Ekonomicznym Unii Europejskiej, reprezentantka Polski *Senior Budget Official* przy OECD. Swoje zainteresowania naukowe koncentruje na analizie procesu zmian w sektorze finansów publicznych i jego wpływu na wzrost konkurencyjności Polski. Autorka wielu publikacji, artykułów, rozdziałów w opracowaniach zbiorowych z zakresu finansów publicznych publikowanych w języku polskim, angielskim, koreańskim, francuskim.

Bibliografia

- Goldratt E.M., Cox J. (1992), *The Goal by Eliyahu*; Paperback, Revised.
- Górniak J. (2007), *Ewaluacja w cyklu polityk publicznych*, [w:] Mazur S. (red.), *Ewaluacja funduszy strukturalnych – perspektywa regionalna*; Kraków.
- Górniak J. (kierownik zespołu) (2008), *Warunki i zakres wprowadzenia audytu efektywnościowego (ewaluacji) w polskiej administracji*, Centrum Ewaluacji i Analiz Polityk Publicznych Uniwersytetu Jagiellońskiego; materiał wewnętrzny wykonany na zlecenie Ministerstwa Finansów.
- Orłowski W. (2008), *Cele formułowane w budżecie zadaniowym jako niezbędny element zarządzania środkami publicznymi przez efekty*, [w:] Postuła M., Perczyński P. (red. nauk.), *Budżet zadaniowy w administracji publicznej*, Ministerstwo Finansów, EIPA, Warszawa.
- Performance Budgeting in OECD Countries* (2007), OECD; Paris.
- Postuła M., Perczyński P. (red. nauk.) (2008), *Budżet zadaniowy w administracji publicznej*, Ministerstwo Finansów, EIPA, Warszawa.
- Programy wieloletnie w układzie zadaniowym; Nowelizacja ustawy budżetowej na 2009 r. – Załącznik 13. Rozporządzenie Ministra Finansów z dnia 9 maja 2008 rok.
- Swora M. (2005), *Państwo kontraktujące. Reforma sektora państwowego w Nowej Zelandii*, „Ruch Prawniczy, Ekonomiczny i Socjologiczny”, z. 4, s. 149 i n.
- Ustawa z dnia 30 czerwca 2005 r. o finansach publicznych (Dz.U. Nr 249, poz. 2104, z późn. zm.).

Część II

Ewaluacja a stanowienie prawa

Stanowienie prawa należy do pierwotnych zadań państwa, będąc jednocześnie jednym z kluczowych wątków filozofii społecznej, poczynwszy od czasów starożytnych. Nauki prawne, wraz z ich licznymi gałęziami oraz różnorodnością podejść, stanowią dziedzinę niezwykle nobliwą. W porównaniu z nimi, ewaluacja, czy też nurt polityk publicznych, jawią się jako dyscypliny dopiero dojrzewające, które kształtując dynamicznie swój warsztat, borykają się wciąż z problemem określenia własnej tożsamości. Pomimo tak jaskrawych różnic, istnieją liczne obszary, na których osadzona w wielowiekowej tradycji myśl prawna oraz typowo współczesne podejścia do problematyki państwa i jego instytucji, nie tylko mogą, ale wręcz powinny się uzupełniać.

Przykładem ciekawej syntezy tego rodzaju jest ocena skutków regulacji (OSR), mająca na celu oszacowanie oddziaływania inicjatyw prawnych na środowisko społeczno-gospodarcze. OSR rozwinęła się na przestrzeni ostatnich dekad w odrębną dziedzinę, cechującą się własną metodyką oraz zorganizowaną w specyficzny dla siebie sposób. Popularność tego podejścia wynika z podobnych przesłanek, jakie stały za rozwojem badań ewaluacyjnych. Wśród nich wskazać należy na zwiększające się systematycznie zaangażowanie państwa w nowe funkcje (zabezpieczenie społeczne, szkolnictwo, opieka zdrowotna) oraz, co się z tym wiąże, poddanie regulacjom prawnym coraz szerszych obszarów życia społecznego.

Prawodawstwo stanęło przed nowymi wyzwaniami, albowiem nigdy uprzednio relacja pomiędzy regulacją a jej praktycznym skutkiem nie była tak problematyczna. Rozwinięta metodologia ewaluacji może wydatnie wzbogacić warsztat stosowany w procesie oceny skutków regulacji, a przez to przyczynić się do wzmocnienia trafności formułowanych w tym procesie wniosków.

Z drugiej strony OSR, ze względu na specyfikę swojego przedmiotu, stanowić może interesujące poszerzenie obszaru zainteresowań dla praktyków badań ewaluacyjnych, zajmujących się obecnie w przeważającej mierze dużymi interwencjami publicznymi finansowanymi z funduszy strukturalnych.

Podejście interdyscyplinarne sprzyja w tym wypadku nie tylko rozwojowi warsztatu, ale przede wszystkim może mieć korzystne oddziaływanie z punktu widzenia poprawy zarządzania finansami publicznymi.

W obecnym kontekście istnienie sprawnego, kompleksowego systemu oceny działań sektora publicznego staje się wręcz koniecznością. Bardzo napięta sytuacja budżetowa większości krajów, w tym Polski, której świadkami będziemy przez przynajmniej kilka najbliższych lat, wymagała będzie zwrócenia zdecydowanie większej uwagi na regulację jako instrument polityk publicznych. Rządcy tracą bowiem obecnie możliwość rozwiązywania palących problemów przy pomocy obfitych strumieni wydatków. W praktyce, jedynym narzędziem, do którego w wielu wypadkach będą mogli się odwołać, pozostaną działania legislacyjne. Odpowiednia jakość oceny będzie, bez wątpienia, jednym z kluczowych warunków ich powodzenia.

Agnieszka Sternik

Ocena Skutków Regulacji – metody i podejścia na przykładzie systemu oceny skutków Komisji Europejskiej

Wprowadzenie

„Prawo europejskie to jeden z głównych wyznaczników wyjątkowego charakteru Unii Europejskiej. Musimy więc zadbać o to, by europejskie przepisy prawne i wykonawcze były dobrze ukierunkowane, wykonywane prawidłowo i na właściwym szczeblu oraz by były adekwatne do potrzeb”¹. Od czasu publikacji, w 2006, powyższa deklaracja Prezydenta Komisji Europejskiej José Manuela Barroso, nie tylko nie straciła na znaczeniu, a wręcz przeciwnie, zyskała nową moc w kontekście przygotowywania priorytetów działań dla nowej Komisji na lata 2009–2014.

Unia Europejska (UE) przez lata rozwijała złożony aparat legislacyjny, służący tworzeniu jednolitego rynku, opracowywaniu szeregu polityk UE i obecnie realizowaniu strategicznych celów, zapisanych w Strategii Lizbońskiej, związanych z rozwojem gospodarczym, społecznym i ochroną środowiska. W miarę zbliżania się do osiągnięcia tych celów stało się jednakże jasne, że sposób, w jaki przyjmowane są rozwiązania legislacyjne ma znaczący wpływ na to czy cele te mogą być zrealizowane, czy nie. W odpowiedzi na te wyzwania, w 2002 roku Komisja Europejska zapoczątkowała dalekosiężny program służący uproszczeniu i ogólnej poprawie otoczenia regulacyjnego, pod nazwą „Lepsze stanowienie prawa” (ang. *Better Regulation*). Inicjatywa ta miała na celu wzmocnienie przestrzegania i efektywności stanowionego prawa, zwiększenie korzyści dla obywateli, a także zminimalizowanie kosztów dla gospodarki i ograniczenie biurokracji, zgodnie z zasadami proporcjonalności i pomocniczości UE. Postępując wedle powyższych zasad, Wspólnota Europejska podejmuje działania legislacyjne tylko wówczas i tylko w takim zakresie, w jakim cele proponowanych działań nie mogą być zrealizowane w sposób wystarczający przez Państwa Członkowskie, natomiast z uwagi na skalę lub skutki proponowanych działań, mogą one zostać lepiej zrealizowane przez Wspólnotę². Prawo stanowione przez UE, nie może wykraczać poza ramy konieczne do osiągnięcia ustanowionych celów politycznych, powinno być efektywne z punktu widzenia kosztów i przyjmować „najlżejszą” formę wymaganej regulacji prawnej³. Program lepszego stanowienia prawa objął szeroki zakres działań, na które składało się przyjęcie planu uproszczenia obowiązującego prawa, rozwiniecie alternatyw dla aktów prawnych, takich jak samoregulacje lub współtworzenie prawa z udziałem zainteresowanych stron, oraz zapoczątkowanie systemu oceny skutków nowo tworzonego prawa.

¹ J.M. Barroso, [w:] *Lepsze stanowienie prawa w prostszych słowach*, Komisja Europejska, 2006.

² Artykuł 2 Traktatu o Utworzeniu Wspólnoty Europejskiej, [w:] *Wersje skonsolidowane Traktatu o Unii Europejskiej i Traktatu Ustanawiającego Wspólnotę Europejską*, Dziennik Urzędowy Unii Europejskiej (C 321) 29.12.2006.

³ Komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów „Realizacja wspólnotowego programu lizbońskiego: Strategia w zakresie uproszczenia otoczenia regulacyjnego” KOM(2005) 97 wersja ostateczna.

Spojrzenie w przeszłość – krótka historia rozwoju systemu oceny skutków w Komisji Europejskiej

Stworzenie sformalizowanego i zintegrowanego systemu oceny skutków, dzięki któremu możliwe stało się wprowadzanie uzasadnionych zmian do nowych propozycji legislacyjnych jeszcze w fazie ich przygotowywania, było praktycznym wyrazem podejmowanych przez Komisję Europejską wysiłków do poprawy i wzmocnienia wewnętrznych zdolności regulacyjnych. Rozwój zintegrowanego systemu oceny skutków w KE należy analizować przez pryzmat dwóch równolegle toczących się procesów.

W 2001 roku Komisja opublikowała Białą Księgę na temat zarządzania w Unii Europejskiej⁴, w której podkreśliła potrzebę poprawy procesu legislacyjnego w instytucjach unijnych oraz uproszczenia obowiązującego systemu prawnego. W tym samym roku na szczycie Unii Europejskiej w Göteborgu Rada Europejska przyjęła Strategię Zrównoważonego Rozwoju, w ramach której zobowiązano się do analizy skutków ekonomicznych, społecznych i środowiskowych nowych propozycji legislacyjnych⁵. Na podstawie publicznej debaty, zapoczątkowanej Białą Księgą oraz konkluzjami ze szczytu w Göteborgu, powstał zarys Planu Działania⁶, który obok postanowień dotyczących uproszczenia i ograniczenia obowiązującej legislacji, zapowiadał stworzenie skonsolidowanego systemu do prowadzenia oceny skutków najważniejszych inicjatyw legislacyjnych. Motorem całego przedsięwzięcia, jak to zostało określone w późniejszym Komunikacie Komisji, było „polepszenie jakości i spójności istniejącego systemu tworzenia prawa poprzez skupienie się na głównych inicjatywach prezentowanych w corocznym planie legislacyjnym Komisji Europejskiej”⁷.

W konsekwencji połączenia, z jednej strony postępującego ruchu związanego z inicjatywą „Lepszego stanowienia prawa” oraz zapisów Strategii Zrównoważonego Rozwoju; a w późniejszym okresie także postanowień Strategii Lizbońskiej dla Wzrostu i Zatrudnienia stworzony system oceny skutków wychodził poza typową analizę aspektów związanych z obciążeniami administracyjnym wynikającymi z ustawodawstwa dla przedsiębiorstw. System oceny skutków, który wybrano w Komisji, miał charakter „zintegrowany”. Przyjęta wówczas i obowiązująca do dziś procedura wymaga oceny ewentualnych rozwiązań, w tym rozwiązań pozalegisacyjnych oraz analizuje ich pozytywne i negatywne skutki we wszystkich trzech kluczowych obszarach z punktu widzenia zrównoważonego rozwoju: ekonomicznym, społecznym i środowiskowym. Zakres dokonywanej oceny uzależniony jest od przedmiotu i zasięgu oddziaływania projektowanego aktu normatywnego i poprzedzony jest obowiązkowymi konsultacjami społecznymi. Nowe, zintegrowane podejście stopniowo zastąpiło funkcjonujący wcześniej rozproszony system ocen sektorowych, które niejednokrotnie analizowały jedynie w sposób prowizoryczny i selektywny skutki nowych propozycji prawnych.

Obecny system oceny skutków rozwijał się w trzech głównych fazach, wyznaczonych przez kolejne korekty i aktualizacje „Wytycznych dotyczących oceny skutków” (ang. *Impact Assessment Guidelines*)⁸ w 2005 i 2008 oraz ustanowienie Rady ds. Oceny Skutków (ang. *Impact Assessment Board*) w 2006 r. Ostateczna konstrukcja systemu oceny skutków została poprzedzona analizą istniejących rozwiązań instytucjonalnych w krajach członkowskich Unii Europejskiej i Organizacji Współpracy Gospodarczej i Rozwoju (OECD), która pokazała, iż w wielu z nich oceny skutków stanowią od lat zinstytucjonalizowaną praktykę,

⁴ European Commission, White Paper on European Governance, COM (2001) 428 final.

⁵ Komisja Europejska, Komunikat Komisji „Zrównoważona Europa dla Lepszego Świata: Strategia Zrównoważonego Rozwoju Unii Europejskiej”, KOM(2001) 264 wersja ostateczna.

⁶ Action Plan, „Simplifying and improving the regulatory environment”, COM(2002) 278 final.

⁷ European Commission, Communication from the Commission on Impact Assessment, COM(2002) 276 final.

⁸ European Commission, Impact Assessment Guidelines, SEC(2009) 92.

jednakże różnią się znacząco między sobą zakresem prowadzonej oceny. O ile w wielu krajach obowiązek wykonywania oceny skutków dotyczy przede wszystkim projektów nowych aktów normatywnych, system obowiązujący w Komisji jest szerszy poprzez włączenie do niego również propozycji pozalegisłacyjnych, takich jak białe i zielone księgi oraz komunikaty. W przypadku wielu krajów analiza skutków przeprowadzana jest wyłącznie w obszarze wpływów ekonomicznych, Komisja tymczasem przywiązuje szczególną wagę do tego, aby analiza była zrównoważona, jeśli chodzi o wszystkie trzy obszary. Pod tym względem można mówić o pewnej wyjątkowości i kompletności systemu przyjętego w Komisji. Elementem łączącym system przyjęty przez Komisję i rozwiązania krajowe jest natomiast to, że stoi za nimi przekonanie, iż najskuteczniejszym sposobem na zapewnienie wysokiej jakości tworzonego prawa jest zobowiązanie osób przygotowujących propozycje legislacyjne do oceny ich potencjalnych skutków. Wkład Komisji w zintegrowaną ocenę skutków umożliwia zatem decydentom dokonywanie wyborów na podstawie solidnych dowodów i gruntownej analizy dostępnych alternatyw oraz ustalania kompromisów tam, gdzie to stosowne⁹.

Kluczowe elementy systemu oceny skutków w Komisji

Przedstawiając system oceny skutków w Komisji Europejskiej konieczne jest zwrócenie uwagi na te elementy, które wpływają zasadniczo na kształt oraz skuteczność tego procesu.

Wybór inicjatyw legislacyjnych do oceny skutków

W systemie, który swoim zakresem nie obejmuje w sposób wyczerpujący całego nowo tworzonego ustawodawstwa, kluczowym stało się określenie zasad wyboru inicjatyw, które poddawane są szczegółowej ocenie skutków. W początkowej fazie działania systemu, w latach 2003–2005, procedura zakładała podejście złożone z dwóch etapów. Po pierwsze wszystkie propozycje legislacyjne zapisane w tworzo- nym co roku Programie Działalności Legislacyjnej Komisji (ang. *Commission Legislative Working Programme – CLWP*) poddawane były ograniczonej ocenie (1–2 strony), która obejmowała analizę problemu, celów, możliwych rozwiązań oraz potencjalnych skutków. Na tej podstawie Kolegium Komisarzy decydowało o przeprowadzeniu pełnej oceny lub jej zaniechaniu. W 2005 r. zreformowano system, zastępując pierwszą, ograniczoną ocenę tzw. mapą drogową (ang. *roadmap*), która szczegółowo opisuje cele, zakres i sposób wdrożenia proponowanej inicjatywy, a przez to umożliwia zidentyfikowanie skali potencjalnych skutków propozycji. Mapa stanowi istotną wskazówkę przy podejmowaniu decyzji o konieczności przygotowania oceny skutków.

W „Wytycznych dotyczących oceny skutków” Komisja, na pierwszym miejscu, zobowiązała się do przeprowadzania analizy inicjatyw, które zgłoszone są w rocznym programie działalności legislacyjnej. W praktyce wymóg ten został rozszerzony, obejmując oceną wszystkie te inicjatywy, również zielone księgi lub inicjatywy o charakterze quasi-legislacyjnym, co do których istnieje prawdopodobieństwo, że mogą mieć znaczący wpływ na sferę gospodarczą, społeczną bądź środowisko.

Rada ds. Oceny Skutków (ang. *Impact Assessment Board*)

W odpowiedzi na postulat polepszenia kontroli nad systemem oceny skutków zgłaszany od początku jego istnienia przez Parlament Europejski i zewnętrznych partnerów, w 2006 r. decyzją Przewodniczą-

⁹ Komunikat Komisji dla Rady i Parlamentu Europejskiego, „Poprawa otoczenia regulacyjnego w dziedzinie wzrostu gospodarczego i zatrudnienia w Unii Europejskiej”, KOM(2005) 97 wersja ostateczna.

cego Komisji Manuela Barroso, powołana została Rada Oceny Skutków. Rada jest ciałem niezależnym, podlegającym bezpośrednio Przewodniczącemu KE. W skład jej wchodzi obecnie czterech członków w randze dyrektora, będących ekspertami w trzech dziedzinach podlegających ocenie: w dziedzinie skutków ekonomicznych (dyrektor z Dyrekcji Generalnej (DG) ds. Gospodarczych i Finansowych oraz DG ds. Przedsiębiorstw i Przemysłu), społecznych (dyrektor z DG ds. Zatrudnienia, Polityki Społecznej i Równych Szans) i środowiskowych (dyrektor z DG ds. Środowiska) oraz zastępca Sekretarza Generalnego KE. Od członków Rady oczekuje się, iż wniosą do jej prac swoją wiedzę i doświadczenie. Jednocześnie, jako że powoływani są *ad personam* a nie jako przedstawiciele swoich Dyrekcji, nie tylko nie mogą przyjmować z ich strony żadnych instrukcji, ale również wstrzymują się od głosu w dyskusjach, które dotyczą ocen skutków przedłożonych przez ich własną instytucję. Rada wspierana jest w swej pracy przez Sekretariat Generalny, międzyinstytucjonalną Grupę Roboczą ds. Oceny Skutków oraz przez pracowników w poszczególnych Dyrekcjach, zajmujących się oceną skutków.

Rada pełni centralną funkcję kontroli jakości przedstawianych analiz oraz ich zgodności z procedurami i normami zapisanymi w „Wytycznych dotyczących oceny skutków”. Status Rady nie przewiduje dla niej formalnej roli w procesie tworzenia prawa, a jedynie uprawnienia do sporządzania opinii na temat poszczególnych ocen skutków. W tym celu jednostki przygotowujące propozycje legislacji, do których prowadzone są oceny skutków, mają obowiązek przedłożenia raportu z oceny skutków pod dyskusję na Radzie. Rada zwraca szczególną uwagę na to, czy raport stworzony był według standardów minimalnych określonych w Wytycznych oraz czy analiza została przeprowadzona w sposób proporcjonalny, zwłaszcza w zakresie oceny skutków preferowanego i alternatywnych rozwiązań legislacyjnych. Po zapoznaniu się z raportem oraz spotkaniu z odpowiedzialną za propozycję legislacyjną i ocenę skutków jednostką, Rada przygotowuje opinię zawierającą rekomendacje identyfikujące główne obszary wymagające dalszej analizy i zmiany w raporcie.

Od momentu powołania, Rada zanalizowała wszystkie raporty z oceny skutków przeprowadzone w Komisji i przedstawiła na ich temat opinie. Tylko w 2008 roku Rada zbadała 135 raportów, czyli o 33 więcej niż rok wcześniej. W 2008 roku wzrosła również, o 10% w stosunku do roku 2007, liczba raportów, wobec których Rada zażądała ponownego przedłożenia. Uwzględniając przypadki ponownego rozpatrywania raportów, w 2008 roku Komisja wydała w sumie 182 opinie. Z powyższych statystyk widać, iż dość duży jest odsetek raportów zwracanych autorom w celu poprawy, nie powinno to być jednak traktowane jako znak pogarszającej się jakości samych ocen skutków, a raczej wzrastających wymagań Rady wobec jakości przedkładanych raportów. Obraz rysujący się na podstawie analizy szeregu opinii wydanych przez Radę wskazuje na główne słabości przedkładanych raportów, wynikające z niejasnego przedstawiania problemu, niedostatecznej lub niezrównoważonej pomiędzy trzema filarami analizy skutków oraz ograniczonej analizy alternatywnych, wobec preferowanych, legislacyjnych rozwiązań¹⁰.

Wpływ systemu oceny skutków na proces legislacyjny

Wytyczne dotyczące oceny skutków jasno określają rolę oceny skutków jako procesu wspierającego podejmowanie decyzji legislacyjnej poprzez dostarczenie wyczerpującej informacji na temat potencjalnych skutków takiej decyzji. Ocena skutków nie stanowi jednakże substytutu samej decyzji, która ma charakter polityczny. Podobnie w przypadku Rady ds. Oceny Skutków, której opinia na temat jakości raportu nie jest prawnie wiążąca i nie ma wpływu ani na ostateczny kształt raportu, ani nie może zablokować politycznej dyskusji nad propozycją aktu normatywnego, a mimo to stworzony mechanizm zapewnia,

¹⁰ European Commission, 2008, Impact Assessment Board Activity Report, SEC(2009) 55.

iz przygotowujący ocenę skutków traktują poważnie przesłuchanie w Radzie. Wynika to z faktu, iż opinia Rady, obok raportu zawierającego ocenę skutków, musi być dołączana do propozycji legislacyjnej przekazywanej do konsultacji międzyinstytucjonalnych. Dodatkowym czynnikiem wzmacniającym rolę oceny skutków w procesie legislacyjnym jest obowiązek publikacji opinii Rady¹¹. Raport z działalności Rady za 2008 rok wskazuje, iż rekomendacje Rady są powszechnie uwzględniane przez autorów raportów, którzy wprowadzają zmiany do ocen według przedstawionych zaleceń, choć zdarza się, że w niektórych przypadkach rekomendacje traktowane są selektywnie¹². Sytuacja ta nie wynika jednakże z negacji słuszności rekomendacji, a raczej z niewłaściwie zaplanowanego całego procesu przygotowywania oceny i *de facto* dużej presji czasu, której poddani są autorzy dokumentu.

Sukces systemu oceny skutków i jego realny wpływ na proces legislacyjny zależy w dużej mierze od przygotowujących raporty oraz uwzględniania przez nich rekomendacji przedstawianych przez Radę. W związku z tym, iż opinia Rady ma wyłącznie charakter konsultatywny, rodzi się pytanie czy aprobatą raportu przez Radę nie powinna być warunkiem, jedynie po spełnieniu którego możliwe byłoby rozpoczęcie konsultacji międzyinstytucjonalnych samej propozycji aktu normatywnego. Doświadczenie z okresu kryzysu na rynkach finansowych, kiedy od Komisji Europejskiej oczekiwano szybkiej reakcji i tworzenia rozwiązań legislacyjnych zapobiegających podobnej sytuacji w przyszłości, pokazało, iż nie zawsze instytucje dysponują czasem wystarczającym na przeprowadzenie dokładnej oceny skutków propozycji regulacji.

Współpraca Komisji Europejskiej z pozostałymi instytucjami EU

Komisja Europejska jest instytucją inicjującą tworzenie prawa w UE, ostateczny jego kształt nadawany jest jednakże przez Radę Unii Europejskiej oraz Parlament Europejski. Ocena skutków projektu aktu prawnego przeprowadzona przez Komisję, jak również sam projekt stanowi jedynie pierwszy etap złożonego i rozciągniętego w czasie procesu, podczas którego pierwotna propozycja podlega, niejednokrotnie, znaczącym modyfikacjom. Z tego względu, niezbędna jest bliska współpraca Komisji z Radą oraz Parlamentem Europejskim, dwoma instytucjami, które decydują o ostatecznej formie regulacji.

W tym celu, w grudniu 2003 roku trzy instytucje zawarły porozumienie określające zaangażowanie Parlamentu Europejskiego i Rady na rzecz lepszego stanowienia prawa¹³. Porozumienie ustanawia ogólną strategię dla lepszego stanowienia prawa poprzez cały proces legislacyjny UE. Jego główne elementy obejmują poprawę międzyinstytucjonalnej koordynacji i przejrzystości procesu poprzez publikację oceny skutków oraz zapewnienie stabilnych ram dla instrumentów prawa, takich jak współregulacja i samo-regulacja (ang. *soft law*). Porozumienie określa również, iż obie instytucje odpowiedzialne są za ocenę skutków zasadniczych zmian dokonywanych w stosunku do pierwotnej propozycji Komisji¹⁴. Zarówno Parlament, jak i Rada zobowiązały się jednocześnie do pełnego brania pod uwagę ocen skutków przygotowywanych przez Komisję, zachowując zarazem prawo do prowadzenia swoich własnych, dodatkowych analiz, jeśli to zostanie uznane za niezbędne.

¹¹ http://ec.europa.eu/governance/impact/practice_en.htm

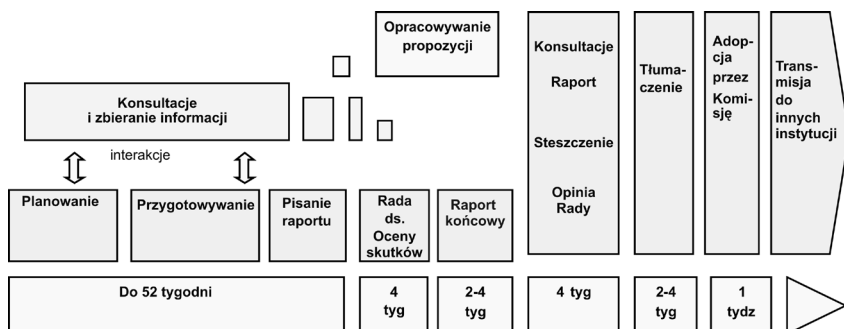
¹² European Commission, 2008, Impact Assessment Board Activity Report, SEC(2009) 55.

¹³ European Commission, 2003, Inter-Institutional Agreement on Better law-making (2003/C/321/01).

¹⁴ Pierwsza taka ocena skutków odbyła się w 2004 r. w sprawie zmian komisyjnej propozycji dyrektywy w sprawie baterii zaproponowanych przez Radę.

System oceny skutków w Komisji krok po kroku

Przygotowanie propozycji legislacyjnej jest procesem długotrwałym, a ocena skutków takiej propozycji także wymaga rozłożonej w czasie analizy. Źródłem szczegółowych informacji na temat poszczególnych etapów procesu są „Wytyczne dotyczące oceny skutków”, dlatego w tej części artykułu zarysowane zostaną jedynie najciekawsze elementy systemu przyjętego w Komisji. W całości proces został przedstawiony na rysunku nr 1.



Rys. 1. Proces przygotowania propozycji legislacyjnej

Źródło: European Commission, Impact Assessment Guidelines, SEC(2009) 92, strona 8 (tłum. autorki).

Planowanie: doświadczenie Komisji wskazuje, iż praca nad oceną skutków w sprzyjających warunkach, czyli przy braku presji ze strony ustawodawców, może trwać nawet do 18 miesięcy. Pamiętać należy jednakże, że jeszcze przed rozpoczęciem właściwej pracy przygotowana jest mapa drogowa inicjatywy, zawierająca opis ogólnych założeń dla przyszłego aktu.

Powołanie Grupy Sterującej ds. oceny skutków: przyjętą powszechnie praktyką, a od ostatniej modyfikacji Wytycznych również obowiązkiem, jest powoływanie grupy sterującej, złożonej z przedstawicieli tych instytucji, których polityki mogą być w największym stopniu narażone na negatywny bądź pozytywny wpływ przygotowywanej inicjatywy legislacyjnej. Zadaniem grupy jest dyskusja nad poszczególnymi elementami raportu, co generalnie sprzyja pełniejszej identyfikacji problemu oraz wspiera ocenę skutków poprzez rozszerzenie perspektywy analizy o wiedzę płynącą z innych obszarów.

Przeprowadzenie konsultacji: przygotowywanie oceny skutków musi iść w parze z szeroko zakrojonymi konsultacjami, które pozwolą w odpowiednim czasie zebrać informacje ułatwiające definicję problemu, doprecyzowanie sposobów działania oraz określenie potencjalnych kosztów i korzyści dla poszczególnych typów podmiotów, na które może wpłynąć przygotowywana regulacja. Konsultowanie tych, na których w sposób bezpośredni bądź pośredni może wpłynąć przygotowywana inicjatywa ustawodawcza jest obowiązkiem ustanowionym w Traktacie. Komisja Europejska przywiązuje szczególną wagę do konsultacji, a sam proces jest jednym z filarów programu „Lepszego stanowienia prawa”¹⁵. W procesie konsultacji wymagane jest, aby dotrzeć do jak najszerszego audytorium, nie tylko zatem do tradycyjnych partnerów dialogu społecznego, ale również podmiotów prywatnych, przedstawicieli władz krajowych, regionalnych i lokalnych oraz obywateli. Wytyczne Komisji rekomendują przeprowadzenie kilkustopniowych konsultacji, w różnych fazach przygotowywania raportu.

¹⁵ European Commission, Towards a reinforced culture of consultation and dialogue – General principles and minimum standards for consultation of interested parties by the Commission” COM(2002) 704 final.

Przeprowadzenie analizy, przygotowanie raportu zawierającego ocenę skutków: ocena skutków przeprowadzana jest w oparciu o różne źródła informacji. Komisja w swojej pracy legislacyjnej, w tym również na etapie przygotowywania oceny skutków korzysta ze wsparcia szeregu grup eksperckich oraz analiz i raportów przygotowywanych przez zewnętrznych konsultantów. Niejednokrotnie Komisja zleca badania, które mają dostarczyć analizy opcji, potencjalnych wpływów. Żelazną zasadą pozostaje jednakże to, iż końcowy raport przygotowywany jest przez samą Komisję, która pozostaje w pełni odpowiedzialna za to, co zamieści w raporcie i ostateczną analizę.

Identyfikacja problemu

Punktem wyjścia dla prac analitycznych, a zwłaszcza podstawowym elementem umożliwiającym określenie celów aktu, zidentyfikowanie alternatywnych opcji rozwiązań i przygotowanie rzetelnej oceny skutków regulacji jest właściwe określenie problemu. Na opis problemu powinny się składać: identyfikacja jego przyczyn, wielkości, to w jaki sposób zmieniał się w czasie, jaki jest przewidywany jego rozwój w przypadku zaniechania działania ze strony Komisji oraz uzasadnienie dla działań przez nią podjętych. Problem definiowany jest w oparciu o dostępne wyniki wcześniejszych badań i ewaluacji. Ewaluacja *ex post* programów i polityk, a zwłaszcza ewaluacja obowiązującej legislacji w przypadku proponowanych nowelizacji stanowi jedno z podstawowych źródeł informacji wykorzystywanych w tym procesie.

Określenie celów regulacji

Definicja celów planowanej interwencji ułatwia określenie ram, które pozwolą na późniejsze zweryfikowanie adekwatności poszczególnych opcji i ocenienie efektów wybranej ostatecznie opcji działania. Cele definiowane powinny być w sposób przejrzysty, w konwencji SMART¹⁶ i przy zastosowaniu hierarchizacji celów, tj. podziału na cele główne, szczegółowe i operacyjne. Cele w sposób bezpośredni powinny odnosić się do problemów zidentyfikowanych na wcześniejszym etapie. Właściwe określenie celów regulacji umożliwia późniejsze monitorowanie ich osiągnięcia poprzez system wskaźników określonych już na tym etapie i późniejszą ewaluację aktu legislacyjnego.

Określenie alternatywnych opcji realizacji celów

Mając opisany problem oraz cele, które ma osiągnąć przyszła inicjatywa, kolejnym krokiem jest zidentyfikowanie potencjalnych opcji, za pomocą których będzie można zrealizować cele regulacji. Proponowane sposoby działania powinny jednakże być blisko związane ze zdefiniowanymi uprzednio celami interwencji.

W przypadku inicjatyw podejmowanych przez Komisję Europejską, kluczowym elementem analizy jest uzasadnienie konieczności działania na poziomie UE, zgodnie ze wspomnianą już zasadą subsydiarności. Na szczeblu europejskim istnieją różne rodzaje aktów prawnych, takie jak rozporządzenia i dyrektywy, dostępne są także rozwiązania alternatywne w stosunku do typowych regulacji. Współregulacje (powierzenie zadania osiągnięcia celów określonych w prawie partnerom społecznym) oraz samoregulacje (dobrowolne umowy między organami zawierane w celu rozwiązania problemów) mogą w sposób bardziej opłacalny i skuteczny prowadzić do osiągnięcia określonych celów. Każde z tych możliwych sposobów działania ma potencjalnie różną skuteczność osiągnięcia celów, co także powinno być wzięte

¹⁶ Konkretny (ang. *Specific*), Mierzalny (ang. *Measurable*), Akceptowalny (ang. *Acceptable*), Realistyczny (ang. *Realistic*), Określony w czasie (ang. *Time-bounded*).

pod uwagę obok potencjalnego ryzyka związanego z ich implementacją i łatwością wdrożenia. Proponowane opcje muszą być realistyczne i na tyle rozwinięte, żeby możliwe było porównanie ich z użyciem kryteriów efektywności, skuteczności i spójności z celami. W przypadku ocen skutków prowadzonych przez Komisję Europejską, obowiązkowo przeprowadza się analizę następujących opcji: zaniechania jakichkolwiek nowych działań na poziomie Unii oraz opcji związanej z dalszym wzmocnieniem wdrażania już obowiązującego prawa.

Ocena skutków

Właściwa ocena skutków obejmuje analizę potencjalnego wpływu we wszystkich trzech obszarach: ekonomicznym, społecznym i środowiskowym oraz identyfikację tych osób i grup, na których planowana inicjatywa legislacyjna będzie oddziaływać w największym stopniu. Krokiem kolejnym, po zidentyfikowaniu przewidywanych skutków i ich jakościowej analizie jest wybranie tych, które wydają się być najbardziej znaczące i przeprowadzenie pogłębionej analizy umożliwiającej dokonanie ich kwantyfikacji. Wytyczne dotyczące oceny skutków zwracają uwagę na specyficzne obszary, które wymagają szczególnej analizy, między innymi obszar praw fundamentalnych, skutków społecznych, sytuacja małych i średnich przedsiębiorstw oraz zapewnić rozwój rynku wewnętrznego EU.

Odrębnym aspektem analizy jest ocena kosztów administracyjnych związanych z przestrzeganiem obowiązujących przepisów prawa. Koszty administracyjne mogą wynikać z prawnego obowiązku dostarczania informacji, etykietowania, monitorowania czy sporządzania sprawozdań, nałożonego na firmy, władze publiczne i obywateli. Jeśli obciążenia są zbyt duże, ich koszt może przewyższać korzyści, jakie płyną z wprowadzenia zasad regulujących daną dziedzinę. Komisja zwraca szczególną uwagę na analizę kosztów administracyjnych i wypracowała odrębną metodę ich pomiaru – model kosztów administracyjnych netto. Model pozwala stwierdzić, jaki udział w kosztach administracyjnych mają przepisy wspólnotowe, a jaki wynika z przepisów wykonawczych Państw Członkowskich związanych z transpozycją i wdrażaniem prawodawstwa UE. W proces zaangażowano grupę ekspertów Grupa Wysokiego Szczebla Niezależnych Partnerów ds. Obciążeń Administracyjnych¹⁷.

Porównanie opcji i rekomendacja najlepszej z nich

Krokiem prowadzącym do wyboru preferowanego rozwiązania zdefiniowanego uprzednio problemu jest porównanie możliwych opcji działania pod kątem ich przewidywanego wpływu na analizowane obszary. Zastosowanie mają tu typowe kryteria ewaluacji *ex ante*: skuteczność i efektywność opcji w wypełnianiu celów inicjatywy, spójność zewnętrzna opcji, analizowana poprzez odniesienie opcji do celów globalnych, strategii i priorytetów polityk UE, a także spójność wewnętrzna charakteryzowana przez porównanie wpływu ekonomicznego, społecznego i środowiskowego poszczególnych opcji. Istotnym narzędziem stosowanym przy porównywaniu opcji jest analiza kosztów i korzyści (*cost benefit analysis*), która jednakże może być stosowana wyłącznie w sytuacji, kiedy możliwe jest dokonanie kwantyfikacji przewidywanego wpływu i wyrażenia go w wartościach monetarnych. W pozostałych przypadkach zastosowanie ma analiza wielostronna (wielokryteriowa, *multi-criteria analysis*). Umożliwia ona porównanie opcji, w których przewidywany wpływ mierzony jest z użyciem zarówno metod ilościowych, jak i jakościowych, ocena poszczególnych opcji różni się stopniem szczegółowości analizy, a całość w ograniczonym stopniu można wyrazić w wartościach monetarnych. W wyniku analizy eliminowane są opcje, które

¹⁷ http://ec.europa.eu/governance/better_regulation/admin_costs_en.htm

w ewidentny sposób nie przynoszą korzyści netto, dla pozostałych zaś sporządza się ranking. Preferowane rozwiązanie regulacyjne jest konkretną opcją lub kombinacją elementów pochodzących z różnych opcji, wobec której wykazano istnienie największego pozytywnego wpływu.

Określenie sposobów monitorowania i oceny legislacji

Każda ocena skutków zawiera rozdział definiujący, w jaki sposób monitorowana i ewaluowana będzie regulacja po jej przyjęciu. Określenie wskaźników monitorowania jest niezbędne, by zapewnić systematyczną ocenę stopnia osiągania założonych celów w trakcie obowiązywania regulacji. Wskaźniki te powinny być w miarę możliwości oparte na istniejących źródłach danych tak, aby nie powodować zwiększenia potencjalnych kosztów administracyjnych dla instytucji publicznych i biznesu. Konsekwentnie, wymóg monitorowania i ewaluacji wpisany powinien być w tekst samej propozycji legislacyjnej, identyfikując instytucje odpowiedzialne za ich prowadzenie, a zwłaszcza zbieranie danych i sposób raportowania, oraz moment przeprowadzenia ewaluacji.

Po zakończeniu prac nad raportem z oceny skutków propozycji regulacyjnej następuje formalny proces jego przyjęcia, a następnie konsultacja propozycji legislacyjnej prowadząca do jej przyjęcia przez Komisję. Rysunek 2 przedstawia kolejne elementy tego procesu.



Rys. 2. Proces prowadzący do przyjęcia propozycji legislacyjnej

Źródło: European Commission, Impact Assessment Guidelines, SEC(2009) 92, s. 10–11 (tłum. autorki).

Przedłożenie raportu do Rady Oceny Skutków, spotkanie z Radą i opracowanie ostatecznej wersji raportu – raport wraz ze streszczeniem przedkłada się pod dyskusję na Radzie Oceny Skutków, przynajmniej na cztery tygodnie przed posiedzeniem Rady. Wytyczne szczegółowo określają format raportu, na przykład to, że jego długość nie powinna przekroczyć 30 stron. Dodatkowe informacje, tabele danych, wyniki konsultacji, na podstawie których przeprowadzono ocenę mogą być umieszczone w załącznikach, zapewniając jednak, że raport jest jednolitym, kompletnym dokumentem, który dostarcza wszystkich niezbędnych informacji bez konieczności zaglądania do załączników.

W pierwszej kolejności Rada w formie pisemnej (ang. *Impact Assessment Quality Checklist*) zgłasza uwagi do treści raportu, a następnie omawia je z przedstawicielami Dyrekcji Generalnej, która przygotowuje inicjatywę i sporządza ocenę jej skutków. Po spotkaniu opracowywana jest Opinia Rady, w której, jeśli to konieczne, Rada zażądać może wprowadzenia znaczących zmian i w konsekwencji ponownego przedłożenia raportu. Jeśli taka sytuacja nie ma miejsca, przygotowuje się końcową wersję raportu zgodnie z rekomendacjami przekazanymi przez Radę.

Następuje przeprowadzenie konsultacji międzysysemowych, przyjęcie propozycji legislacyjnej przez Kolegium Komisarzy, przekazanie oceny skutków oraz inicjatywy legislacyjnej do Rady i Parlamentu Europejskiego, publikacja raportu oraz Opinii Rady ds. Oceny Skutków na stronie Komisji Europejskiej.

Ukończony raport przekazywany jest wraz z Opinią Rady oraz propozycją legislacyjną do konsultacji międzyinstytucjonalnych. To, że do propozycji legislacyjnej dołączany jest nie tylko raport, ale także jego ocena, wzmacnia rolę kontroli jakości, jaką przeprowadza Rada. Jednym z pierwszych elementów, na które zwracają uwagę przedstawiciele pozostałych instytucji, jest bowiem zgodność końcowej wersji raportu z uwagami zgłoszonymi przez Radę. Raport z oceny wpływu wraz z Opinią jest publikowany na stronach Komisji Europejskiej.

Rola ewaluacji w procesie przygotowywania oceny skutków propozycji legislacyjnych

Ocena skutków regulacji jest procesem, który z jednej strony w znacznym stopniu czerpie z metodologii, techniki i praktyki prowadzenia ewaluacji, z drugiej zaś wykorzystuje badania ewaluacyjne jako narzędzie dostarczania informacji służących definiowaniu problemu, formułowaniu celów i opcji działania.

Przygotowując ocenę skutków wykorzystywane są te same kryteria spójności, skuteczności i efektywności oraz narzędzia zbierania i analizowania danych. Zastosowanie mają tu zarówno podstawowe narzędzia badawcze takie jak wywiady, grupy fokusowe i badania ankietowe, które służą przede wszystkim opracowaniu definicji problemu oraz prowadzeniu konsultacji grup społecznych, jak i bardziej skomplikowane narzędzia do pomiaru wpływu: modele ekonometryczne, analizy kosztów i korzyści oraz różnego rodzaju specyficzne narzędzia do prowadzenia ocen sektorowych.

Ponadto w procesie przygotowywania oceny skutków zastosowanie mają różne rodzaje ewaluacji. O ile w przypadku analizy problemu ekstensywnie wykorzystuje się ewaluację typu *on going* i *ex post*, to ocena wpływu propozycji legislacyjnej we wszystkich trzech obszarach: ekonomicznym, społecznym i środowiskowym jest procesem, który w dużej mierze odpowiada klasycznemu rozumieniu ewaluacji *ex ante*, z tą jedyną różnicą, że przedmiotem analizy nie jest konkretny program lub projekt, a propozycja aktu normatywnego.

W przypadku Komisji Europejskiej, związek oceny skutków z ewaluacją ma dodatkowo wymiar czyisto formalny. Zgodnie z obowiązującymi zasadami finansowymi, przyjęcie inicjatyw, które pociągają za sobą zobowiązanie budżetowe ze strony KE wymaga przeprowadzenia klasycznej ewaluacji *ex ante*. Jeśli jednak podjęta została decyzja o konieczności wykonania oceny skutków dla tej samej propozycji legislacyjnej oraz przeprowadzona ocena będzie w zgodzie z zasadami prowadzenia analizy *ex ante*, instytucja zostaje zwolniona z konieczności dokonywania również ewaluacji. Do pewnego stopnia zatem ocena skutków traktowana jest jako forma ewaluacji *ex ante*.

W 2009 r., w kontekście toczącej się publicznej debaty nad priorytetami kolejnej Komisji Europejskiej, rola ewaluacji w zintegrowanym systemie oceny skutków nabrała nowego wymiaru. Zauważono, iż wraz z powołaniem Rady ds. Oceny Skutków zwiększyło się znaczenie ewaluacji *ex ante* oraz stopniowo poprawił się poziom wiedzy przygotowujących oceny na temat metodologii badawczej. Znacznemu zwiększeniu uległa również liczba prowadzonych ewaluacji i ocen skutków. Tymczasem, w obliczu zwiększającej się liczby regulacji, konieczne stało się zwrócenie szczególnej uwagi na ewaluację *ex post* obowiązujących aktów prawnych, której ograniczona praktyka uznana została przez niektórych ekspertów za główną słabość obecnego systemu¹⁸. Odpowiedzią Komisji na tę krytykę jest zapowiedź wprowadzenia zasady, iż każda kolejna nowelizacja aktu prawnego poprzedzona musi być ewaluacją jego dotychczasowych rezultatów i wpływu. Wpisanie wymogu dotyczącego monitorowania i ewaluacji legislacji w trakcie jej

¹⁸ A. Sapir (red.), *Bruegel memos to the new Commission - "Europe's economic priorities 2010–2015"*, Bruegel, 2009.

obowiązywania zatem jest wyrazem zmieniającego się podejścia do stanowienia prawa, które powinno być oparte na szczegółowej analizie dotychczasowych rozwiązań i ewaluacji ich skuteczności i efektywności.

Spojrzenie w przyszłość – główne wyzwania dla obecnego systemu w Komisji

Komisja Europejska zbudowała system oceny skutków regulacji, który – pod kątem jego przejrzystości i kompletności – jest powszechnie uznawany za system na bardzo wysokim poziomie zarówno na płaszczyźnie europejskiej, jak i międzynarodowej. Wyrazem wagi, jaką Komisja przywiązuje do poprawy stanowienia prawa w UE, jest bezpośrednie odniesienie się do tego zagadnienia w „Założeniach Politycznych dla następnej Komisji” ogłoszonych przez Prezydenta Barroso we wrześniu 2009 r. Barroso nawiązując do obowiązujących już zasad dotyczących konsultacji społecznych i oceny skutków przy przyjmowaniu nowych wniosków prawodawczych oraz dążenia do znacznego uproszczenia obowiązującego prawa wspólnotowego, wskazuje na potrzebę zastosowania koncepcji lepszych uregulowań prawnych, tak aby były one skuteczne, proporcjonalne i miały szerokie zastosowanie: „lepsze uregulowania prawne powinny chronić konsumenta, skutecznie przyczyniać się do osiągnięcia celów polityki publicznej, nie utrudniając jednocześnie funkcjonowania podmiotów gospodarczych, takich jak MŚP, i nie ograniczając bezpodstawnie ich konkurencyjności”¹⁹. Jego zdaniem następna Komisja powinna wypełnić zobowiązanie do ograniczenia obciążeń administracyjnych oraz powinna, co podkreślił Prezydent Barroso: „połączyć tę wielką inwestycję w ocenę *ex ante* z równoważnym wysiłkiem włożonym w ocenę *ex post* – by zagwarantować, że nasze propozycje faktycznie przynoszą obiecywane korzyści i umożliwić sobie ich zmianę i korektę, gdyby było inaczej”. Wśród obszarów, gdzie dalszy rozwój i poprawa działania wydają się konieczne, należy zwrócić uwagę na następujące:

Niezbędne wydaje się być dalsze promowanie **tworzenia i stosowania narzędzi lepszego otoczenia regulacyjnego** na poziomie Unii Europejskiej, zwłaszcza w zakresie ocen skutków i upraszczania obecnego ustawodawstwa. Ewaluacja systemu przeprowadzona w 2007 roku zwróciła uwagę na to, że rzeczywisty wpływ przeprowadzonej oceny skutków na jakość propozycji legislacyjnych różnił się znacznie w zależności od dyskutowanych raportów, jakoś tych zaś powiązana była ściśle z rodzajem analizy, która im towarzyszyła. I tak oceny skutków propozycji aktów normatywnych były prowadzone w sposób bardziej zrównoważony i metodologicznie poprawny niż na przykład oceny dotyczące skutków regulacji niewiążących prawnie, takich jak komunikatów Komisji²⁰. W tym wypadku Rada ds. Oceny Skutków odgrywa znaczącą rolę stymulując i wspierając rozwój kultury oceny w samej Komisji. Niezbędne jest również wzmocnienie roli jednostek ds. oceny skutków w poszczególnych dyrekcjach generalnych, dalsze kształcenie osób przygotowujących oceny, a także zinstytucjonalizowanie grup sterujących procesem przygotowywania ocen.

Specyficznym priorytetem dla Komisji oraz sukcesywnych Prezydencji jest dalszy rozwój **oceny skutków społecznych** propozycji legislacyjnych. Wprawdzie w najnowszych Wytocznych, dotyczących Oceny Skutków, zawarto szczegółową listę zagadnień przedstawionych w formie pytań, które powinny być poddane analizie w momencie przygotowywania oceny skutków społecznych, konieczne jest jednak wypracowanie konkretnej metodologii, wskaźników i kryteriów pomiaru. W 2009 r. Dyrekcja General-

¹⁹ J.M. Barroso, *Założenia polityczne dla następnej Komisji*, 2009.

²⁰ The Evaluation Partnership, *Evaluation of the Commission's Impact Assessment System*, 2007.

na ds. Zatrudnienia, Polityki Społecznej i Równych Szans opracowała wstępny „Poradnik służący ocenie skutków społecznych”²¹ oraz prowadzi analizy mające na celu zidentyfikowanie metod i technik, które umożliwiłyby kwantyfikację kosztów i korzyści społecznych²². Dodatkowo, wraz z przyjęciem Europejskiej Karty Praw Podstawowych, kraje członkowskie zobowiązały się do pełnego respektowania tych praw. W konsekwencji, nieodłącznym elementem oceny skutków jest weryfikacja potencjalnego wpływu, jaki inicjatywa może mieć na prawa fundamentalne. Również ten obszar wymagać będzie dalszego rozwoju metodologicznego.

Poprawa otoczenia regulacyjnego jest kwestią, która dotyczy nie tylko Unii Europejskiej. Należy wspierać **ściślejszą współpracę Komisji z Państwami Członkowskimi w celu zapewnienia** jednolitego w całej Unii Europejskiej stosowania zasad **lepszego otoczenia regulacyjnego** przez wszystkie organy nadzorujące. Ustawodawstwo europejskie w dużej mierze podlega transpozycji przez Państwa Członkowskie co, obok krajowych inicjatyw w dziedzinie uregulowań, ma również bezpośrednie konsekwencje, nie tylko dla administracji krajowych i obywateli, ale także dla przedsiębiorstw na terenie całej Unii Europejskiej. Obecnie nie we wszystkich Państwach Członkowskich wnioski dotyczące aktów prawnych są poddawane ocenie, dlatego Komisja zaproponowała, aby poprawa otoczenia regulacyjnego stała się częścią krajowych programów rozwoju w ramach Strategii Lizbońskiej. Docelowo zakłada się, iż wszystkie państwa stworzą krajowe strategie dotyczące poprawy otoczenia regulacyjnego, a w szczególności zintegrowane systemy oceny skutków gospodarczych, społecznych i na środowisko, w oparciu o krajowe struktury dostosowane do warunków krajowych. Wsparciem dla tego procesu jest powołana w roku 2005 grupa krajowych specjalistów wysokiego szczebla, która służy rozpowszechnianiu w UE najlepszych praktyk i know-how w dziedzinie poprawy otoczenia regulacyjnego.

Spojrzenie na Polskę – w jakim kierunku podążać?

Na przestrzeni ostatnich lat przywódcy europejscy i Komisja zwracali szczególną uwagę na usprawnienie otoczenia regulacyjnego Unii Europejskiej w celu zwiększenia jego skuteczności. Komisja, Parlament Europejski i Rada zainicjowały szereg inicjatyw mających na celu kodyfikację, konsolidację i uproszczenie istniejącego prawodawstwa oraz lepszą ocenę możliwych wpływów: ekonomicznego, społecznego i środowiskowego nowych propozycji uregulowań.

Doświadczenia zgromadzone w tym procesie stanowią dobrą podstawę usprawnienia systemu Oceny Skutków Regulacji (OSR) w Polsce. System OSR zapoczątkowany został w 2001 w oparciu o rekomendacje OECD, dotyczące procesu przygotowywania się Polski do członkostwa w UE. Podstawą prawną prowadzenia OSR jest Uchwała Rady Ministrów, która stwierdza, iż ocena skutków stanowi integralną część procesu opracowywania, opiniowania, uzgadniania i rozpatrywania projektów aktów normatywnych przyjmowanych przez Radę Ministrów²³. W 2006 r. Ministerstwo Gospodarki przyjęło nowe „Wytyczne do Oceny skutków regulacji”²⁴, które w sposób szczegółowy określają zasady dokonywania OSR.

²¹ European Commission, Guidance for Assessing Social Impacts within the Commission Impact Assessment system, 2009.

²² Obecnie toczą się dwa badania: „Study on Social Impact Assessment as a tool for mainstreaming social inclusion and social protection concerns in public policy in EU Member States” oraz „Review of Methodologies applied for the assessment of employment and social impacts”. W 2008 roku zakończono badanie: „Assessing the Employment and Social Impacts of Selected Strategic Commission Policies”. Dodatkowo w 2007 na zlecenie Dyrekcji Generalnej ds. Sprawiedliwości, Wolności i Bezpieczeństwa przygotowano „Study to develop methods for the quantification of social impacts of JLS policies”.

²³ Uchwała nr 49 Rady Ministrów z 19 marca 2002 r. – Regulamin Pracy Rady Ministrów (M.P. Nr 13, poz. 221).

²⁴ „Wytyczne do Oceny skutków regulacji” zastąpiły „Metodologiczne podstawy do oceny skutków regulacji” przyjęte przez Radę Ministrów 1 lipca 2003 roku.

Obowiązujące w Polsce podejście, podobnie jak to stosowane w Komisji Europejskiej, oparte jest na zintegrowanej ocenie kosztów i korzyści, we wszystkich trzech obszarach: ekonomicznym, społecznym i środowiskowym. Swoim zakresem obejmuje także, bardzo ważną z punktu widzenia Komisji Europejskiej, analizę kosztów administracyjnych dla przedsiębiorstw, obywateli i administracji publicznej. Wstępne wyniki analizy przeprowadzanej na zlecenie Dyrekcji Generalnej ds. Zatrudnienia, Polityki Społecznej i Równych Szans wskazują jednakże na konieczność wzmocnienia elementu oceny skutków społecznych oraz systematyczne włączanie analizy kosztów i korzyści społecznych do przygotowywanych OSR²⁵.

W przeciwieństwie do Komisji Europejskiej, system polski nie zakłada jednakże odrębnego elementu niezależnej oceny jakości i rzetelności OSR przed przedłożeniem go do Kancelarii Prezesa Rady Ministrów. Częściowa ocena dokonywana jest przez samą Kancelarię, nie ma ona jednak sformalizowanej struktury i nie ma aż takiego znaczenia jak pozycja Rady ds. Oceny Skutków w Komisji. Wprowadzenie do polskiego systemu instytucji odgrywającej rolę stróża jakości przygotowywanych raportów wpłynąć mogłoby na zwiększenie przejrzystości systemu i wzmocniłoby również znaczenie OSR w procesie legislacyjnym.

Zakończenie

Rola, jaką odgrywa ocena skutków w ulepszeniu propozycji legislacyjnych oraz szerzej w całym procesie tworzenia prawa, jest dobrze rozpoznana w Komisji Europejskiej. Zapoczątkowany w 2001 roku program poprawy otoczenia regulacyjnego, którego elementem była zintegrowana ocena skutków, służyć miał zwiększeniu wzrostu gospodarczego i zatrudnienia przy jednoczesnym stałym uwzględnianiu celów społecznych i środowiskowych oraz korzyści dla obywateli i administracji krajowej pod względem poprawy zarządzania. W kilka lat od wprowadzenia zintegrowanego systemu oceny skutków trudno jeszcze w sposób jednoznaczny ocenić jego wpływ na poprawę otoczenia regulacyjnego, jednakże oczekuje się, iż ta reforma, polegająca na zapewnieniu jakości ram regulacyjnych, przyniesie korzyści wszystkim.

Zmiany wprowadzone w procesie tworzenia nowego prawa oraz uproszczenia w obszarze obowiązującego ustawodawstwa mają szansę uczynić z Unii Europejskiej nie tylko bardziej atrakcyjne miejsce inwestowania, ale również bardziej atrakcyjne miejsce pracy dla jej obywateli, poprzez pozytywny wpływ na podstawowe warunki wzrostu gospodarczego, zatrudnienia i wydajności. Poprawa jakości prawodawstwa, między innymi dzięki systematycznej ocenie *ex ante* jej potencjalnych skutków, stwarza właściwe bodźce dla przedsiębiorstw, eliminuje niepotrzebne koszty i usuwa przeszkody na drodze dostosowań i innowacji. Zapewnia również stabilność systemu regulacyjnego, a dzięki temu skuteczne stosowanie i egzekwowanie prawa w całej Unii Europejskiej. W tym kontekście jednym z ważniejszych priorytetów Komisji Europejskiej na przyszłość pozostaje dalsze promowanie rozwoju inicjatyw w zakresie poprawy otoczenia regulacyjnego zarówno na poziomie europejskim, jak i w Państwach Członkowskich.

Agnieszka Sternik – absolwentka Wydziału Socjologii na Uniwersytecie Warszawskim (2001) oraz Polsko-Francuskich Studiów Europejskich w Szkole Głównej Handlowej (2002). Laureatka Chevening Scholarship przyznawanego przez rząd brytyjski, dzięki któremu ukończyła roczne studia z zakresu badań społecznych i ewaluacji na University of Bath. Od pięciu lat zajmuje się ewaluacją programów europejskich, przede wszystkim finansowanych z Europejskiego Funduszu Społecznego (EFS). W 2004 r., jako pracow-

²⁵ European Commission, DG Employment Social Policy and Equal Opportunities, 2009, *“Review of Methodologies applied for the assessment of employment and social impacts”*. Interim report, Polish case study. Not published.

nik Instytucji Zarządzającej EFS w Polsce, stworzyła jednostkę ewaluacyjną dla EFS oraz, współpracując z Krajową Jednostką Oceny, wspierała budowę potencjału ewaluacyjnego w Polsce. Od 2006 r. pracuje w Wydziale Ewaluacji i Oceny Skutków Dyrekcji Generalnej ds. Zatrudnienia, Polityki Społecznej i Równych Szans w Komisji Europejskiej. Przeprowadza ewaluacje programów i polityk w obszarze zatrudnienia i integracji społecznej, a także współpracuje z krajami członkowskimi UE przy ewaluacji EFS. Od ponad roku zaangażowana jest w proces przeprowadzania ocen skutków propozycji legislacyjnych Komisji Europejskiej. W ciągu ostatnich 3 lat przeprowadziła wiele szkoleń z ewaluacji polityk publicznych, głównie w nowych Państwach Członkowskich i państwach kandydujących. Współautorka kilku publikacji w obszarze ewaluacji. Członek Polskiego Towarzystwa Ewaluacyjnego.

Bibliografia

- Barroso, J.M. (2009), *Założenia polityczne dla następnej Komisji*.
- European Commission (2001), White Paper on European Governance, COM (2001) 428 final.
- European Commission (2002), Communication from the Commission on Impact Assessment, COM(2002) 276 final.
- European Commission (2002), Action Plan „Simplifying and improving the regulatory environment”, COM(2002) 278 final.
- European Commission (2002), Towards a reinforced culture of consultation and dialogue – General principles and minimum standards for consultation of interested parties by the Commission” COM(2002) 704.
- European Commission (2003), Inter-Institutional Agreement on Better law-making (2003/C/321/01);
- European Commission (2007), Evaluation of the Commission's Impact Assessment System, The Evaluation Partnership.
- European Commission (2008), Impact Assessment Board Activity Report, SEC(2009) 55.
- European Commission (2009), Impact Assessment Guidelines, SEC(2009) 92.
- European Commission (2009), Guidance for Assessing Social Impacts within the Commission Impact Assessment system.
- European Commission, DG Employment Social Policy and Equal Opportunities (2009), *“Review of Methodologies applied for the assessment of employment and social impacts”*. Interim report, Polish case study. Not published.
- Komisja Europejska (2001), Komunikat Komisji „Zrównoważona Europa dla Lepszego Świata: Strategia Zrównoważonego Rozwoju Unii Europejskiej”, KOM(2001)264 wersja ostateczna.
- Komisja Europejska (2005), Komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno Społecznego i Komitetu Regionów “Realizacja wspólnotowego programu lizbońskiego: Strategia w zakresie uproszczenia otoczenia regulacyjnego” KOM(2005) 97 wersja ostateczna;
- Komisja Europejska (2005), Komunikat Komisji dla Rady i Parlamentu Europejskiego „Poprawa otoczenia regulacyjnego w dziedzinie wzrostu gospodarczego i zatrudnienia w Unii Europejskiej”, KOM(2005) 97 wersja ostateczna.
- Komisja Europejska (2006), Lepsze stanowienie prawa w prostszych słowach. Bruksela.
- Komisja Europejska (2008), Komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno Społecznego i Komitetu Regionów „Drugi strategiczny przegląd lepszego stanowienia prawa w Unii Europejskiej” KOM(2008) 32 wersja ostateczna.

Komisja Europejska (2008), Komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno Społecznego i Komitetu Regionów „Trzeci strategiczny przegląd lepszego stowienia prawa w Unii Europejskiej” KOM(2009) 15 wersja ostateczna.

Ministerstwo Gospodarki (2006), *Wytyczne do oceny skutków regulacji (OSR)*, Warszawa.

Ministerstwo Gospodarki (2009), *Zasady konsultacji przeprowadzonych podczas przygotowywania dokumentów rządowych*. Warszawa.

Sapir A. (ed) (2009), Bruegel memos to the new Commission - „Europe’s economic priorities 2010–2015”, Bruegel.

The Evaluation Partnership (2007), Evaluation of the Commission’s Impact Assessment System.

Traktat o Utworzeniu Wspólnoty Europejskiej (2006), Unia Europejska – Wersje skonsolidowane Traktatu o Unii Europejskiej i Traktatu Ustanawiającego Wspólnotę Europejską, Dziennik Urzędowy Unii Europejskiej (C 321) 29.12.2006.

Ocena Skutków Regulacji (OSR) – regulacja instrumentem polityki

Wstęp

To, że system prawny jest jednym z wręcz podstawowych obszarów realizowania celów polityki (i że jest również przestrzenią tworzenia kosztów sektora finansów publicznych), nie ulega żadnej wątpliwości. Istotniejsze problemy w pełnym zrozumieniu relacji pomiędzy polityką (w rozumieniu angielskiego określenia *policy-making*) a systemem prawnym pojawiają się wtedy, gdy próbujemy dokładniej poznać mechanizm transmisji bodźców pomiędzy tymi obszarami. Jest to trudne, gdyż system prawny i polityczny są odrębne – także w rozumieniu Luhmannowskim¹; posługują się więc odrębnym i funkcjonującym dla innych celów i różnie nakierowanym systemem komunikacyjnym. Z tego punktu widzenia za pewną ironię można uznać to, że system polityczny korzysta właśnie z systemu prawnego do oddziaływania na konkretne obszary życia społecznego i gospodarczego. Z operacyjnego punktu widzenia bowiem jest pożądane, by – jakkolwiek odrębne – kody i programowanie obu systemów były w koniecznym zakresie zbieżne, tzn. by prawo osiągało te cele, które zostały wskazane w systemie tworzenia polityk a w wymaganym zakresie droga dojścia systemu prawnego do tych celów nie odbiegała od drogi założonej politycznie. Badanie, czy tak jest stanowi istotę oceny skutków regulacji (OSR), co sprawia, że lokuje się ona w obrębie całościowego systemu ewaluacji funduszy publicznych. W kontekście tematyki niniejszego tomu istotne jest również to, że specyfika OSR pozwala wnieść unikatowy wkład w poprawę gospodarowania „groszem publicznym”.

Celem tego artykułu jest przedstawienie podstawowych problemów stosowania OSR oraz wskazanie głównych ich rozwiązań – tak na płaszczyźnie teoretycznej, jak i praktycznej (ze szczególnym uwzględnieniem tej drugiej). W opracowaniu podjęto przy tym próbę uwypuklenia tych aspektów stosowania OSR, które mają największe znaczenie dla wymiaru finansowego planowanych interwencji regulacyjnych.

Istota oceny skutków regulacji i jej relacja do polityki

Nie ma jednolitej i powszechnie uznanej definicji oceny skutków regulacji. Samo znaczenie językowe tego terminu wskazuje jednak, że chodzi o prakesologicznie ukierunkowane określenie jakiejś relacji pomiędzy regulacją (a więc komunikowanym abstrakcyjnym co do natury zasobem normatywnym) a stanem rzeczywistości, do której owa regulacja ma się odnosić (w stanie postulowanym) lub faktycznie odnosi się (w odniesieniu do oddziaływania faktycznego). Już takie określenie podstawowej właściwości OSR pozwala na sformułowanie dodatkowych wniosków dotyczących tej oceny. Pierwszy jest taki, że OSR może dotyczyć któregośkolwiek etapu poprzedzającego uchwalenie danego aktu prawnego

¹ N. Luhmann, *Law As a Social System*, Oxford University Press, Oxford, 2004, s. 76–141.

i jego wejście w życie (wtedy mamy do czynienia z OSR *ex ante*), a także stanu rzeczywistości po wejściu w życie tego aktu (co oznacza, że OSR jest *ex post*). Drugi wniosek wynikający z przedstawionej prostej definicji oceny skutków regulacji jest taki, że już w niej samej tkwi podstawowe kryterium stosowane w OSR: zgodności oddziaływania regulacyjnego postulowanego z oddziaływaniem faktycznym, a co za tym idzie również jakiejś kwalifikacji zakresu ewentualnie stwierdzonych rozbieżności z punktu widzenia osiągnięcia pożądanych celów prawodawczych. Już w takim ujęciu staje się jasne, że OSR jest rodzajem ewaluacji, rozumianej jako „zróżnicowany zestaw metod i działań zorientowany na krytyczną refleksję nad wartością i jakością interwencji publicznych – zarówno procesów ich wdrażania, jak i ich efektów”².

W takim ujęciu OSR może nawiązywać do konstrukcji teoretycznoprawnej doskonałego prawodawcy, tzn. takiego, który jest onnipotentny, w tym zna wszelkie konsekwencje swoich działań³ – wtedy ocena skutków regulacji jest testem kwalifikującym jakość kompetencji faktycznego legislatora w stosunku do owego konstruktu idealnego. Ponieważ prawodawca doskonały jest racjonalny, OSR w tym ujęciu sprowadza się do oceny racjonalności prawa. W takim ujęciu analiza naukowa stanowiąca trzon OSR napotyka na istotną i niedającą się pokonać barierę niepełnej kognicji wiedzy samego prawodawcy doskonałego. Stąd na gruncie teorii prawa formułuje się postulat racjonalności prawa najczęściej dużo skromniej – jako zgodność z jakimiś spełniającymi wymagania logiczne i aksjologiczne założeniami, przy czym rozróżnia się w tym zakresie także ocenę racjonalności „zewnątrznej” (gdy ocena odwołuje się do kompetencji w tym zakresie podmiotu działającego – a więc samego prawodawcy) oraz „wewnętrznej” (gdy to odwołanie odnosi się do kompetencji jakiegokolwiek innej niż podmiot działający osoby).

W ujęciu Michalskiej i Wronkowskiej:

„Podmioty tworzące prawo działają racjonalnie wówczas, jeżeli ich decyzje podejmowane są na gruncie najlepszej w danych warunkach wiedzy o okolicznościach działania oraz w sposób odpowiadający przyjmowanym społecznie preferencjom i przyjmowanej taktyce podejmowania decyzji. Inaczej mówiąc prawo jest tworzone racjonalnie, jeżeli ustanowienie jakichś norm prawnych jest, na gruncie najlepszej w danych okolicznościach wiedzy, środkiem dla realizacji wyznaczonych i akceptowalnych celów”.

W ujęciu mniej abstrakcyjnym, bo nawiązującym już do jakości normatywnie określonej praktyki postępowania, OSR ujmuje się jako systematyczną, obowiązkową, jednolitą i wewnętrznie spójną ocenę wpływu regulacji prawnych i polityk rządowych w zakresie ich wpływu na otoczenie społeczne, gospodarcze lub na środowisko naturalne⁴. W ujęciu Jacobsa, OSR stanowi narzędzie służące podejmowaniu decyzji, stanowiące metodę systematycznego i konsekwentnego badania wybranych potencjalnych skutków, wynikających z działań podejmowanych przez rząd, którego wyniki są przedmiotem informacji skierowanej do osób podejmujących decyzje prawotwórcze i polityczne⁵. W tym wymiarze rozszerza się niekiedy ową funkcję informacyjną OSR tak, że staje się ona podstawowym wymiarem realizacji zasady przejrzystości działalności legislacyjnej władzy publicznej⁶.

² K. Olejniczak, *Wprowadzenie do zagadnień ewaluacji*, [w:] K. Olejniczak, M. Kozak, B. Ledzion (red.), *Teoria i praktyka ewaluacji interwencji publicznych. Podręcznik akademicki*, Wyd. Akademickie i Profesjonalne, Warszawa, 2008 s. 19.

³ A. Michalska, S. Wronkowska, *Zasady tworzenia prawa*, Wyd. UAM, Poznań, 1980, s. 27–33.

⁴ C.M. Radaelli, *European Commission Conference on 'Impact Assessment in the European Union: Innovations, Quality and Good Regulatory Governance'* Background Report 20, 2003; C. Radaelli, B. Dente, S. Jacobs, C. Kirkpatrick, A. Meuwese, A. Renda, *Diadem Handbook: How to Perform the Diadem Data Collection*, 2006, www.enbr.org.

⁵ S.H. Jacobs, *Przegląd oceny skutków regulacji (OSR) w krajach OECD*, Organizacja Współpracy Gospodarczej i Rozwoju, Radom Warszawa, s. 14.

⁶ Tak np. w ujęciu zaprezentowanym [w:] European Commission (2009), *Impact Assessment Guidelines 15 January 2009*, dokument SEC(2009) 92, s. 4.

Wszystkie te definicje wręcz zakładają istnienie w ogóle fundamentalnej dla bytu regulacji prawnej relacji pomiędzy tą regulacją a wyborami politycznymi realizowanymi przez rząd (rozumiany ogólnie, a więc bez wskazywania na jego poziom czy charakter). Innymi słowy założenie jest takie, że prawo po prostu jest narzędziem polityki, a więc jest wykorzystywane do tego, by osiągać cele polityczne decydentów. W tym kontekście ocena skutków regulacji jest metodą weryfikacji, czy relacja między owym narzędziem polityki oraz jej celem jest taka, jak ją zakłada decydent polityczny i czy zgodny z tymi założeniami jest również mechanizm transmisji bodźców narzędziowych w owe cele. Ponieważ w praktyce trudno o w pełni poinformowanego decydenta politycznego, OSR może być także definiowana jako metoda wręcz dostarczająca informacji o tym, jaki może być prawdopodobny przebieg transmisji pomiędzy bodźcami, którymi posługuje się regulacja prawna a tymi elementami rzeczywistości, na które taka regulacja ma oddziaływać lub rzeczywiście oddziałuje.

W ujęciu *ex ante* OSR może więc służyć zapobieganiu przyjmowania takich rozwiązań prawnych, które nie spełniają określonych wymogów ekonomiki regulacyjnej (tzn. jakiegoś akceptowalnego poziomu ogólnego kosztu danej regulacji do korzyści z niej wynikających) lub które powodują niepożądane przez decydentów politycznych skutki – także takie, które niekoniecznie mają charakter czysto ekonomiczny, lecz choćby nie odzwierciedlają preferencji, których platformą realizacyjną jest to ugrupowanie polityczne, do którego należą w danym kontekście decydenci. W ujęciu *ex post*, OSR może zaś służyć podjęciu działań korygujących błędy legislacyjne i implementacyjne, przede wszystkim wprowadzeniu niezbędnych zmian regulacyjnych lub przyjęcia zmian w modelu wdrożenia prawa.

W rozumieniu tego opracowania pojęcie „polityka” oznacza – w najogólniejszym ujęciu – po pierwsze, określone pole oddziaływania rządu, a po drugie, wszystkie środki (metody, strategie i techniki?) wykorzystywane na tym polu po to, by osiągnąć jakieś cele (a więc i nadające się do oceny prakseologicznej, dotyczącej faktycznego „nadawania się” środków do uzyskania zamierzonego rezultatu). Należy jednak pamiętać, że procesy decyzyjne, wykorzystywane przez rząd w wyborze i pola oddziaływania, i środków oddziaływania wymaga „chronologicznej kontekstualizacji” procesów i decyzji w określonym czasie objętym perspektywą decyzyjną. Stąd w prawidłowo prowadzonej polityce liczy się nie jeden proces wiodący do jednej decyzji, ale raczej sieć powiązanych ze sobą decyzji wyznaczających określoną ścieżkę działania⁸. Dopiero taki obraz polityki wyznacza właściwe pole oceny dla OSR.

W takim kontekście, z punktu widzenia teorii prawa, relacja pomiędzy polityką i prawem może być interpretowana w różnych kategoriach, zasadniczo w zależności od tego, czy paradygmatem danego nurtu teoretycznego jest pełna autonomizacja prawa, czy nie. Interpretacja oceny skutków regulacji, jako zabiegu praktycznego, z takiego teoretycznego punktu widzenia jest formułowana w ramach teorii odrzucających pełną autonomizację prawa. Stąd właśnie do grupy takich teorii należy ta sformułowana przez Finnis, dla którego punktem wyjścia jest właśnie istnienie narzędziowej relacji pomiędzy polityką a systemem prawnym, w której prawo jest otwarte na swoiste przetworzenie decyzji politycznych na bardziej sztywne i systemowo spójne twierdzenia normatywne opatrzone sankcją wymuszającą ich przestrzeganie przez adresatów. Proces transformacji jednych twierdzeń w drugie Finnis określa *determinatio*⁹. Zgodnie z jego teorią *determinatio* przejawia się jako praktyczne uzasadnienie norm prawnych przyjmowanych w celu osiągnięcia jakichś legitymowanych celów społecznych. Za wyborem tych celów kryją się zasadniczo preferencje polityczne dotyczące wartości. Stąd istotą *determinatio* jest „zakorzenianie”

⁷ D. Easton, *The Political System: An Inquiry into the State of Political Science*, University of Chicago Press, Chicago, 1981, s. 130.

⁸ Zob. np. M.J. Hill, *The Policy Process in the Modern State*, Prentice Hall, Englewood Cliffs, 1997, s. 16–17; M. Zamboni, *The Policy of Law: A Legal Theoretical Framework*, Hart Publishing, Oxford, 2007, s. 102–103.

⁹ J. Finnis, *Natural Law and Natural Rights*, Clarendon Press, Oxford, 1980, szczególnie s. 276–282.

w systemie prawnym tych wartości, które są przedmiotem wyboru politycznego. Nawiązując do teorii ewaluacji możemy przyjąć, że *determinatio* stanowi szczególną formułę określania *value* (wartość), czyli normatywnego (aksjologicznego) odniesienia ewaluacyjnych kryteriów oceny (np. w przypadku przedsięwzięć finansowanych z funduszy strukturalnych rolę *value* pełnią kierunki strategiczne NSRO, a w odniesieniu do polityki rodzinnej *value* stanowią w pewnym zakresie wartości wynikające ze światopoglądu dominującego w wymiarze politycznym w wyniku wyborów).

Determinatio funkcjonuje jako społeczny mechanizm „rozstrzygnięcia autorytatywnego”, w którym chodzi o wybór norm prawnych, które „najlepiej” (z punktu widzenia prakseologicznego) mogą służyć do osiągnięcia dobra wspólnego¹⁰. W takim kontekście OSR może być interpretowana jako zbiór systemowych, a więc zinstytucjonalizowanych i mających swoje własne normy prawidłowości, zabiegów służących realizacji Finnisowskiego *determinatio*. Innymi słowy, to co Finnis ujmuje w kategoriach abstrakcyjnych, z punktu widzenia systemu, w którym stosuje się OSR, jest faktycznie realizowane właśnie jako ocena skutków regulacji.

Dla Zamboniego mniej więcej taka właśnie relacja polityki i prawa przybiera postać modelu, w którym rezultaty dyskursu politycznego są transformowane (podlegają transferowi) w politykę prawa stanowiącą zbiór procesów i rezultatów służących już transpozycji tych twierdzeń w elementy systemu prawnego – kategorie prawne i koncepcje prawne.



Rys. 1. Miejsce polityki prawa pomiędzy polityką a prawem

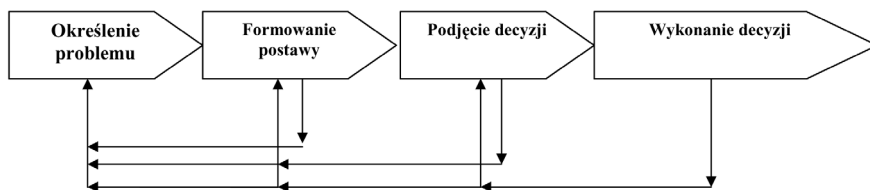
Źródło: Zamboni M., *The Policy of Law: A Legal Theoretical Framework*, Hart Publishing, Oxford, 2007, s. 129.

W ujęciu przedstawionym na rys. 1 polityka prawa jest szczególnym rodzajem przekształcania (a więc czegoś, co dzieje się w danym czasie, w postaci procesów oraz czegoś, co jest już zamkniętym wynikiem tych procesów – a więc efektów) wyborów politycznych w narzędzie ich realizacji, jakim jest prawo. W tym układzie na etapie wyboru politycznego, idee i wartości są formowane w miarę elastycznych ramach, co także dotyczy ram samego procesu. Z drugiej strony nie muszą one być – przynajmniej na wczesnym etapie procesu politycznego – formułowane narzędziowo. Raczej są kategoriami bardzo masywnymi aksjologicznie, ale przy tym niekoniecznie wystarczająco dokładnie określonymi. Po ich przekształceniu – w ramach polityki prawa – w samo prawo, stają się one już jakimiś realizacyjnymi konstrukcjami prawnymi o ustalonym kształcie normatywnym (co przesądza zarówno o ich relatywnej sztywności, jak i zarazem pewności). Owe konstrukcje prawne nie są przypadkowym przekształceniem idei i wartości wybranych w procesie politycznym – raczej służą one narzędziowo ich realizacji. W takim kontekście ocena skutków regulacji dotyczy tego, czy zamierzony związek przekształcenia rzeczywiście istnieje (w tym, czy wręcz system prawny nadaje się do realizacji wybranych w procesie politycznym idei i wartości) oraz czy realizujące funkcje prawa są sprawne z punktu widzenia tego związku. Innymi słowy, w szerszych kontekstach, OSR jest formą kontroli jakości polityki prawa w danej dziedzinie bądź wręcz całej polityki prawa w danym państwie.

¹⁰ J. Finnis, *The Authority of Law in the Predicament of Contemporary Social Theory*, „Notre Dame Journal of Law”, Ethics and Public Policy”, No. 1, 1984, s. 133.

Ocena skutków regulacji w procesie legislacyjnym

Ogólnym zadaniem oceny skutków regulacji jest wsparcie politycznie programowanych procesów podejmowania decyzji. W takim ujęciu zakłada się systemową iteracyjność tych procesów, która pozwala na odpowiednią korektę podjętych już w toku trwania procesu decyzyjnego działań oraz na powrót do tego ich etapu, który wymaga – jako najwcześniejszy – powtórzenia. Na przykładzie prostego modelu podejmowania decyzji wpływ OSR przedstawiałby się więc następująco:



Rys. 2. Ogólny model iteracyjnego procesu decyzyjnego

Źródło: Adaptacja w oparciu o: Zaltman G., Duncan R., *Strategies for Planned Change*, John Wiley&Sons, New York, 1997, s. 33–58.

W zaprezentowanym modelu informacja przedstawiona w wyniku oceny skutków regulacji jest wykorzystana do przejścia z tego etapu procesu podejmowania decyzji, na którym ten proces jest do etapu, w którym ta informacja wywołuje konieczność korekty. W odniesieniu do trzech pierwszych etapów, stanowiących razem fazę formowania decyzji, ocena skutków regulacji ma charakter *ex ante*. Na etapie wykonania decyzji (dotyczącej przyjęcia danego aktu prawnego) korekta powodująca skutki iteracyjne (powrotu do któregośkolwiek etapu fazy formowania decyzji) wynikać musi już z OSR *ex post* (co wymaga monitorowania funkcjonowania określonej regulacji prawnej).

Bardziej precyzyjny i dodatkowo osadzony w konkretnym kontekście instytucjonalno-prawnym model procesu decyzyjnego dotyczącego już określonej legislacji inicjowanej przez rząd może być zrekonstruowany w oparciu o polską regulację i praktykę zarządzania procesem prawotwórczym (jednak tylko w wymiarze zaczynającym się i kończącym na rządowym centrum zarządzania państwem, choć w literaturze podkreśla się jego przydatność także w innych kontekstach stanowienia prawa¹¹). W tym przypadku procedura niekoniecznie jest w pełni iteracyjna (choć hipotetycznie może za taką być uważana). Dotyczy ona jedynie ustaw i innych projektów aktów normatywnych, których dysponentem są organy podlegające Prezesowi Rady Ministrów. Składa się ona z dwu zasadniczych etapów:

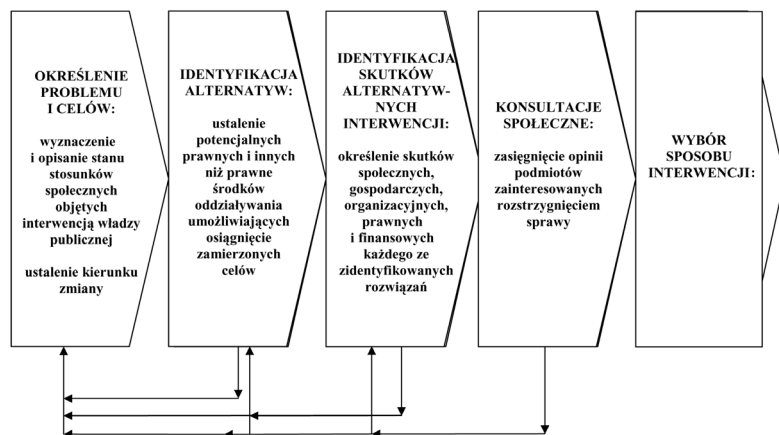
- przygotowania do podjęcia decyzji o sporządzeniu projektu aktu prawotwórczego;
- przygotowania zasadniczego rozwiązania prawnego wtedy, gdy na etapie przygotowania do podjęcia decyzji okazało się, że uchwalenie ustawy jest pożądane.

Pomiędzy oboma wskazanymi etapami może zachodzić relacja iteracji (co zakłada możliwość cofnięcia procesu, jeśli zachodzą odpowiednie do tego przesłanki), jednak w praktyce nie odnotowuje się jakiegóż znaczącej praktyki tego rodzaju. Także w innych zakresach (tj. w ramach szczegółowych działań podejmowanych na każdym etapie procesu legislacyjnego) iteracji także się nie dostrzega.

Etap przygotowania do podjęcia decyzji o sporządzeniu projektu aktu prawotwórczego można przedstawić graficznie w postaci rysunku 3. W układzie zaprezentowanym na tym rysunku ewaluacja (*ex ante*) skutków interwencji regulacyjnej właściwa OSR jest trzecią fazą procesu przygotowania do pod-

¹¹ S. Wronkowska, M. Zieliński, *Komentarz do zasad techniki prawodawczej*, Wyd. Sejmowe, Warszawa, 2004, s. 20–21.

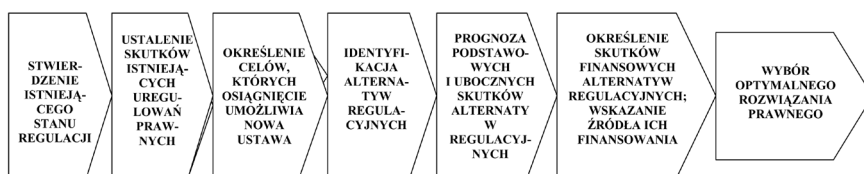
jęcia decyzji o przyjęciu danego aktu prawnego lub podjęcia innych działań, do których rząd ma kompetencje. W tym sensie jest ona także metodą weryfikacji prawidłowości przyjętych na potrzeby przyszłej interwencji publicznej założeń, która może także wieść do wniosku negatywnego z punktu widzenia rozważanej interwencji legislacyjnej – może bowiem taką interwencję odrzucać jako niewytrzymującą konkurencji z innymi sposobami rozwiązania danego problemu społecznego bądź wręcz od samego początku niepotrzebną.



Rys. 3. Etap przygotowania do podjęcia decyzji w polskiej regulacji dotyczącej zasad techniki prawodawczej

Źródło: Opracowanie własne.

W przypadku konieczności przejścia do następnego etapu procesu legislacyjnego, a więc wtedy, gdy podjęto jednak decyzję o przygotowaniu projektu ustawy, polska regulacja dotycząca zasad techniki prawodawczej wymaga przejścia następujących siedmiu faz analizy (przy czym nie jest oczywiste, że pomiędzy wszystkimi tymi fazami istnieje relacja sekwencyjna), przedstawionych graficznie na rysunku 4:



Rys. 4. Etap przygotowania projektu ustawy w polskiej regulacji dotyczącej zasad techniki prawodawczej

Źródło: Opracowanie własne.

Analizując poszczególne fazy przygotowania projektu ustawy już po podjęciu decyzji o interwencji regulacyjnej, warto zwrócić uwagę, że pomiędzy dwoma etapami procesu wiodącego do podjęcia konkretnej akcji legislacyjnej i przyjęcia określonej ustawy istnieje silna relacja analogii. Na drugim bowiem etapie (przedstawionym na rysunku 4), tak jak na etapie pierwszym (rysunek 3) także mamy do czynienia z diagnozą stanu wyjściowego, określeniem celu zmiany, identyfikacją alternatywnych rozwiązań, szacunkiem prognozowanych skutków poszczególnych alternatyw oraz wyborem określonego sposobu działania. Różnica polega jedynie na tym, że drugi etap dotyczy już tylko analizy systemu prawnego oraz

jego ewentualnych modyfikacji, podczas gdy pierwszy musi zakładać także pozaregulacyjne metody osiągnięcia wskazywanych celów politycznych. Z uwagi na stwierdzoną analogię, metodologia właściwa OSR powinna pojawiać się w praktyce polskiej przynajmniej na dwu etapach każdego procesu potencjalnie wiodącego do przyjęcia ustaw. Niepełne stosowanie ustalonej metodologii prowadzi do tego, że nie zawsze tak jest¹². Problemem jest najprawdopodobniej złożoność samego procesu określonego w zasadach techniki prawodawczej (co powoduje również znaczne koszty transakcyjne reiteracji) oraz poważne trudności metodologiczne, które z jego zastosowaniem muszą się wiązać – polegające choćby na identyfikacji i różnicowaniu skutków regulacji podstawowych i ubocznych. Niepełna realizacja wytycznych wynikających z zasad techniki prawodawczej jest także zdeterminowana przez kulturę polityczną i administracyjną – może wynikać z preferowania doraźnych (i zasadniczo formalnie zdefiniowanych) rozwiązań prawodawczych, w odniesieniu do których liczy się na szybki efekt pozytywnej percepcji otoczenia, jak również ze złego planowania działalności legislacyjnej, w której występuje nadmierna liczba projektów pilnych (a więc takich, w których przeprowadzenie pełnej OSR nie jest możliwe). Mówimy więc tu o objawach wadliwości polityki prawa.

Tak czy inaczej polski model analizy poprzedzającej przyjęcie aktu prawnego odzwierciedla *mainstream* praktyk krajów OECD, przynajmniej w tych przypadkach, gdy chodzi o przyjęte w nich takie zasady oceny skutków regulacji, które dotyczą w miarę szerokiego zakresu analizy (w niektórych państwach, choćby w Wielkiej Brytanii, szczególny nacisk kładzie się np. na kwestię wpływu regulacji na działalność gospodarczą, zaś efekty w pozostałych obszarach życia społecznego traktuje się mniej priorytetowo). Gdy porówna się go do standardów tam wypracowanych lub zgłaszanych przez doktrynę OSR można jednak zauważyć, że występują tam także dodatkowe elementy analizy. W szczególności słusznie przewiduje się wskazanie w OSR na szacunek okresu, w jakim dana regulacja miałaby funkcjonować oraz opis zastosowanej w ewaluacji metodologii. Bywa i tak, że wymaga się nie jedynie określenia podmiotów, których regulacja dotyczy, lecz ujmuje się ten wymóg kwalifikująco: należy określić tylko te grupy podmiotów, na które proponowana regulacja wpływa szczególnie silnie (przy czym często zakłada się, że do tej grupy mogą – przynajmniej w odniesieniu do regulacji gospodarczej i socjalnej – należeć małe i średnie przedsiębiorstwa).

Funkcje oceny skutków regulacji

W odniesieniu do konkretnych zadań oceny skutków regulacji zwykle dostrzega się następujące jej zadania z punktu widzenia rządów¹³:

- (a) po pierwsze, OSR ma pozwolić na lepsze zrozumienie wpływu działań rządu na rzeczywistość społeczno-gospodarczą, w tym zarówno korzyści, jak i koszty tych działań;
- (b) po drugie, OSR ma umożliwić integrację złożonych celów polityki (w której to funkcji OSR spełnia zarówno rolę narzędzia analitycznego, jak i koordynującego, umożliwiającego – już w ramach procesu politycznego – pogodzenie (harmonizację) zróżnicowanych, zidentyfikowanych interesów);
- (c) po trzecie, OSR ma służyć zapewnieniu przejrzystości procesu legislacyjnego (czy nawet szerzej – procesu kształtowania polityki w danej dziedzinie oddziaływania rządu) oraz doskonaleniu konsultacji proponowanych rozwiązań z partnerami społecznymi;

¹² Zob. np. A. Palukiewicz, M. Gancarz, *Nowa metodologia oceny skutków regulacji na tle poprzednich doświadczeń*, [w:] W. Szpringer, W. Rogowski (red.), *Ocena skutków regulacji – poradnik OSR, doświadczenia, perspektywy*, C. H. Beck, Warszawa, 2007, s. 142–145; J. Wierzbicki, *Ocena skutków regulacji w projektach aktów prawnych dotyczących finansów publicznych – obserwacje i wnioski de lege ferenda*, [w:] *Ibidem*, s. 158–161.

¹³ Zob. S.H. Jacobs, *op. cit.*, s. 17–18.

(d) po czwarte, OSR ma także powodować zwiększenie odpowiedzialności za prowadzoną przez niego politykę poprzez formułowanie kryteriów oceny i standardów odnoszących się do zapewniania rozsądnej równowagi pomiędzy korzyściami i kosztami decyzji legislacyjnych.

Wymienione zadania pokrywają się z najczęściej wskazywanymi funkcjami ewaluacji, a więc funkcją poznawczą, formatywną (kształtowanie polityk publicznych) oraz rozliczeniową (wzmacnianie społecznej odpowiedzialności instytucji publicznych). Spełnienie tych oczekiwań zakłada rozpoznanie w ramach technologii analitycznych zastosowanych w OSR związków przyczynowo-skutkowych pomiędzy bodźcem, którym jest zamierzona bądź już przyjęta (w tym drugim przypadku w ocenie *ex post*) regulacja a jej rozmaitymi polami oddziaływania. Już w zakresie identyfikacji skutków regulacyjnych pojawia się jednak istotny problem określenia kryteriów, według których owe skutki powinny być identyfikowane, który można nazwać determinacją mapy analitycznej OSR. Jest to problem wieloaspektowy. Najważniejsze aspekty tego problemu są trzy:

- (a) po pierwsze, dotyczą wyznaczenia obszarów, które podlegają analizie co do skutków regulacji (co także determinuje wtórnie wybór kryteriów oceny tych skutków);
- (b) po drugie, odnoszą się do głębokości analizy związków przyczynowo-skutkowych pomiędzy regulacją prawną a jej skutkami w wyznaczonych obszarach (co wymaga nie tylko identyfikacji skutków, ale również ich podzielenia ze względu na znaczenie – przy czym wymaga to oczywiście przyjęcia kryteriów określenia ważności skutków);
- (c) po trzecie, dotyczą wyboru rozległości powiązań pomiędzy poszczególnymi obszarami oddziaływania regulacji prawnej w ogóle, które będą poddane analizie w ocenie skutków konkretnej regulacji.

Podstawowym ograniczeniem wyboru bardzo rozległej (i głębokiej) mapy analitycznej OSR jest rachunek korzyści i kosztów OSR jako takiego, obejmujący w szczególności koszty transakcyjne. Oczywiście w tym przypadku jest to, że im bardziej rozległa i głęboka jest analiza, tym jej koszty są większe, zaś krańcowe przyrosty użyteczności mniejsze. Wynika to m.in. stąd, że relacja pomiędzy bodźcem regulacyjnym (przyczyna) a jej (pożądanym bądź niepożądanym) skutkiem w postaci jakiejś modyfikacji rzeczywistości przebiega wedle zasady, że relacja jest istotna w odniesieniu do stosunkowo niewielkiej grupy bodźców i skutków. Im większa jest liczba rozważanych relacji, tym mniejsza istotność kolejno analizowanych związków. Sprowadza się to do uproszczonej zasady, że w ocenie skutków regulacji należy zidentyfikować te 20% związków przyczynowo-skutkowych, które są w 80% istotne ze społecznie istotnych (i z góry znanych) punktów widzenia.

Odniesienie się do wskazanych wyzwań natury zasadniczo teoretycznej wpływa na treść zadań stawianych OSR. W tym zakresie można jedynie hipotetycznie zakładać, że wybory dotyczące wszystkich trzech kluczowych aspektów tej analizy będą przedmiotem działań eksperckich. W krajach stosujących rutynowo w miarę rozwinięte mechanizmy OSR w praktyce jakiegokolwiek wybory dotyczące rozległości mapy analitycznej OSR są przedmiotem jakiejś formy jej uzgodnienia na arenie politycznej, wszakże przy wzięciu pod uwagę argumentacji eksperckiej (przede wszystkim przedstawianej przez urzędników). Jest to więc rodzaj dialogu pomiędzy zamawiającymi OSR (jako taki), czyli gremiami politycznymi, a tymi, którzy to zamówienie na informację będą realizować, czyli osobami, które OSR będą wykonywać. Pewna część tak pojętych uzgodnień zostaje również „uszywniona” w postaci przejęcia ich przez regulację prawną (różnej mocy). Przykładem takiego rozwiązania jest np. przyjęta w polskim prawie zasada, że ocena skutków regulacji powinna określać:

- (a) podmioty, na które oddziałuje projektowane prawo,
- (b) wpływ funkcjonowania regulacji na finanse publiczne,
- (c) wpływ regulacji na rynek pracy,

- (d) wpływ na konkurencyjność gospodarki,
- (e) wpływ na sytuację i rozwój regionów.

Dodatkowo, zgodnie z uregulowaniem polskim, ocena skutków regulacji powinna zawierać informację o przeprowadzonych w trakcie przygotowywania projektu danego aktu konsultacjach społecznych i ich wynikach.

Podstawową cechą tak ukształtowanej regulacji prawnej dotyczącej OSR jest jej naturalne (bo opierające się na regułach prawa) usztywnienie. Jednocześnie regulacja ta wyznacza zakres analizy, będącej narzędziem oceny skutków regulacji – co oczywiście jest wyborem pragmatycznym, lecz i zawężającym mapę analityczną OSR w zbyt dużym zakresie wtedy, gdy projektowany akt prawny oddziałuje jednak na inne obszary. Dodatkowo, przyjęta regulacja prawna dotycząca zakresu oceny skutków regulacji nie obejmuje kwestii analizy wzajemnego oddziaływania projektowanej regulacji i regulacji już istniejących, przez co ogranicza możliwość realizacji tej funkcji OSR, która polega na integracji złożonych celów polityki (która to funkcja może być częściowo realizowana w przypadku prawidłowo prowadzonych i autentycznych konsultacji społecznych). Jeszcze jednym ograniczeniem systemowym OSR w Polsce jest odstąpienie od wymagania oceny skutków regulacji w przypadku poselskich projektów aktów prawnych. Wszystko to powoduje oczywiście – legislacyjnie zdeterminowane – ograniczenie funkcji informacyjnej OSR.

Ograniczenie mapy analitycznej (a więc i realizowanych funkcji) zdradzają także inne modele OSR stosowane w świecie. I tak w modelu amerykańskim podkreśla się funkcję minimalizacji sprzeczności (i zwiększenia spójności) przyjmowanych regulacji prawnych, przy jednoczesnym ograniczeniu obowiązku przejścia ewaluacji jedynie dla tych projektów aktów prawnych, które spełniają wymóg „istotności wpływu” (chodzi o tzw. *significant regulatory actions*). Za takie uznano w regulacji EO (*Evaluation Orientation*) 12866 projektów, których wpływ finansowy w którymkolwiek roku wynosi co najmniej 100 milionów dolarów lub które negatywnie wpływają na gospodarkę jako całość, sektor gospodarki, produktywność, konkurencję, ilość miejsc pracy, zdrowie lub bezpieczeństwo publiczne, na funkcjonowanie państwa, jak również funkcjonowanie administracji i społeczności lokalnych i plemiennych¹⁴. Dodatkowo należy wspomnieć, że przynajmniej od 1980 r. w amerykańskim systemie OSR widać wyraźną tendencję do analizy regulacyjnej nakierowanej na eliminację regulacji zbędnych (np. nakładających zbyteczne obowiązki sprawozdawcze na przedsiębiorców) bądź takich, które są niedostosowane do specyfiki małych i średnich przedsiębiorstw. System oceny regulacji jest znacząco otwarty także na działania lobbingsowe.

Podstawowe funkcje OSR w modelu brytyjskim obejmują zarówno działania w zakresie kontroli prawidłowości regulacji już ustanowionej (mają więc wymiar *ex post*), jak i planowanej. Stąd podstawowe funkcje OSR polegają na umożliwieniu racjonalnego uchwalania prawa oraz zmiany istniejącego prawa w celu zapewnienia jego większej racjonalności. W tym kontekście, co najmniej od pierwszych lat XXI w. zwraca się w szczególności uwagę na działania legislacyjne, które:

- (a) skutkują eliminacją niespójności i anomaliiw legislacyjnych,
- (b) skutkują eliminacją ciężarów wobec adresatów prawa, w przypadku gdy jedynie same organy państwa odnoszą z tych ciężarów jakieś korzyści (co oznacza brak szerszego interesu społecznego w utrzymaniu tych ciężarów),
- (c) jeżeli nakładają dodatkowe ciężary na adresatów prawa, to spełniają kryterium proporcjonalności (a więc konieczności regulacji, jej niezbędności i adekwatności do realizowanego, legitymowanego interesu¹⁵),

¹⁴ Za A. Renda, *Impact Assessment in the EU: The State of the Art. And the Art of the State*, CEPS, Brussels, 2006, s. 13–14.

¹⁵ Co do dokładnej treści tej ważnej dla OSR zasady zob. np. E. Ellis, (ed.), *The Principle of Proportionality in the Laws of Europe*, Hart, Oxford, 1999.

(d) dookreślają jakieś obowiązki w zakresie, w jakim nie ma co do nich wystarczających podstaw prawnych lub zakres obowiązywania prawa jest niejasny, jak również wprowadzają mechanizm pozwalający na bardziej elastyczną reakcję legislacyjną w przyszłości poprzez ustanowienie zasad późniejszego dostosowania przyjętej regulacji drogą późniejszych zmian prawa.

Poważniejsza zmiana brytyjskiego systemu OSR nastąpiła w 2005 r. i była częścią realizacji Planu Działania na rzecz Lepszej Regulacji (*Better Regulation Action Plan*). Polegała ona na wprowadzeniu do metodologii ewaluacyjnej metod oceny ryzyka. Istotą nowego podejścia w tym zakresie jest konieczność identyfikacji ryzyka naruszenia zasady proporcjonalności w stanowionym prawie, jak również eliminacji tych regulacji, które wiążą się z nadmiernymi w tym kontekście ciężarami wobec ich adresatów. Politycznym hasłem wprowadzenia tej poważnej zmiany metodologicznej był slogan „Mniej to więcej” oraz bardziej konkretny: „nie ma inspekcji bez uzasadnienia, nie ma formularzy bez uzasadnienia i nie ma zbierania informacji bez uzasadnienia” (*„No inspection without justification, no form filling without justification and no information requirements without justification”*)¹⁶. Stąd podstawowymi, priorytetowymi funkcjami OSR stało się umożliwienie osiągnięcia w działalności legislacyjnej takiego standardu, by:

- (a) wprowadzane akty prawne ustanawiały środki proporcjonalne także do ryzyka zagrożenia określonego legitymowanego interesu, w tym ograniczyły uprawnienia inspekcyjne organów państwa i ich kompetencje w zakresie wymuszania przestrzegania prawa do niezbędnego minimum;
- (b) wprowadzane zmiany stale zmniejszały koszty administrowania daną regulacją – i by było to nakierowane na osiągnięcie uzgodnionych w ramach administracji państwowej celów redukcji ciężarów administracyjnych.

Warto zwrócić uwagę, że przyjęte rozwiązania brytyjskie są ilustracją tego, że w swoim podstawowym zakresie kryteria oceny stosowane w OSR w znacznym stopniu podlegają fluktuacji politycznej – odzwierciedlają one przecież priorytety rządu, w tym przypadku realizującego stosunkowo liberalną wizję ograniczenia ingerencji państwa. Nie jest to ilustracja nietypowa – preferencje polityczne muszą mieć odzwierciedlenie w ocenie skutków regulacji; jednocześnie należy pamiętać, że podstawowym celem OSR powinno w każdym momencie być oszacowanie stosunku korzyści danej interwencji legislacyjnej do jej kosztów – przy uwzględnieniu w ich definicji wspomnianego już czynnika politycznego. Samą analizę politycznej akceptowalności poszczególnych opcji legislacyjnych, które są wynikiem OSR niektóre ze stosowanych metodologii zalecają ująć jednak nie w samym sprawozdaniu ewaluacyjnym, lecz w osobnym dokumencie¹⁷.

Podstawowe problemy metodologii oceny skutków regulacji

Ocena skutków regulacji wiąże się z poważnymi problemami metodologicznymi. Wynika to z wielu powodów, z których wiele tkwi w samym postulatcie racjonalności systemu prawnego, stanowiącego przecież jedno z dóbr, które przez OSR musi być niejako chronione. Wymóg racjonalności systemu prawnego jest spełniony wtedy, gdy wykazuje on odpowiednie wewnętrzne zorganizowanie – przy czym przybiera ono postać powiązań pomiędzy normami prawnymi na ten system się składającymi. Powiązania te, o charakterze kompetencyjnym (dynamicznym) i treściowym (statycznym), muszą realizować pewne postulaty racjonalności tworzenia i odtwarzania systemu (w tym drugim przypadku następuje

¹⁶ Za A. Renda, *op. cit.*, s. 34.

¹⁷ Tak np. w praktyce Komisji Europejskiej. Zob. European Commission, *Impact Assessment Guidelines* 15 January 2009, dokument SEC(2009) 92, s. 12.

to w procesie stosowania poszczególnych norm prawnych – wtedy gdy poprzez skonfrontowanie z zakładaną sytuacją faktyczną aktualizuje się obowiązek stosowania normy prawnej)¹⁸. Do podstawowych postulatów w tym zakresie należy postulat wzajemnego logicznego i prakseologicznego ich powiązania (co prowadzi do spełniania wymogów sprawności ekonomicznej systemu¹⁹, jego spójności²⁰ oraz formułowanej niekiedy jako postulat dobrej legislacji – „poręczności”²¹). To z kolei zakłada, że prawodawca je ustanawiający dobrze zidentyfikował i zweryfikował w oparciu o uznane kryteria aksjologiczne cele społeczne, które mają być osiągnięte za ich pomocą. A to wymaga, by prawodawca miał najlepszą w danych warunkach wiedzę o możliwych powiązaniach przyczynowo-skutkowych poszczególnych norm systemu. Ten wymóg jednak nie jest spełniony nawet w odniesieniu do w pełni i prawidłowo przeprowadzonej w ramach OSR ewaluacji, choćby dlatego, że oceny ewaluatorów są zniekształcane w warunkach ograniczonego (bo kosztownego) dostępu do informacji – a więc spełnienia przesłanek ograniczonej (a nie pełnej) racjonalności²². Jest tak tym bardziej, że pewna część wnioskowania, z którego korzysta metodologia OSR (w jakimkolwiek konkretnym ujęciu) opiera się wręcz na metodach intuicyjnych²³. Ewaluatorzy nie są także w stanie w pełni dokonać rekonstrukcji tych przyszłych (a zapewne w poważnej części koniecznych) zmian systemu prawnego, które dopiero będą dokonane.

Pozostałe trudności mają już charakter bardziej techniczny. Pierwszorzędny problem tego typu polega na tym, że identyfikacja wszystkich potencjalnych korzyści i kosztów danej regulacji oraz ich uszeregowanie pod względem ważności lub wartości jest zadaniem bardzo złożonym. Dodatkowo, bardzo trudno jest ustalić taką metodologię ewaluacji w tym zakresie, by różnorodne oceny poczynione w tym zakresie sprowadzić do jakiegoś wspólnego mianownika. Należy przecież pamiętać, że – jak to wskazano w rozdziale 1 tego opracowania – prawo realizuje idee i wartości polityczne, z których część nie daje się łatwo albo wręcz wcale kwantyfikować i porównywać. Mówimy bowiem często o ideach i wartościach transcendentnych (z których niektóre nie poddają się łatwo wypracowanym już metodom wyceny, choć większość istniejących metod OSR zasadniczo nakazuje ich „monetaryzację”, tj. ujęcie pieniężne)²⁴.

Drugą istotną trudnością metodologiczną wiążącą się z rachunkiem korzyści i kosztów interwencji legislacyjnej (także konfrontowanej w tym przypadku z kosztami i korzyściami brakiem takiej interwencji) jest przyjęcie odpowiedniej perspektywy czasowej dla takiej oceny, jak również wyznaczenie jej perspektywy. Prawidłowość doboru w tym zakresie może bowiem decydować o ostatecznym wyniku ewaluacji. Dodatkowo, w przypadku gdy w już prawidłowo dobranej perspektywie czasowej dla różnych alternatywnych rozwiązań występuje różna sekwencja różnowartościowych korzyści i kosztów, powstaje konieczność ich szacunku wedle jakiegoś zobiektywizowanego kryterium. W doniesieniu do wielkości, które da się skwantyfikować w pieniądzu, należy zastosować metodologię dyskontowania tak, by uzyskać wielkości przyszłe sprowadzone do ich wartości obecnej netto (NPV, *net present value*). Stosuje się wtedy następujący rachunek:

¹⁸ A. Nowak-Far, *Optimalizacja tworzenia i stosowania prawa*, [w:] Ministerstwo Gospodarki (red.), *Reforma Regulacji w praktyce – doświadczenia nowych państw członkowskich Unii Europejskiej*, materiały konferencyjne 9 października 2006, 2007, s. 9–10.

¹⁹ Wyrażanej w kategoriach efektywności i wydajności.

²⁰ Rozumianej jako prakseologiczna zgodność z innymi normami systemu lub jako zgodność z celami danej polityki lub grupy polityki publicznych).

²¹ Rozumianej jako łatwość korzystania przez wszystkich użytkowników danej regulacji. Zob. A. Nowak-Far, *Zarządzanie strategiczne a analiza efektywności reformy administracji publicznej*, „Służba Cywilna” 2001, nr 2, s. 134.

²² O przyczynach i właściwościach błędów w tych warunkach zob. [w:] D. Kahneman, A. Tversky, *Choices, Values and Frames*, Cambridge University Press, Cambridge, 2000, s. 17–66.

²³ W tym zakresie rozległe spectrum analizy wpływu na wnioskowanie przedstawiona została [w:] T. Gilovich, D. Griffin, D. Kahneman, *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge University Press, Cambridge, 2002.

²⁴ Przegląd podstawowych metod w tym zakresie zob. [w:] A. Surdej, *Metoda analizy kosztów i korzyści: imperatyw efektywności w politykach publicznych*, [w:] A. Haber, M. Szałaj, *Środowisko i warsztat ewaluacji*, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa, 2008, s. 112–118.

$$V = \sum_{t=1}^n \frac{R^t}{(1+i)^t}$$

gdzie

V – obecna wartość w pieniądzu przyszłych korzyści/kosztów,

R^t – wielkość przyszłych korzyści/kosztów w tym okresie, w którym zostaną one zrealizowane,

t – dany rok w przyszłości,

n – okres przynoszenia korzyści/ponoszenia kosztu,

i – stopa dyskonta.

Przy stosowaniu metodologii NPV trzeba jednak pamiętać, że rodzi ona poważny problem wyboru właściwej stopy dyskonta, jak również jej stosowanie nie może być wszechstronne – niekiedy wręcz nie można jej w OSR (ani w żadnej ewaluacji dotyczącej jakości planowanej legislacji) stosować. Jest tak wtedy, gdy przyszłe, wystarczająco poważne koszty mają tak znaczną perspektywę czasową wystąpienia, że będą już ponoszone przez przyszłe pokolenie. Ich zdyskontowanie w takim przypadku prowadziłoby do decyzji naruszających zasadę solidarności międzypokoleniowej, która musi być przez każde państwo realizowana.

Warto wspomnieć, że metoda NPV może być łączona z rozwiniętą metodą modelu kosztu standardowego (*Standard Cost Model, SCM*), ale jedynie w odniesieniu do kosztów powtarzających się i przewidywanych w przyszłości – z wyjątkiem więc kosztów jednorazowych ponoszonych przy wprowadzeniu danego środka legislacyjnego w życie. Jak wiadomo, SCM (w wąskim ujęciu) polega na określeniu kosztów administracyjnych, wynikających dla podmiotów nieadministracyjnych z nałożonych na nie obowiązków przekazywania informacji, dla których ostatecznie całkowity koszt społeczny będzie wynosił:

$$TSC = \Sigma C \times (Q \times F)$$

gdzie

TSC – całkowity koszt społeczny,

C – koszt standardowy działania podmiotu zobowiązanego,

Q – liczba podmiotów zobowiązanych,

F – częstotliwość działań będących istotą obowiązku.

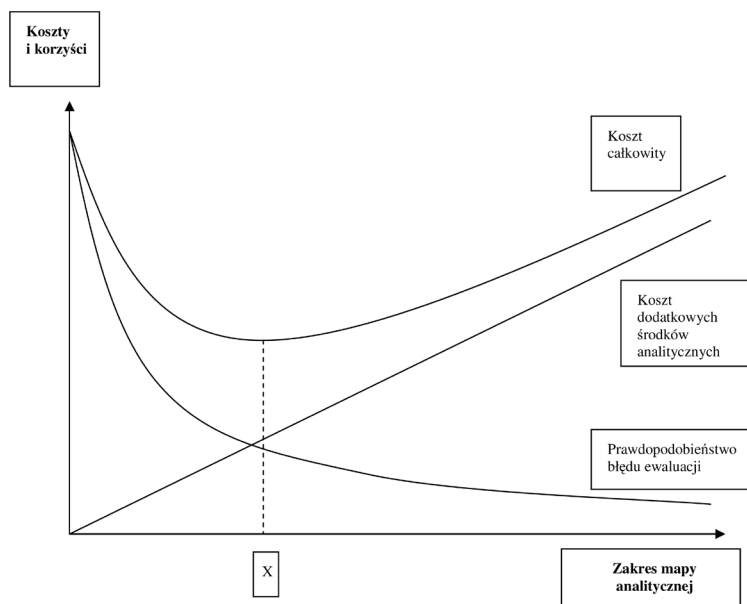
Tak więc w ujęciu, w którym uwzględnia się szeregi czasowe, NPV kosztu standardowego będzie wynosić

$$V = \sum_{t=1}^n \frac{R^t}{(1+i)^t}$$

Bardzo poważnym problemem jest także sam koszt procedur, które mają być zastosowane w OSR (i każdej innej ewaluacji). Im bardziej rozległa i wnikliwa analiza w OSR wykorzystywana, tym mniejsze prawdopodobieństwo wystąpienia tu jakiegoś poważniejszego błędu. Jednocześnie jednak w miarę zwiększania mapy analitycznej koszt krańcowy uzyskiwania dodatkowych korzyści (użyteczności) wzrasta. Relację tę można przedstawić posługując się – drogą analogii – w postaci krzywych Calabresiego (rys. 5).

Na rys. 5 optymalny wybór w zakresie rozmiarów mapy analitycznej stosowanej w OSR dokonany będzie przy wielkościach kosztów i korzyści (tu mierzonych coraz mniejszym – w miarę jej poszerzania

– prawdopodobieństwem wystąpienia błędu ewaluacji) odpowiadających punktowi X. W tym bowiem miejscu wielkość całkowitych kosztów jest najmniejsza. Oznacza to jednak, że wybierając punkt optymalny, należy równocześnie zrezygnować z wielu informacji, które byłyby dostępne, gdyby poniesiono większe koszty OSR. Należy również zwrócić uwagę, że prawidłowość opisana za pomocą krzywych Calabresiego nie wykaże anomalii jedynie wtedy, gdy działania analityczne w ramach OSR zostaną prawidłowo uszeregowane (z punktu widzenia istotności badanych zjawisk, a nie tylko zachowania jakiejś sekwencji).



Rys. 5. Koszt działań analitycznych OSR w ujęciu przyrostowym

Źródło: Adaptacja modelu G. Calabresiego opisanego [w:] Calabresi G., *The Cost of Accidents*, Yale University Press, New Haven, 1970, s. 237–288.

Zachowanie właściwej proporcji korzyści i kosztów samego OSR jest tym istotniejsze, że przecież dotyczy ona nie jednej opcji regulacyjnej, ale kilku opcji do wyboru. Dodatkowo obejmują one zawsze również tzw. „opcję zerową” (w nomenklaturze stosowanej przez Komisję Europejską określaną także jako „baseline scenario”), polegającą na pozostawieniu danej kwestii w obszarze ambiwalencji prawnej lub po prostu bez zmian. Czyni się tak wtedy, gdy OSR wskazuje, że dana sytuacja (która zasadniczo wymaga zmiany) może ukształtować się w przyszłości tak, że jakaś nowa interwencja publiczna lub określające ją działania legislacyjne nie są wymagane.

Właśnie najczęściej w odniesieniu do „opcji zerowej” sugeruje się zastosowanie – stosowanej i szerszej – analizy wrażliwości (występującej także w postaci analizy elastyczności). Ta nakazuje zbadanie wpływu istotnych czynników zewnętrznych na założenia skłaniające do uznania sytuacji za niewymagającą dodatkowej interwencji publicznej i zbadaniu ewentualnego wpływu zmiany tych czynników właśnie na te założenia. We wspomnianym ujęciu dotyczącym elastyczności badanie dotyczy wręcz intensywności reakcji założeń na zmianę jakiegoś konkretnego czynnika definiującego przyjęte założenie. Podstawowym problemem tej analizy jest konieczność zbudowania odpowiedniego modelu funkcjonowania określo-

nego, ocenianego systemu, w którym prawidłowo zidentyfikowane są przynajmniej podstawowe relacje przyczynowo-skutkowe. W takim ujęciu model powinien być również przedmiotem oceny ryzyka – tzn. powinien umożliwiać określenie prawdopodobieństwa wystąpienia scenariuszy wiodących do zjawisk niepożądanych z punktu widzenia prawodawcy rozważającego jakąś interwencję legislacyjną. Najbardziej typowym przykładem w zakresie legislacji finansowej jest choćby określenie w tym kontekście momentu złamania krzywej Laffera, wyznaczającej zależność pomiędzy wysokością stawek podatkowych a dochodami budżetowymi. Adekwatna metodologia oceny ryzyka w tym obszarze nabiera szczególnego znaczenia w kontekście planowania działań antykryzysowych i określenia w nich roli instrumentów fiskalnych.

Innym poważnym problemem metodologicznym w ocenie skutków regulacji, zwłaszcza w wymiarze gospodarczym (ale i społecznym), jest identyfikacja modelu ekonomicznego (lub społecznego), który będzie służyć za ramy wnioskowania o związkach przyczynowo-skutkowych pomiędzy planowaną interwencją regulacyjną a rzeczywistością, której ta interwencja ma dotyczyć. Jak się wydaje, na potrzeby OSR nie można przyjmować modelu czysto naukowego – raczej musi to być taki model, który opiera się na obiektywnych twierdzeniach naukowych (uznawanych za prawidłowe i wiodących w danym czasie), ale który jednocześnie uwzględni te istotne (i bywa że intuicyjne) przekonania, które tkwią u podstaw wyborów politycznych. Należy się jednak starać o rozsądny zakres uwzględnienia wspomnianego elementu intuicyjnego – jedynie w takim zakresie, gdy nie kłóci się on z podstawowymi założeniami modelu naukowego. W każdym przypadku należy pamiętać o konsekwentnym stosowaniu przyjętego modelu do wszystkich alternatyw rozważanych w ramach OSR. Trzeba również mieć na uwadze i to, że w przypadku wykonywania na gruncie prawnym aktów prawa międzynarodowego (i wspólnotowego) w zakresie, w którym nie są one – odpowiednio – samowykonalne albo nie stosują się bezpośrednio, podstawowym problemem OSR jest odtworzenie jako parametru tego modelu ekonomicznego, który jest dla danej regulacji założony. W takim przypadku należy więc zebrać i przeanalizować dokumenty źródłowe, pozwalające odtworzyć ten model.

Wszystkie przedstawione w tym rozdziale argumenty wskazują, że bardzo ważnym składnikiem OSR (jak i każdej innej metody ewaluacyjnej) musi być przedstawienie jej metodologii. Jest to konieczne dla zrozumienia wniosków OSR i realizacji jednej z jej podstawowych funkcji, jaką jest umożliwienie decydentom politycznym podejmowania racjonalnych i przygotowanych informacyjnie decyzji dotyczących interwencji publicznych.

Wnioski

Ocena skutków regulacji (OSR) jest jedną z metod ewaluacji interwencji publicznych. Przedmiot owej ewaluacji – interwencja prawna – jest bowiem narzędziem takich interwencji, które – w ramach realizowanej polityki prawa – zasadniczo realizują idee i wartości polityczne, tj. takie, które znajdują się w spektrum powszechnego, demokratycznego wyboru i przez taki wybór są legitymowane.

Jednocześnie jedną z podstawowych funkcji OSR jest zapewnienie, by system prawny (traktowany jako dobro publiczne samo w sobie) spełniał kryteria racjonalności. W najbardziej podstawowym zakresie oznacza to spełnienie wymogu logicznego wzajemnego powiązania wszystkich norm składających się na system prawa. W mocniejszej wersji postulat ten wymaga, aby normy prawne były zdadne do osiągnięcia wyznaczonych i odpowiednio dostosowanych do siebie celów społecznych – w tym zakresie mówimy więc o postulacie logicznej spójności wszystkich skutków osiąganych w procesie stosowania prawa.

Ocena skutków regulacji ma także służyć wyborowi optymalnych wyborów w zakresie interwencji regulacyjnych, w szczególności tych, które niosą ze sobą poważne skutki finansowe dla budżetów publicznych. Optymalizacja w tym zakresie polega na powstrzymaniu się od interwencji regulacyjnej wtedy, gdy dany cel może być osiągnięty bardziej efektywnie i wydajnie środkami pozaregulacyjnymi, ograniczeniu interwencji stosownie do potrzeb (określanym zgodnie z zasadą proporcjonalności), jak również na wyborze tej spośród wielu propozycji legislacyjnych, która reprezentuje najkorzystniejszą proporcję korzyści i kosztów.

Techniki i metodologia OSR są zasadniczo nieźle teoretycznie rozpoznane i – co najważniejsze – odzwierciedlone w regulacjach prawnych dotyczących tworzenia prawa rozwiniętych państw świata (w tym zwłaszcza krajów OECD). Trudności w praktyce sprawia pełne stosowanie tych zasad. Występują one wtedy, gdy wadliwie prowadzona jest polityka prawa (a więc zasadniczo niesprawne są mechanizmy transpozycji idei i wartości politycznych w impulsy regulacyjne, np. ze względu na złą kulturę prawno-polityczną przedkładającą szybkość reakcji regulacyjnej nad jej jakością). Znaczna część trudności jest inherentnie związana z szacunkiem korzyści i kosztów w polityce prawa: problemem jest m.in. adekwatna identyfikacja modelu stanowiącego paradygmat danej interwencji, oszacowanie i porównanie kosztów i korzyści niedających się łatwo ująć w pieniądzu, wybór perspektywy czasowej ewaluacji oraz porównania korzyści i kosztów nierównowartościowych i niewystępujących równolegle. Inherentnym problemem OSR jest także konieczność dopuszczenia ocen i szacunków intuicyjnych, co prowadzi do tego, że ewaluacja w jej ramach jest szczególnie podatna na błędy wnioskowania.

Kwestią problemową w odniesieniu do OSR jest także jego koszt tym większy, im rozleglejsza i głębsza jest jej mapa analityczna. Podobnie jak to jest z innymi problemami, zasadniczo nie da się ich w pełni uniknąć, jednak istnieją metody optymalizacji OSR jako procesu, które warto zastosować.

Artur Nowak-Far – doktor habilitowany nauk prawnych, doktor nauk ekonomicznych, profesor Szkoły Głównej Handlowej w Warszawie, kierownik Katedry Prawa Europejskiego SGH. Autor ponad 160 prac naukowych, w tym 7 książek z dziedziny prawa europejskiego, strategii administracji publicznej oraz ekonomicznej analizy prawa. Członek Kolegium Najwyższej Izby Kontroli.

Bibliografia

- Calabresi G. (1970), *The Cost of Accidents*, Yale University Press, New Haven.
- Easton D., (1981), *The Political System: An Inquiry into the State of Political Science*, University of Chicago Press, Chicago.
- Ellis E., (ed.) (1999), *The Principle of Proportionality in the Laws of Europe*, Hart, Oxford.
- European Commission (2009), *Impact Assessment Guidelines 15 January 2009*, dokument SEC(2009) 92.
- Finnis J. (1980), *Natural Law and Natural Rights*, Clarendon Press, Oxford.
- Finnis J. (1984), *The Authority of Law in the Predicament of Contemporary Social Theory*, "Notre Dame Journal of Law, Ethics and Public Policy", No. 1.
- Gilovich T., Griffin D., Kahneman D. (2002), *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge University Press, Cambridge.
- Haber A., Szałaj M. (2008), *Środowisko i warsztat ewaluacji*, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa.

- Hill M.J. (1997), *The Policy Process in the Modern State*, Prentice Hall, Englewood Cliffs.
- Jacobs S.H. (2008), *Przegląd oceny skutków regulacji (OSR) w krajach OECD*, Organizacja Współpracy Gospodarczej i Rozwoju, Radom-Warszawa.
- Kahneman D., Tversky A., *Choices, Values and Frames*, Cambridge University Press, Cambridge.
- Michalska A., Wronkowska S. (1980), *Zasady tworzenia prawa*, Wyd. UAM, Poznań, 2000.
- Luhmann N. (2004), *Law As a Social System*, Oxford University Press, Oxford.
- Nowak-Far A. (2001), *Zarządzanie strategiczne a analiza efektywności reformy administracji publicznej*, „Służba Cywilna” 2001, nr 2.
- Nowak-Far A. (2007), *Optymalizacja tworzenia i stosowania prawa*, [w:] Ministerstwo Gospodarki (red.), *Reforma Regulacji w praktyce – doświadczenia nowych państw członkowskich Unii Europejskiej*, materiały konferencyjne 9 października 2006.
- Olejniczak K. (2008), *Wprowadzenie do zagadnień ewaluacji*, [w:] Olejniczak K., Kozak M., Ledzion B. (red.), *Teoria i praktyka ewaluacji interwencji publicznych. Podręcznik akademicki*, Wyd. Akademickie i Profesjonalne, Warszawa.
- Olejniczak K., Kozak M., Ledzion (red.), *Teoria i praktyka ewaluacji interwencji publicznych. Podręcznik akademicki*, Wyd. Akademickie i Profesjonalne, Warszawa.
- Palukiewicz A., Gancarz M. (2007), *Nowa metodologia oceny skutków regulacji na tle poprzednich doświadczeń*, [w:] Szpringer W., Rogowski W. (red.), *Ocena skutków regulacji – poradnik OSR, doświadczenia, perspektywy*, C. H. Beck, Warszawa.
- Radaelli C.M. (2003), *European Commission Conference on 'Impact Assessment in the European Union: Innovations, Quality and Good Regulatory Governance'* Background Report 20.
- Radaelli C., Dente B., Jacobs S., Kirkpatrick C., Meuwese A., Renda A. (2006), *Diadem Handbook: How to Perform the Diadem Data Collection*, www.enbr.org.
- Podręcznik akademicki*, Wyd. Akademickie i Profesjonalne, Warszawa.
- Renda A. (2006), *Impact Assessment in the EU: The State of the Art. And the Art of the State*, CEPS, Brussels.
- Surdej A. (2008), *Metoda analizy kosztów i korzyści: imperatyw efektywności w politykach publicznych*, [w:] Haber A., Szałaj M., *Środowisko i warsztat ewaluacji*, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa.
- Wierzbicki J. (2007), *Ocena skutków regulacji w projektach aktów prawnych dotyczących finansów publicznych – obserwacje i wnioski de lege ferenda*, [w:] Szpringer W., Rogowski W. (red.), *Ocena skutków regulacji – poradnik OSR, doświadczenia, perspektywy*, C. H. Beck, Warszawa.
- Wronkowska S., Zieliński M. (2004), *Komentarz do zasad techniki prawodawczej*, Wydawnictwo Sejmowe, Warszawa.
- Zaltman G., Duncan R. (1977), *Strategies for Planned Change*, John Wiley&Sons, New York.
- Zamboni M. (2007), *The Policy of Law: A Legal Theoretical Framework*, Hart Publishing, Oxford.

Część III

**Metodologia badań ewaluacyjnych
– wokół jakości i technologii**

Metodologia jest jednym z najbardziej newralgicznych aspektów wszelkich badań społecznych, a więc również tych stosowanych na potrzeby ewaluacji. Jakość i adekwatność podejść badawczych stosowanych na tym gruncie jest bowiem każdorazowo wystawiana na ciężką próbę, w obliczu zmieniających się warunków społeczno-gospodarczych, czy też kulturowych niuansów. Odpowiedni poziom metodologii jest również kluczowym czynnikiem rozstrzygającym o wiarygodności danego badania ewaluacyjnego. Przekonanie o relacji do rangi zagadnień metodologicznych jest szeroko podzielane przez polską społeczność ewaluacyjną. Wychodząc naprzeciw związanym z tym oczekiwaniom, dedykujemy problematyce metodologicznej, podobnie jak w przypadku wcześniejszych tomów, odrębny rozdział tegorocznej publikacji. Prezentowane w nim teksty poruszają trzy główne wątki – wykorzystanie badań internetowych, budowanie jakości w procesie realizacji badania, oraz aspekt przyczynowości w badaniach ewaluacyjnych.

Badania internetowe są obecnie najdynamiczniej rozwijającą się gałęzią badań marketingowych i społecznych. W niektórych krajach rozwiniętych zyskały już nawet pozycję dominującą, zdobywając nawet bastiony opanowane wcześniej przez badania telefoniczne czy wywiady fokusowe. Trend ten obserwowany jest również na gruncie polskim. Dane gromadzone przez branżę badawczą wskazują, iż na przestrzeni ostatnich kilku lat systematycznie rośnie udział badań internetowych w wolumenie zamówień realizowanych przez polskie instytuty. Internet odgrywa również coraz większą rolę w badaniach naukowych oraz w systemach statystyki publicznej.

Popularność badań prowadzonych w sieci wynika po części z roli, jaką Internet odgrywa we współczesnym społeczeństwie. Zakres zmian wywołanych jego powszechnym stosowaniem jest trudny do przecenienia. Ich skala pozwala mówić wręcz o swoistej rewolucji technologicznej, której owocem są zupełnie nowe jakościowo formy komunikacji społecznej, stąd Internet musi silną rzeczą wywierać wpływ na sposób realizacji badań, w tym również badań ewaluacyjnych.

Polskie doświadczenia dotyczące badań internetowych w ewaluacji nie są, jak dotąd, szczególnie bogate. Bez wątpienia, bariery ograniczające możliwości ich zastosowania wynikają z niewielkiego, w porównaniu z krajami najbardziej rozwiniętymi wykorzystania Internetu. Należy jednak w tym miejscu przyznać, iż na podobne ograniczenia napotykają także polskie instytuty badawcze i naukowe. Istnieją przesłanki pozwalające przyjąć, iż możliwości wykorzystania badań internetowych w ewaluacji są większe niż w wielu innych obszarach. Wynika to chociażby z tego, iż badania ewaluacyjne realizowane są w dobrze zdefiniowanych populacjach, a dotarcie do nich jest ułatwione dzięki bazom danych prowadzonym na potrzeby wdrażania programów. Odpowiednio zaplanowane i dostosowane do specyfiki danej ewaluacji wykorzystanie metod internetowych mogłoby istotnie wzbogacić zakres pozyskiwanych informacji, przy jednoczesnym ograniczeniu kosztów badania.

Kolejnym podejmowanym w tej części publikacji wątkiem jest zagadnienie jakości w środowisku ewaluacji. Jego znaczenie ulega istotnemu wzmocnieniu w kontekście dynamicznego wzrostu liczby badań ewaluacyjnych zlecanych przez polskie instytuty publiczne w ostatnich dwóch latach. Towarzyszy temu dynamiczny rozwój rynku usług ewaluacyjnych wraz z dużą liczbą nowych podmiotów i ekspertów. Specyfika badań ewaluacyjnych sprawia, iż zapewnienie ich właściwej jakości wymaga podjęcia innych działań niż w przypadku standardowych zamówień publicznych. Kluczowym czynnikiem sukcesu jest w tym wypadku zaangażowanie oraz właściwy podział ról i odpowiedzialności pomiędzy zamawiającymi i wykonawcami. Celem włączenia tego wątku do niniejszej publikacji było przekrojowe ukazanie rozwiązań, które znaleźć powinny zastosowa-

nie dla zapewnienia odpowiedniej jakości procesu realizacji badania ewaluacyjnego oraz kontroli wpisanych w jego logikę ryzyk.

Trzeci z głównych wątków tego rozdziału odnosi się do problematyki poruszanej także w naszych poprzednich publikacjach, w których znalazły się teksty poświęcone technice propensity score matching (PSM). Tym razem prezentujemy szersze ujęcie problemu przyczynowości. Postulat wpisania ewaluacji w nowe obszary działalności, stawiany w tekstach zamieszczonych w dwóch poprzednich rozdziałach, pociąga za sobą potrzebę spojrzenia na kluczowe problemy badawcze ewaluacji także z innej perspektywy, niż przyjmowana na etapie ex-post, gdzie wykorzystuje się wspomnianą technikę PSM. Ujęcie problemu przyczynowości, które przedstawiamy, odnosi się więc do ewaluacji prowadzonych w różnych celach, na różnych etapach realizacji polityk publicznych.

CAWI – specyfika, wykorzystanie, perspektywy rozwoju¹

Rozwój badań online

Od kilku lat obserwujemy w Polsce znaczący wzrost zainteresowania badaniami prowadzonymi za pośrednictwem Internetu. Coraz więcej podmiotów prowadzących badania – czy to rynku, psychologiczne, społeczne, czy w szczególności ewaluacyjne – zwraca się w stronę Internetu jako medium pozwalającego na uzyskanie ilościowych i jakościowych danych na temat badanych populacji oraz interesujących zjawisk.

Badania *online* to bardzo pojemna kategoria, w której mieszczą się zarówno stosowane od lat tradycyjne techniki badawcze (odpowiednio dostosowane do specyfiki Internetu), jak i zupełnie nowe typy badań specyficzne tylko dla tego medium. W jej zakres wchodzi zarówno techniki ilościowe, jak i jakościowe badania, polegające na bezpośredniej interakcji z respondentem, jak i badania treści generowanej przez użytkowników Internetu, a także badania ruchu w Internecie i oglądalności stron internetowych. Z jednej strony, internet umożliwia wzbogacanie tradycyjnych technik badawczych (takich jak np. badania ankietowe, FGI, IDI) o dodatkowe elementy, z drugiej strony otwiera przed badaczami niedostępne wcześniej możliwości obserwacji zachowań interesujących grup i analizy wytwarzanych przez nich treści.

Prowadzone przez Internet badania ankietowe, w których respondenci wypełniają ankietę online, zaś wpisywane przez nich odpowiedzi automatycznie zapisują się w bazie danych – *Computer Assisted Web Interview* (CAWI) – to jedna z najczęściej wykorzystywanych internetowych technik badawczych. Badania te powstały wraz z upowszechnieniem się Internetu – w latach 90. – jednak wzrost ich znaczenia obserwujemy dopiero w ostatnich latach.

Rynek badań CAWI na świecie i w Polsce

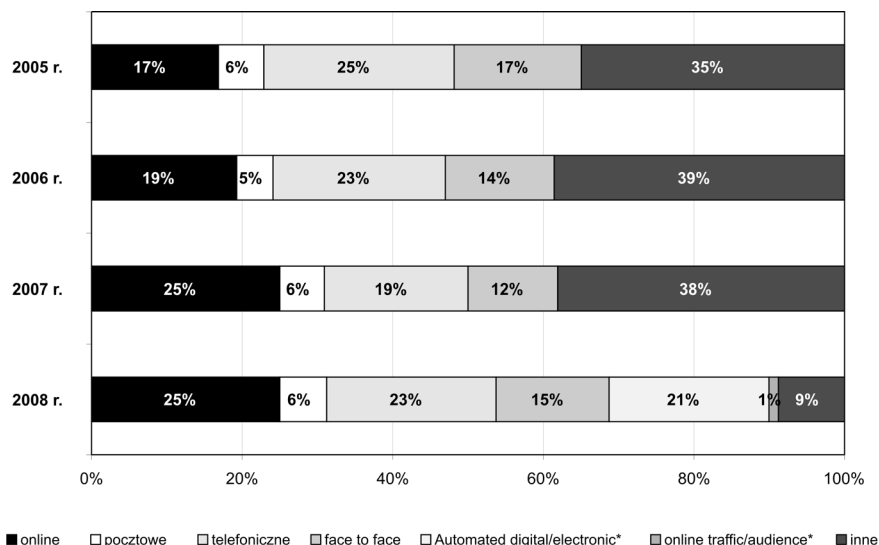
Dane dotyczące struktury wydatków na badania rynku pokazują, że na świecie CAWI nie są już traktowane jako ciekawostka, lecz jako jeden z podstawowych typów badań. Jak informuje ESOMAR², wydatki na ilościowe badania online stanowiły w 2008 roku 20% wszystkich wydatków na badania rynku. Jest to jedna czwarta wszystkich wydatków na badania ilościowe, co plasuje badania online na pierwszym miejscu wykorzystywanych technik ilościowych. Na tradycyjne techniki prowadzenia wywiadów – telefoniczne (*Computer Assisted Telephone Interview* – CATI) i osobiste (*face to face*)³ przeznaczono odpowiednio 18% i 12% całości budżetów badawczych (czyli 23% i 15% wydatków na badania ilościowe).

¹ Autorka pragnie podziękować dr. Janowi Zającowi za cenne uwagi pomocne przy tworzeniu tekstu.

² ESOMAR, *Global Market Research 2009. Esomar Industry Report*, Amsterdam, 2009; dane dotyczą wyłącznie badań oprowadzonych przez agencje i instytuty badawcze i nie uwzględniają działań niekomercyjnych podmiotów, takich jak uniwersytety czy instytucje publiczne.

³ W tej kategorii zawierają się zarówno badania PAPI- Paper and Pencil Interview, jak i CAPI – Computer Assisted Personal Interview

Uwzględniając wydatki na badania z lat ubiegłych obserwujemy ugruntowanie się pozycji badań online: w latach 2005–2007 widoczny był stały wzrost udziału wydatków na te badania w ogólnej puli wydatków; w 2008 roku trend ten uległ zahamowaniu.



Rys. 1. Struktura wydatków na badania ilościowe na świecie

Źródło danych: ESOMAR, *Global Market Research 2009. Esomar Industry Report*, Amsterdam, 2009
 ESOMAR, *Global Market Research 2008. Esomar Industry Report*, Amsterdam, 2008
 ESOMAR, *Global Market Research 2007. Esomar Industry Report*, Amsterdam, 2007.

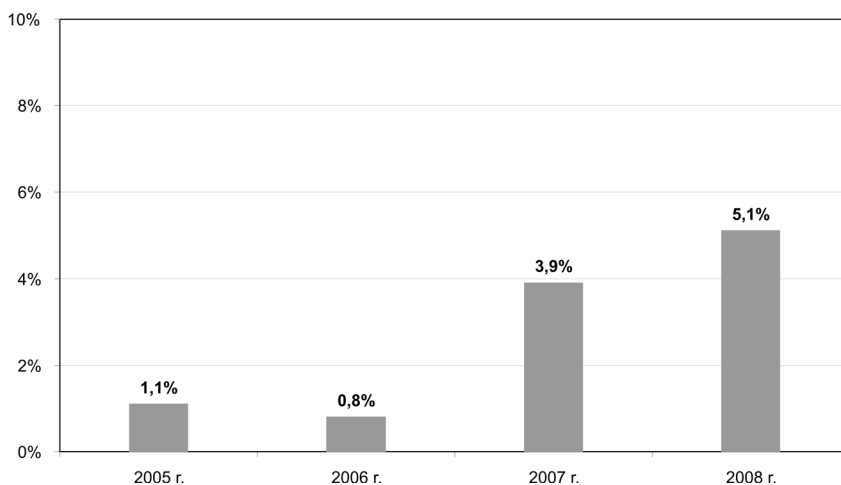
* kategorie „online traffic/audience” i „automated digital/electronic” zostały wyodrębnione z kategorii „inne” dopiero w badaniu dot. 2008 roku.

Warto zwrócić uwagę na fakt, że istnieją rynki, gdzie wydatki na badania ilościowe online stanowią nawet ok. jednej trzeciej wszystkich wydatków na badania, co przekłada się na 40%–50% wydatków na badania ilościowe (Kanada, Japonia i Australia).

Obraz rynku badań w Polsce odbiega od opisanego powyżej – odsetek wydatków na badania online jest wielokrotnie niższy niż średnia światowa⁴. Trend wzrostowy jest jednak wyraźny: według danych Polskiego Towarzystwa Badaczy Rynku i Opinii opartych na ankiecie wśród 45 firm badawczych, osoby, które wzięły udział w badaniach prowadzonych techniką CAWI, stanowią 5,1% wszystkich osób uczestniczących w badaniach w 2008 roku. Udział ten jest znacznie wyższy niż w ubiegłych latach; w 2007 roku wynosił 3,9%, w 2006 – 0,8%, w 2005 – 1,1%. Oznacza to, że w 2008 roku w badaniach CAWI udział wzięło ponad 300 tysięcy respondentów, przy czym należy podkreślić, że faktyczna liczba jest wyższa ze względu na brak udziału w ankiecie niektórych firm badawczych prowadzących badania online.

Wydaje się więc, że najszybszy rozwój badań online czeka Polskę w najbliższych latach. Żeby jednak móc na ten temat spekulować, warto przyjrzeć się specyfice badań online, źródłom ich odmienności od badań *face to face* i CATI, przyczynom wzrostu znaczenia badań online, a także problemom i barierom ich rozwoju.

⁴ Według ESOMARU, w Polsce wydatki na ilościowe badania online stanowią zaledwie 1% wszystkich wydatków badawczych; zdecydowanie dominują badania *face to face*, na które przeznaczanych jest 32% wydatków.



Rys. 2. Odsetek liczby badanych osób metodą CAWI w ogólnej liczbie badanych

Źródło: PTBriO.

Atuty badań CAWI

Próbując streścić w jednym zdaniu przyczyny rosnącej popularności badań online, można by powiedzieć: Opierając się na nowych technologiach i umożliwiając atrakcyjną z punktu widzenia respondenta formę kontaktu, pozwalają one na szybkie uzyskiwanie dużych prób stosunkowo niskim kosztem. Przyjrzyjmy się jednak dokładniej tym atutom.

Wykorzystanie Internetu i najnowszych technologii pozwala na przygotowanie interaktywnego narzędzia badawczego, w którym możliwe jest prezentowanie badanym różnorodnych materiałów multimedialnych (zdjęć, filmów, ścieżek dźwiękowych), a także zasobów internetowych, do których respondenci mogą być odsyłani w trakcie badania lub pomiędzy kolejnymi jego etapami. Zbliżone możliwości dają też badania CAPI, jednak już w CATI możliwe jest jedynie prezentowanie materiałów dźwiękowych.

Internet umożliwia też prowadzenie badań o dużym zasięgu geograficznym. Jednocześnie dzięki wykorzystaniu informacji z adresu IP użytkownika, możliwe jest dotarcie do osób z określonego⁵ miasta lub regionu świata, co pozwala np. na prowadzenie badań na emigrantach⁶. CAWI umożliwiają też badanie innych trudno dostępnych grup celowych: np. osób zainteresowanych określoną tematyką (poprzez wysyłkę zaproszeń do badania uczestnikom określonych forów / grup dyskusyjnych lub emisję ankiety w formie *pop-up* do użytkowników witryn związanych z daną tematyką) lub osób rzadko przebywających w domu.

Z punktu widzenia badanych, technika ta jest nieinwazyjna, wypełnianie ankiety nie wiąże się z ingerencją w życie prywatne obcej osoby (ankietera), komfortowa jest też możliwość wypełnienia ankiety w dogodnym czasie⁷. Brak ankietera sprawia też, że ograniczony jest efekt dostosowywania odpowiedzi do norm społecznych i panujących poglądów (tzw. *social desirability effect*), zmniejszona jest także

⁵ Rozpoznawanie lokalizacji oparte na tych danych nie jest jednak w 100% skuteczne. W badaniach zadaje się więc respondentom dodatkowe pytania, które pozwalają zweryfikować informację opartą na adresie IP.

⁶ Przykładem może być przeprowadzone Interaktywny Instytut Badań Rynkowych w 2008 roku badanie Polaków mieszkających w Wielkiej Brytanii i Irlandii; por. M. Milewski, J. Ruszczak-Żbikowska, *Motywacje do wyjazdu, praca, więzi społeczne i plany na przyszłość polskich migrantów przebywających w Wielkiej Brytanii i Irlandii*, CMR Working Papers, Nr 35/(93) Warszawa, 2008.

⁷ Tylko w przypadku części ankiet CAWI.

tendencja do udzielania odpowiedzi stawiających badanego w korzystnym świetle. Wszystko to daje możliwość zbierania szczerych odpowiedzi i badania kwestii drażliwych. Większa jest też standaryzacja badania dzięki eliminacji wpływu ankietera, który przez swoje zachowanie czy intonację może sugerować (często nieświadomie) konkretne odpowiedzi.

Dzięki jednoczesnemu dotarciu do dużej liczby respondentów skrócić można czas zbierania danych. Innym powodem krótszego czasu badania jest brak konieczności wprowadzania odpowiedzi respondentów do bazy danych.

Bezpośredni zapis odpowiedzi badanych w bazie danych bez pośrednictwa ankietera i osoby kodującej ma też poważniejsze konsekwencje: pozwala wyeliminować błędy mogące wystąpić podczas kodowania danych, a także ustrzec się fałszerstw ze strony ankieterów.

Warto też wspomnieć, że podobnie jak i w innych typach badań wspomaganych komputerowo, reguły przejścia między kolejnymi pytaniami i blokami pytań (filtry) są zaprogramowane w ankiecie, co pozwala na uniknięcie błędów wynikających z niepełnego opanowania tych reguł przez ankieterów. Wykorzystywana jest też automatyczna rotacja odpowiedzi w kafeterii, a także rotacja pytań i bloków pytań.

Badania CAWI dają jeszcze jedną istotną możliwość: oprócz odpowiedzi badanych, zapisywane są też inne dane dotyczące procesu odpowiadania (tzw. *paradata*), jak np. dokładny czas udzielania odpowiedzi na kolejne pytania czy moment przerwania wywiadu⁸. Pozwala to na analizę zachowań respondentów i ewentualne poprawienie źle zaprojektowanych pytań.

W przypadku badań dotyczących zachowań internetowych, ważnym atutem badań CAWI jest też możliwość połączenia danych deklaracyjnych z informacjami o zachowaniu użytkowników na witrynach zebranych w oparciu o pliki *cookie*⁹, co z jednej strony może pozwolić na precyzyjny dobór uczestników badania (uzależniony od odwiedzenia konkretnych zasobów internetowych lub wykonania określonej akcji), z drugiej – na konfrontację deklaracji i zachowań respondentów.

Na koniec warto podkreślić argument ekonomiczny, nierzadko kluczowy przy podejmowaniu decyzji o wyborze techniki badawczej: niski koszt badania. Szacuje się, że badanie internetowe jest o ok. jedną czwartą tańsze od wywiadu telefonicznego i o połowę tańsze od wywiadu *face to face*¹⁰. W szczególności dotyczy to badań zakrojonych na dużą skalę, gdyż najczęściej dotarcie do kolejnych respondentów (po zaprojektowaniu i uruchomieniu badania) nie wiąże się już z dużym kosztem. W konsekwencji możliwe jest uzyskiwanie dużych prób.

Specyfika badań prowadzonych metodą CAWI

Wymienione wyżej atuty badań ilościowych prowadzonych przez Internet nie oznaczają jednak, że jest to metoda, którą można powszechnie stosować do wszelkiego typu badań ankietowych. Konieczne jest zrozumienie ograniczeń tej techniki badawczej, a także wskazanie metod radzenia sobie z problemami specyficznymi dla badań prowadzonych online.

⁸ Takie możliwości oferują też badania poza Internetem, w których respondent sam udziela odpowiedzi bezpośrednio wprowadzając je do komputera.

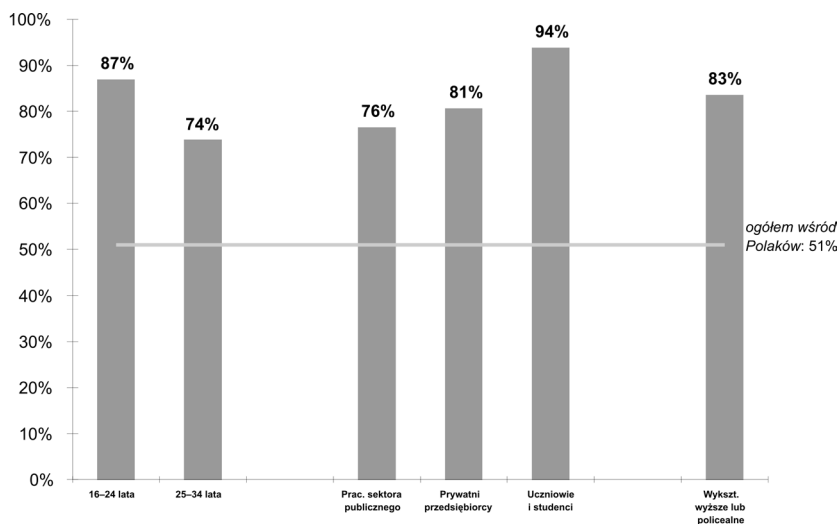
⁹ Możliwe jest to tylko w sytuacjach, w których jednoznaczność cookie i użytkownika jest wysoko prawdopodobna. Ograniczeniem są tu zjawiska współkorzystania z komputerów (przeglądarek) przez kilka osób, korzystania z wielu komputerów (przeglądarek) przez jedną osobę oraz kasowalności cookie.

¹⁰ ESOMAR, *Global Prices Study 2007*, 2007 za: ESOMAR, *Global Market Research 2008. Esomar Industry Report*, Amsterdam, 2008.

Porównując i oceniając różne techniki prowadzenia badań ankietowych, wśród potencjalnych źródeł błędów wskazuje się¹¹: (1) stopień pokrycia badanej populacji, (2) sposób doboru próby, (3) stopień realizacji próby (występowanie jednostek niedostępnych; błąd braku odpowiedzi – *nonresponse*) oraz (4) sam pomiar, czyli odpowiedzi udzielane przez badanych.

(1) W badaniach prowadzonych metodą CAWI respondentami z definicji mogą być tylko osoby korzystające z Internetu¹². W sytuacji gdy na podstawie danych z badania CAWI wnioskuje się o populacjach, które tylko w części mają dostęp do tego medium (np. wszystkich Polakach), wystąpić może tzw. błąd pokrycia (*coverage error*), będący konsekwencją różnic między osobami, które miały szansę uczestnictwa w badaniu, a tymi, które z racji nieużywania Internetu takiej możliwości nie miały. Wielkość tego błędu zależy z jednej strony od skali tych różnic, z drugiej – od proporcji tych grup, czyli od penetracji Internetu w badanej populacji. Generalizacja wniosków z badań CAWI np. na całą populację Polaków jest na tym etapie rozwoju Internetu co najmniej problematyczna biorąc pod uwagę silną korelację dostępu do internetu ze zmiennymi demograficznymi (takimi jak wiek, wykształcenie czy zarobki), a także stylem życia¹³.

Badania CAWI powinno się więc wykorzystywać tylko gdy badanie dotyczy grup, w których penetracja Internetu jest wysoka. W szczególności warto zwrócić uwagę na osoby młode, dobrze wykształcone, a także przedsiębiorców czy uczniów i studentów.



Rys. 3. Penetracja internetu w wybranych grupach demograficznych

Źródło: Batorski D., *Korzystanie z technologii informacyjno komunikacyjnych*,

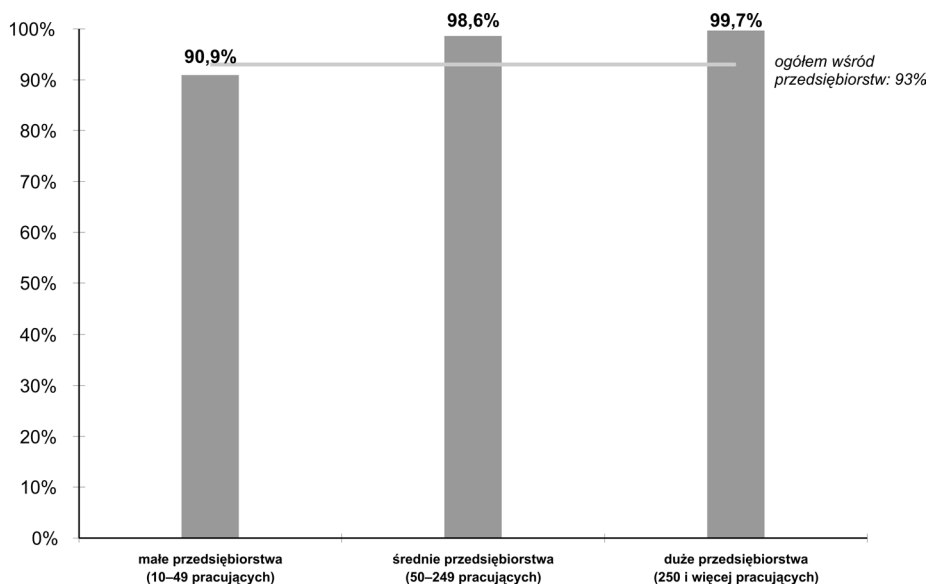
[w:] Czapiński J., Panek T., *Diagnoza społeczna 2009. Warunki i jakość życia Polaków*, s. 279–307, Warszawa, 2009.

¹¹ Por. m.in. M. Couper, *Web surveys: a review of issues and approaches*, Public Opinion Quarterly, 64 (4), 2000, s. 464–481; D. Dillman, D. Bowker, *The Web Questionnaire Challenge to Survey Methodologists*, [w:] U. Reips, M. Bosnjak (red.), *Dimensions of Internet Science*, Lengerich, Niemcy, 2001.

¹² Możliwe są wprowadzenie badania, w których wylosowanym do badania osobom zapewnia się dostęp do internetu specjalnie na potrzeby badania (i faktycznie takie próby mają miejsce – por. np. D. Batorski, M. Olcoń, *Prowadzenie badań przez Internet: Podstawowe zagadnienia metodologiczne*. Studia Socjologiczne, 3(182), 2006, s. 99–132), jednak tego typu badania są na tyle kosztowne, że nie wydaje się by miało to być wykorzystywane na szerszą skalę.

¹³ Dane o różnicach między internautami a osobami niekorzystającymi z Internetu na podstawie D. Batorski, *Korzystanie z technologii informacyjno komunikacyjnych*, [w:] J. Czapiński, T. Panek, *Diagnoza społeczna 2009. Warunki i jakość życia Polaków*, Warszawa, 2009, s. 279–307.

Uzasadnione jest też wykorzystywanie techniki CAWI do badania przedsiębiorstw. Jak informuje GUS¹⁴, 93% przedsiębiorstw zatrudniających co najmniej 10 pracowników ma dostęp do Internetu. Biorąc pod uwagę, że dane te pochodzą ze stycznia 2008 (nowsze dane nie są jeszcze dostępne), można szacować, że dziś penetracja Internetu bliska jest już 100% i błąd pokrycia przy prowadzeniu badań CAWI praktycznie nie występuje.



Rys. 4. Odsetek przedsiębiorstw z dostępem do Internetu

Źródło: Główny Urząd Statystyczny, *Wykorzystanie technologii informacyjno-telekomunikacyjnych w przedsiębiorstwach, gospodarstwach domowych i przez osoby prywatne w 2008 r.*, Warszawa 2008.

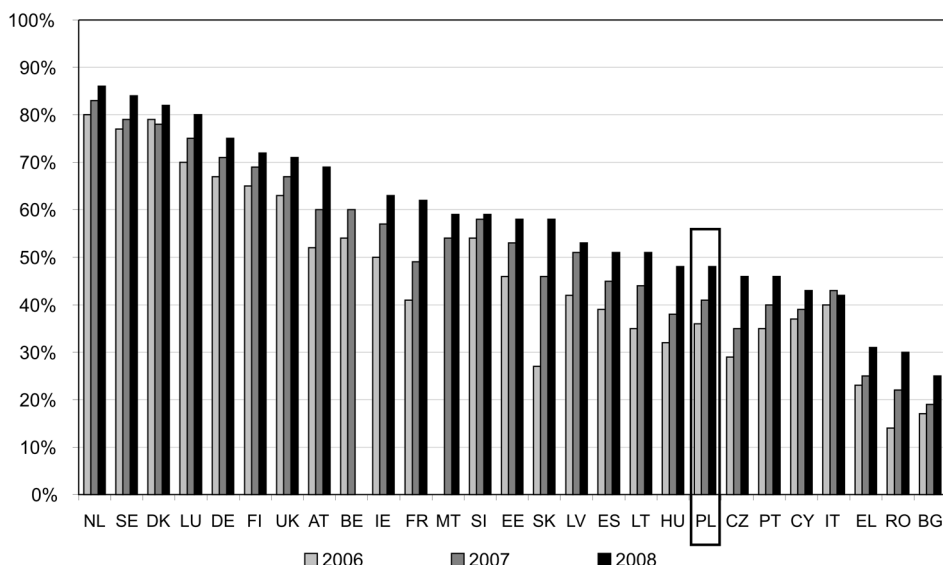
Nie można też wykluczyć możliwości realizacji za kilka lat badań CAWI uogólnianych na generalną populację Polaków. Penetracja Internetu w naszym kraju rośnie bardzo szybko. Już połowa Polaków korzysta z Internetu¹⁵, co oznacza dwukrotny wzrost w ciągu ostatnich 6 lat; jeśli zaś spojrzeć na dostęp gospodarstw domowych do Internetu, wzrost w tym czasie jest aż trzykrotny: obecnie ponad połowa gospodarstw domowych ma dostęp do Internetu¹⁶. Nic nie wskazuje na to, by ten trend miał się zatrzymać; obserwując zmiany w innych krajach europejskich widać, iż wzrost dostępu jest charakterystyczny także dla krajów o znacznie większej penetracji Internetu niż Polska.

(2) Kolejną istotną kwestią przy projektowaniu badań CAWI jest dobór próby. W przypadku znaczącej części populacji badanych online brak jest operatu losowania (listy jednostek, z których można odpowiednio dobrać próbę). Wprawdzie jednostki z populacji możemy rozpoznawać dzięki różnym identyfikatorom (przede wszystkim e-mailom oraz plikom *cookie*), rzadko jednak badacz dysponuje pełną listą tych identyfikatorów dla populacji, którą chce badać. W wielu sytuacjach konieczne jest więc stosowanie nieprobabilistycznych metod doboru próby, przy czym nie jest to sytuacja specyficzna dla CAWI, lecz

¹⁴ Główny Urząd Statystyczny, *Wykorzystanie technologii informacyjno-telekomunikacyjnych w przedsiębiorstwach, gospodarstwach domowych i przez osoby prywatne w 2008 r.*, Warszawa, 2008.

¹⁵ Wyniki podawane przez agencje i instytucje badawcze nieco różnią się między sobą, jednak wszystkie oscylują wokół 50%. Z najważniejszych pomiarów wymienić warto: Diagnoza Społeczna 2009 – penetracja równa 51% w grupie osób 16 lat i więcej (marzec 2009); NetTrack Millward Brown SMG/KRC – penetracja równa 49% w grupie osób w wieku 15 lat i więcej (luty-kwiecień 2009);

¹⁶ D. Batorski, ibidem.



Rys. 5. Odsetek gospodarstw domowych z dostępem do internetu w krajach europejskich

Źródło danych: Loof A.: Eurostat, *Science and technology. Population and social conditions. Industry, trade and services, Data in focus 46/2008*, 2008.

dotyczy także badań *face to face* czy CATI, których duża część robiona jest z wykorzystaniem prób kwotowych. Choć statystyczna teoria dotycząca uogólniania badań reprezentacyjnych na populację dotyczy tylko probabilistycznych metod doboru próby (i to w pełni zrealizowanych), w praktyce z powodzeniem udaje się określać parametry w populacji na podstawie prób dobieranych kwotowo i/lub analizowanych z użyciem wag analitycznych.

Realizacja badań typu CAWI najczęściej wykorzystuje jeden z trzech sposobów doboru próby i dotarcia do respondentów:

- a. Możliwy jest dobór próby spośród członków zrekrutowanego wcześniej panelu badawczego (tzw. *access panel*¹⁷) i wysyłka zaproszeń do badania w formie e-maili. Dobierając próbę do konkretnego badania wykorzystuje się wiedzę na temat członków panelu (informacje z ankiety rekrutacyjnej czy ankiet weryfikacyjnych) oraz dane zewnętrzne na temat badanej populacji (przede wszystkim struktura demograficzna); najczęściej stosuje się dobór kwotowy. Jakość takich prób uzależniona jest m.in. od sposobu rekrutacji członków panelu, algorytmu doboru do konkretnego badania, wielkości panelu, czy dostępności danych na temat uczestników panelu¹⁸. Największe na świecie agencje badawcze chwalą się panelami liczącymi kilkadziesiąt lub kilkaset tysięcy, a nawet kilka milionów uczestników¹⁹.

¹⁷ Nie należy mylić z panelami służącymi pomiarowi oglądalności witryn (w Polsce jest to Megapanel PBI/Gemius); więcej na temat różnych paneli internetowych m.in. [w:] R. Mącik, *Wykorzystanie internetu w badaniach marketingowych*, Wydawnictwo UMCS, Lublin 2005, s. 133–140.

¹⁸ Por. m.in. ESOMAR (2008), *26 Questions To Help Research Buyers Of Online Samples*, dostępne: <http://www.esomar.org/uploads/pdf/professional-standards/26questions.pdf>

¹⁹ Przykładowo, amerykański Harris Interactive podaje informację o łącznej liczbie uczestników paneli w 125(!) krajach – jest to ponad 6 milionów osób; brytyjski YouGov chwali się panelem mieszkańców Wielkiej Brytanii liczącym ponad 250 tys. uczestników.

- b. Inną metodą jest losowy²⁰ dobór *cookies* odwiedzających określone strony, a następnie wyświetlenie użytkownikom identyfikowanym przez te *cookies* zaproszenia do badania w formie okienka *pop-up* lub innych form kreacji reklamowych (określane też jako *real-time sampling*). Jeśli badanie dotyczy populacji użytkowników tych stron, taki typ doboru próby klasyfikowany jest przez badaczy jako losowy²¹, przy czym trzeba pamiętać o ograniczeniach tej losowości związanych z faktem, iż *cookie* nie zawsze jest unikalnym identyfikatorem użytkownika²². Ten typ doboru próby nie ogranicza respondentów do osób zarejestrowanych w panelu, co pozwala uniknąć obciążenia wyników (tzw. *panel bias*); z drugiej strony, motywacja respondentów badanych tą metodą jest często niższa niż w przypadku panelu, co powoduje niższy stopień realizacji próby.
- c. W przypadku posiadania bazy jednoznacznych identyfikatorów (najczęściej e-maili) wszystkich członków badanej populacji możliwe są losowe metody doboru próby i wysyłka zaproszeń do badania do tak określonej listy adresatów. Taki dobór próby jest możliwy tylko w przypadku badania niektórych populacji ze względu na trudność dostępu do danych kontaktowych; z całą pewnością jest on jednak pożądany i budzący najmniej kontrowersji. Metoda taka możliwa jest np. przy badaniu zarejestrowanych użytkowników jakiegoś serwisu czy usługi, przy badaniach studentów (z których wszyscy posiadają oficjalny uczelniany adres e-mail) czy też w badaniach ewaluacyjnych, gdy dostępne są dane kontaktowe beneficjentów.

(3) Odpowiedni dobór próby nie zapewni jednak wysokiej jakości próby, jeśli wylosowane do próby jednostki nie wezmą udziału w badaniu. Jest to tzw. problem braku odpowiedzi (*non-response error*), którego skala zależy z jednej strony od stopnia realizacji próby (*response rate*), z drugiej – od różnic między osobami, które wypełniły ankietę i tymi, które jej nie wypełniły. Warto tu rozróżnić trzy możliwe problemy na kolejnych etapach realizacji badania:

- kwestia dotarcia do wybranych respondentów z zaproszeniem do badania: część maili z zaproszeniem może nie dotrzeć do adresatów (nieaktualne adresy, filtry antyspamowe itp.) część zaproszeń np. w formie okienka *pop-up* może się nie wyświetlić wybranym użytkownikom (np. z powodu blokady *pop-upów*);
- kwestia skorzystania z zaproszenia do badania, czyli rozpoczęcie wypełniania ankiety: nie wszyscy zaproszeni do badania decydują się wziąć w nim udział;
- kwestia wypełnienia do końca kwestionariusza badawczego: część osób rezygnuje z udziału w badaniu w trakcie wypełniania ankiety – tzw. *drop-out*.

Niski stopień realizacji próby, nierzadko na poziomie zaledwie kilku procent jest jednym z najpoważniejszych problemów badań CAWI²³ i w ostatnich latach prowadzonych jest wiele badań mających na celu opracowanie technik podnoszących *response rate*²⁴.

²⁰ Możliwy jest zarówno prosty dobór losowy, jak i stosowanie algorytmów losowo-warstwowych, gdzie do warstwowania wykorzystywana jest informacja przechowywana w *cookie*, por. J. Przewłocka, M. Półtorak: Gemius SA, *Towards Unbiased Sampling in Online Survey Research. Use of Information Stored in Cookies as a Method of Improving Sample Structure Quality*, plakat na konferencji General Online Research, Hamburg, 2008.

²¹ Por. S.J. Best, B.S. Krueger, *Internet data collection*. Sage Publications, Londyn, 2004; J.M. Zając, D. Batorski, *Jak skłonić do udziału w badaniach internetowych: Zwiększanie realizacji próby*, „Psychologia Społeczna”, nr 3, 2007, 234–247.

²² Związane jest to ze zjawiskiem współkorzystania z komputerów przez kilka osób i korzystanie przez jedną osobę z więcej niż jednego komputera.

²³ Raportowane przez różnych badaczy i ośrodki badawcze wskaźniki RR są bardzo zróżnicowane: od kilku do kilkudziesięciu procent (J.M. Zając, D. Batorski, *Jak skłonić do udziału w badaniach internetowych: Zwiększanie realizacji próby*, „Psychologia Społeczna”, nr 3, 2007, s. 234–247). Jednocześnie warto pamiętać, że problem niskich RR nie jest unikalnym problemem badań CAWI; publikowane są informacje o RR w badaniach CATI na poziomie kilkunastu procent (P. Kellner, *Can online Polls produce accurate findings?*, „International Journal of Market Research”, Vol. 46, Quarter 1, 2004; C. Roster, R. Rogers, G. Album, D. Klein, *A comparison of response characteristics from Web and telephone surveys*, „International Journal of Market Research”, Vol. 46, Quarter 3, 2004).

²⁴ Konkretnie rozwiązania przytaczane są np. [w:] J.M. Zając, D. Batorski, *Jak skłonić do udziału w badaniach internetowych: Zwiększanie realizacji próby*, „Psychologia Społeczna”, nr 3, 2007, s. 234–247 oraz w pracy tych autorów w tym tomie.

Rzetelnie prowadzone badania internetowe powinny minimalizować każdy z opisanych trzech rodzajów błędów (błąd pokrycia, błąd związany z doбором próby oraz jednostkami niedostępnymi) – poprzez stosowanie techniki CAWI tylko do badania grup z wysoką penetracją Internetu, poprzez wybór właściwych algorytmów doboru próby i poprzez maksymalizację stopnia realizacji próby. W praktyce jednak wszystkie badania obarczone są przynajmniej jednym (a nierzadko wszystkimi) z tych błędów. W mniejszym lub większym stopniu dotyczy to wszystkich typów badań ankietowych, także prowadzonych metodami tradycyjnymi (poza Internetem), jednak faktem jest, że w przypadku badań CAWI problemy te są szczególnie dolegliwe i znacznie poważniejsze niż w przypadku badań bezpośrednich. Standardowo stosuje się więc korekcję tych błędów jeszcze na etapie zbierania danych (np. poprzez zastąpienie jednostek niedostępnych innymi respondentami o podobnej charakterystyce) oraz na etapie analizy już zebranych danych poprzez nadanie respondentom wag analitycznych, obliczanych na podstawie danych referencyjnych dot. badanej populacji. Najczęściej stosowane metody obliczania tych wag to ważenie oparte o rozkład łączny cech strukturalnych oraz ważenie według rozkładów brzegowych, a także – choć znacznie rzadziej – ważenie oparte o wskaźnik dostępności (*propensity-score weighting*). Kluczowy dla skuteczności ważenia jest dobór właściwych zmiennych – skorelowanych zarówno z badanymi zjawiskami, jak i wpływających na niedostępność jednostek.

Sukcesy²⁵ badań prowadzonych metodami CAWI pokazują, że mimo opisanych wyżej problemów badania te mogą precyzyjnie mierzyć zjawiska społeczne.

(4) Niezależnie od problemów związanych z badanymi jednostkami, w badaniach ankietowych występuje też ryzyko udzielania przez badanych odpowiedzi nie w pełni odzwierciedlających ich opinie lub źle raportujących fakty, o które są pytani. W przypadku badań prowadzonych metodą CAWI może to być spowodowane z jednej strony niewłaściwym zaprojektowaniem kwestionariusza powodującym, że respondenci mylnie interpretują pytania i odpowiedzi, z drugiej strony – związanymi z brakiem ankietera anonimowością i niższą motywacją. Nierzetelne odpowiedzi mogą też wiązać się z chęcią wzięcia udziału w badaniu w przypadku paneli badawczych, których członkowie wynagradzani są za wypełnianie ankiet. Konieczne jest stosowanie technik wykrywania nierzetelnych respondentów i usuwania ich z bazy danych; techniki te opierają się m.in. na umieszczaniu w kwestionariuszu dodatkowych pytań i/lub porównywaniu odpowiedzi na poszczególne pytania²⁶.

Konsekwencje braku ankietera

Problem braku odpowiedzi oraz specyfika błędów pomiaru w badaniach CAWI są silnie związane z podstawową cechą, jaka odróżnia te badania od tradycyjnych badań *face to face* i CATI: w badaniach internetowych (podobnie jak np. pocztowych) nie występuje ankieter. Warto przyrzeć się bliżej konsekwencjom tego stanu rzeczy.

Po pierwsze – i najważniejsze z punktu widzenia problemu *nonresponsu* – wywiad w Internecie nie jest sytuacją społeczną (z którą mamy do czynienia w przypadku wywiadu ankieterskiego). Niższe jest więc zaangażowanie respondenta, a odmowa udziału lub rezygnacja w trakcie wypełniania ankiety nie pociągają za sobą żadnych konsekwencji. Z drugiej strony, wyeliminowany jest opisywany wcześniej nie-

²⁵ Np. bardzo precyzyjnie przewidziane przez Harris Interactive wyniki ostatnich wyborów prezydenckich w USA (więcej: <http://ir.harrisinteractive.com/releasedetail.cfm?ReleaseID=396524>), czy sukcesy agencji YouGov w przewidywaniu wyników wyborów w Wielkiej Brytanii (P. Kellner, *Can online Polls produce accurate findings?*, International Journal of Market Research, Vol. 46, 2004, Quarter 1).

²⁶ Por. M. Milewski, J. Wojciechowski, *Na tropach „pozerów”, „klikaczy” i „psotników”. Techniki identyfikacji nierzetelnych respondentów w badaniach internetowych*, prezentacja na X Kongresie Badaczy Rynku i Opinii, Warszawa, 15–16 października 2009.

korzystny wpływ ankietera na odpowiedzi badanych; uniemożliwione są jakiekolwiek sugestie czy odpowiedzi, które modyfikować mogą odpowiedzi badanych.

Oznacza to jednak także, że w badaniach CAWI jedynym źródłem informacji dla respondenta jest zaproszenie do badania i sama ankieta. Ich przygotowanie jest więc kluczowe dla przekonania wybranego do próby respondenta do wzięcia udziału w badaniu, utrzymania jego motywacji przez cały wywiad i właściwej interpretacji pytań przez respondenta²⁷. Warto tu podkreślić, że w ankiecie do samodzielnego wypełniania istotne są nie tylko słowa, ale cały język wizualny wykorzystany w kwestionariuszu – wygląd i format pytań, oznaczenia numeryczne, symbole, kolory, *layout* całej ankiety²⁸. W związku z kontekstem badania (respondenci biorą udział w badaniu między lub nawet w trakcie wykonywania innych czynności w Internecie – przeglądania stron, pisanie maili itp.) ważne jest też dostosowanie kwestionariusza do reguł panujących w Internecie, w tym uwzględnienie zasad prezentacji treści w tym medium.

Powyższy skrótowy przegląd najważniejszych kwestii dotyczących badań CAWI pokazuje wielki potencjał tych badań a jednocześnie szereg ograniczeń, z których część z pewnością zostanie wyeliminowana wraz z dalszym upowszechnianiem się Internetu. Warto zwrócić uwagę na fakt, że rosnące znaczenie internetu wpływa na rozwój badań CAWI z kilku względów:

- Dostęp do internetu ludzi i instytucji stale rośnie; coraz więcej populacji może być badanych tą drogą; Zarówno w życiu prywatnym, jak i zawodowym komputery i Internet używane są coraz intensywniej (wystarczy choćby spojrzeć na średni czas spędzany w Internecie – na przestrzeni 4 lat, od listopada 2004 do listopada 2008 wzrósł on z nieco ponad pół godziny do blisko półtorej godziny dziennie²⁹). W konsekwencji zwiększają się kompetencje związane z wykorzystywaniem Internetu, komunikacja zapośredniczona przez komputer staje się dla użytkowników Internetu coraz bardziej naturalna, dzięki czemu udział w badaniach internetowych, w tym samodzielne wypełnianie ankiet CAWI, staje się dla potencjalnych respondentów prostsze i bardziej naturalne;
- Zwiększa się znaczenie Internetu w życiu codziennym, publicznym, społecznym, rozwijają się usługi online, zwiększa się rola komunikacji za pośrednictwem Internetu oraz jego rola informacyjna. Powoduje to, że zjawiska związane z Internetem budzą coraz większe zainteresowanie badaczy. Internet jako przedmiot zainteresowania może być oczywiście badany także metodami tradycyjnymi (poza Internetem), jednak szerokie możliwości badań CAWI oraz możliwość badania zjawisk w ich naturalnym kontekście powodują, że to one szczególnie często wybierane są do badania tych właśnie tematów.

Nic nie wskazuje na to, by trendy te miały ulec zmianie. W połączeniu z opisanymi wcześniej zaletami badań prowadzonych metodą CAWI można więc oczekiwać dalszego wzrostu popularności tych technik. W szczególności dotyczy to Polski, gdzie jak widzieliśmy, skala stosowania tych badań jest jeszcze bardzo mała w porównaniu z największymi rynkami światowymi.

²⁷ Wpływ formatu i wyglądu pytań na *drop-out* i odpowiedzi badanych omawiany jest m.in. [w:] Dillman D., Christiam M., J. Smyth, *Helping respondents get it right the first time: the influence of words, symbols, and graphics in web surveys*, Public OpinionQuarterly, Vol. 71, No. 1, Spring 2007, pp. 113–125; J. Przewłocka, P. Janczewski: Gemius SA, *Jak e-Kowalski uczył nas zadawać pytania*. Referat wygłoszony na VIII Kongresie Badaczy Rynku i Opinii, Warszawa, 2007.

²⁸ L. Christian, D. Dillman and J. Smyth *The Effects of Mode and Format on Answers to Scalar Questions in Telephone and Web Surveys*, wersja robocza prezentacji na 2nd International Conference on Telephone Survey Methodology, January 12, 2006 in Miami, Florida, 2006,

²⁹ Dane Megapanel PBI/Gemius opublikowane w Gemius, *Polski internet 2008/2009*, Warszawa, 2009.

Jadwiga Przewłocka – absolwentka socjologii na Uniwersytecie Warszawskim i matematyki na Uniwersytecie Paris 13. Starszy Badacz w Dziale Metodologii Gemius SA, firmy specjalizującej się w badaniach internetowych. Badaniami online zajmuje się od czterech lat, w tym czasie realizowała przede wszystkim projekty z wykorzystaniem techniki CAWI, prowadziła też badania jakościowe (FGI online), analizy ruchu na witrynach oraz analizy skuteczności kampanii reklamowych online. Obecnie zajmuje się rozwojem metodologii CAWI i opracowywaniem nowych internetowych produktów badawczych.

Bibliografia

- Batorski D. (2009), *Korzystanie z technologii informacyjno komunikacyjnych*, [w:] Czapiński J., Panek T., *Diagnoza społeczna 2009. Warunki i jakość życia Polaków*, Warszawa.
- Batorski D., Olcoń M. (2006). *Prowadzenie badań przez internet: Podstawowe zagadnienia metodologiczne*. Studia Socjologiczne, 3(182).
- Best S. J., Krueger B. S. (2004), *Internet data collection*. Sage Publications, Londyn.
- Christian L., Dillman, D. and Smyth, J. (2006), *The Effects of Mode and Format on Answers to Scalar Questions in Telephone and Web Surveys*, wersja robocza prezentacji na 2nd International Conference on Telephone Survey Methodology, January 12, 2006 in Miami, Florida, dostępne: <http://www.sesrc.wsu.edu/dillman/papers.htm>
- Christian M., (2000), *The Influence of Visual Layout on Scalar Questions in Web Surveys*. Niepublikowana praca magisterska, Washington State University, USA.
- Couper M. (2000), *Web surveys: a review of issues and approaches*, Public Opinion Quarterly. 64 (4).
- Dillman D., Bowker D. (2001), *The Web Questionnaire Challenge to Survey Methodologists*, [w:] Reip U.S. & Bosnjak M. (red.), *Dimensions of Internet Science*, Lengerich, Niemcy.
- ESOMAR (2007) *Global Market Research 2007. Esomar Industry Report*, Amsterdam.
- ESOMAR (2008) *Global Market Research 2008. Esomar Industry Report*, Amsterdam.
- ESOMAR (2008), *Questions To Help Research Buyers Of Online Samples*, Amsterdam.
- ESOMAR (2009) *Global Market Research 2009. Esomar Industry Report*, Amsterdam.
- Gemius (2009). *Polski internet 2008/2009*, Warszawa.
- Główny Urząd Statystyczny (2008). *Wykorzystanie technologii informacyjno-telekomunikacyjnych w przedsiębiorstwach, gospodarstwach domowych i przez osoby prywatne w 2008 r.*, Warszawa.
- Gun H. (2002), *Changing the Survey Process*, Research Paper, dostępne: <http://www.accesswave.ca/~hgunn/special/papers/websurv/>.
- Kellner P. (2004), *Can online Polls produce accurate findings?*, International Journal of Market Research, Vol. 46, Quarter 1.
- Loof A.: Eurostat (2008), *Science and technology. Population and social conditions. Industry, trade and services, Data in focus 46/2008*.
- Manfreda K. L., Batagelj Z., Vehovar V. (2002), *Design of Web Survey Questionnaires: Three Basic Experiments*, Journal of Computer Mediated Communication, 7, 3.
- Mącik R. (2005), *Wykorzystanie internetu w badaniach marketingowych*, Wydawnictwo UMCS, Lublin 2005.
- Milewski M., Ruszczak-Żbikowska J. (2008); *Motywacje do wyjazdu, praca, więzi społeczne i plany na przyszłość polskich migrantów przebywających w Wielkiej Brytanii i Irlandii*, CMR Working Papers, Nr 35/(93) Warszawa.

- Milewski M., Wojciechowski J. (2009) *Na tropach „pozerów”, „klikaczy” i „psotników”. Techniki identyfikacji nietrzetelnych respondentów w badaniach internetowych*. Prezentacja na X Kongresie Badaczy Rynku i Opinii, Warszawa, 15–16 października 2009.
- Przewłocka J., Janczewski P. (2007). *Jak e-Kowalski uczył nas zadawać pytania*. Referat wygłoszony na VIII Kongresie Badaczy Rynku i Opinii, Warszawa.
- Przewłocka J., Półtorak M.(2008), *Towards Unbiased Sampling in Online Survey Research. Use of Information Stored in Cookies as a Method of Improving Sample Structure Quality*, plakat na konferencji General On-line Research, Hamburg.
- PTBRiO (2008), *Katalog PTBRiO 2008. Rynek badań. Badacze. Firmy Badawcze. XVIII edycja*.
- Roster C., Rogers R, Album G., Klein D., *A comparison of response characteristics from Web and telephone surveys*, International Journal of Market Research, Vol. 46, Quarter 3.
- Vehovar V., Manfreda K. L., Batagelj Z. (2001), *Errors In Web surveys*, 53 sesja International Statistical Institute, Seul.
- Zajac J. M., Batorski D. (2007). *Jak skłonić do udziału w badaniach internetowych: Zwiększanie realizacji próby*. „Psychologia Społeczna”, nr 3.

Zastosowanie techniki CAWI w badaniach ewaluacyjnych

Wprowadzenie

Badania marketingowe i społeczne, realizowane w oparciu o technikę CAWI (*Computer Assisted Web Interviewing*), jak pokazują analizy rynku badań w Polsce¹, ostatnimi czasy systematycznie zyskują na znaczeniu. Wzrost zapotrzebowania na badania internetowe bez wątpienia idzie w parze z postępującym rozwojem społeczeństwa informacyjnego w Polsce (w tym z dynamicznym rozwijaniem technologii informacyjno-komunikacyjnych, zwiększaniem dostępności szerokopasmowego Internetu, wykorzystywaniem sieci w codziennej pracy i sprawach prywatnych przez coraz szersze grupy społeczne). Faktem jest, że aktywność ludzka powszechnie wkracza i rozwija się w wirtualnej rzeczywistości, a różnorakie przejawy tej aktywności na polu działalności społecznej, gospodarczej, politycznej i kulturowej to nieprzebrane morze pytań badawczych, które coraz częściej mogą być skutecznie zadawane tylko w e-wymiarze.

Odnosnie do wykorzystania badań internetowych w ewaluacji, mimo ich stosunkowo krótkiej historii w badaniach społecznych, a w Polsce także młodej dziedziny badań, jaką jest ewaluacja, zdążyło pojawić się już sporo uprzedzeń. Po kilku pierwszych, niezbyt udanych próbach zamawiających z wynikami ewaluacji, podpartymi badaniami CAWI, podmioty zlecające ewaluacje szczególnie ostrożnie podchodzą do badań tego typu, oferowanych przez wykonawców w propozycjach technicznych (metodologicznych) przeprowadzenia ocen. Technika CAWI jednak w wielu przypadkach może zapewnić najbardziej efektywny sposób pomiaru efektów programu, zebrania ocen i opinii beneficjentów czy nawet spełnić szeroko rozumiane funkcje sprawozdawcze.

Warto już na wstępie zwrócić uwagę, że CAWI, jak każda inna technika badań surveyowych, ma swoje mocne i słabe strony. Choć przedmiotem niniejszego artykułu jest możliwość jej zastosowania w ewaluacji, należy podkreślić, że nie można rekomendować jej stosowania we wszystkich rodzajach badań tego typu. W badaniach ewaluacyjnych, gdzie kryterium trafności zaadresowania programu wsparcia ma tak istotne znaczenie, należy podkreślić, że wybór techniki w ramach projektowanego badania ewaluacyjnego powinien być również w pełni relewantny², tj. wynikać z uprzednio odpowiednio przeprowadzonego rozpoznania badanego obszaru i profilu respondenta. CAWI warto stosować w tych ewaluacjach, gdzie

¹ Por. Katalog PTBriO 2008. Rynek badań. Badacze. Firmy Badawcze. XIII edycja (http://www.ptbrio.pl/files/Katalog_PTBRiO_2008_na_strone.pdf z dnia 11 października 2009 r.). Poza danymi dostarczonymi przez raporty roczne Polskiego Towarzystwa Badaczy Rynku i Opinii (PTBriO) w zakresie wielkości rynku badań internetowych (oraz badań samego Internetu), znaczenie to potwierdza niedawno dodany rozdział poświęcony badaniom CAWI w „Polskich Standardach Jakości Realizacji Badań Rynku i Opinii Społecznej w Terenie” opublikowanych przez Organizację Firm Badania Opinii i Rynku oraz pierwsze certyfikaty PKJPA uzyskane w 2008 r. przez instytuty badawcze w zakresie badań CAWI (w ubiegłym roku certyfikaty PKJPA w zakresie badań CAWI uzyskało 8 firm badawczych – por. http://www.ofbor.pl/public/File/Wyniki_VII_edycji_audytu_PKJPA.pdf).

² Por. ang. *relevance* – odpowiedniość, trafność, adekwatność. Słowa, które weszły do kanonu najważniejszych pojęć w ewaluacji. Kryterium adekwatności jest uznawane za klasyczne kryterium ewaluacyjne, razem z kryterium skuteczności, efektywności, użyteczności i trwałości.

można wykorzystać mocne strony techniki, a oddziaływanie tych słabszych zminimalizować lub całkowicie pozbawić znaczenia. Istotnymi przy wyborze pozostają przede wszystkim typy ewaluacji, rodzaje programów objętych oceną, profil respondentów, ale również szereg innych zmiennych. W niniejszym artykule zidentyfikowane zostaną te obszary ewaluacji, gdzie naprawdę warto wybrać CAWI, biorąc pod uwagę korzystną relację uzyskiwanych efektów do nakładów. Przestrzeganie tylko kilku istotnych zaleceń dotyczących stosowania CAWI w badaniach ewaluacyjnych, może zaprocentować wysoką jakością zrealizowanych badań, przy niskich, nienadwyżających budżetu, kosztach projektu.

Badania internetowe w ewaluacji

Obecnie, w badaniach ewaluacyjnych na polskim rynku, można wyróżnić dwa główne rodzaje stosowanych technik internetowych (CAWI). Pierwsza z nich – najbardziej powszechna – to badania CAWI z wykorzystaniem baz danych (zlecniodawcy lub zewnętrznych), druga z kolei to „ankiety wyświetlane w zasobach Internetu (pop-ups)”. Marginalne znaczenie dla ewaluacji jeszcze w chwili obecnej ma technika internetowych badań panelowych³.

Zaczynając od końca, być może stosunkowo trudno sobie wyobrazić utrzymywanie aktywnego panelu – w szczególności w skład którego wchodziłoby wybrani beneficjenci jakiegoś programu – którego rolą byłoby wypowiadanie się w zadanej formie (np. poprzez ankietę *on-line*), na określony temat dotyczący ewaluowanego programu. Ze względu na to, że aktualnie większość programów, zwłaszcza tych unijnych, jest realizowana w ramach zamkniętych okresów programowania (kiluletnich perspektyw finansowych), kolejne procesy ewaluacji (przede wszystkim badań tematycznych) z perspektywy beneficjenta mają charakter punktowy – udział w wywiadzie, wypełnienie ankiety za każdym razem na inny temat, mimo że w zasadzie dotyczący tego samego projektu. Inaczej sytuacja wygląda w przypadku ewaluacji bieżących (*on-going*), gdzie teoretycznie beneficjent może być proszony o wypowiedź na tematy tego samego rodzaju (np. procesy, efekty) na różnym etapie realizacji projektu (np. poprzez ankiety CAWI, internetowe formularze sprawozdawcze albo w formie bardziej interaktywnej – *on-line focus group / in-depth interview*, dyskusje panelowe poprzez różnego typu komunikatory, fora czy specjalnie zaprojektowane portale tematyczne). Można więc w pewnym sensie potraktować go jako panelistę, który w okresie realizacji i trwałości projektu, dobrowolnie zgadza się, przystępując do programu, na udział w tego typu badaniach (poświęcając swój czas i dzieląc się wiedzą w dziedzinie projektu).

Jednak taki układ w rzeczywistości zazwyczaj przybiera inną relację. Beneficjent dowiaduje się, że oczekuje się od niego udziału w wywiadzie lub wypełnienia ankiety, w momencie gdy otrzymuje list zapowiadni (drogą elektroniczną lub pocztową), dotyczący konkretnego badania ewaluacyjnego. Proces rekrutacji do takiego „panelu” przybiera zupełnie inną formę niż w zwykłych badaniach rynkowych. Beneficjent podpisując umowę o dofinansowanie, staje się potencjalnym respondentem (informatorem w zakresie swojego projektu, szerszego programu czy elementów całego systemu udzielania wsparcia). Niejednokrotnie jednak rekrutacja ta odbywa się nie w pełni świadomie z punktu widzenia respondenta – jest to jedna z wielu konsekwencji udziału w programie publicznym.

Wykorzystanie internetowych badań panelowych jest bardziej popularne w przypadku, gdy badanie ewaluacyjne zakłada metodę oceny w formie panelu ekspertów, zwłaszcza realizowaną specyficzną metodą delficką. W dużym uproszczeniu, „metoda *Delphi* opiera się na ustrukturyzowanym procesie zbierania wiedzy grupy ekspertów za pomocą serii kwestionariuszy połączonych z kontrolnym zbieraniem

³ Por. Klasyfikacja badań CAWI, zawarta w „Polskich Standardach Jakości Realizacji Badań Rynku i Opinii Społecznej w Terenie” (www.ofbor.pl z dnia 14 października 2009 r.).

opinii zwrotnych. Kwestionariusze są przedstawiane w formie anonimowej i wielokrotnie powtarzanej procedury konsultacyjnej w trakcie badania (pocztą lub e-mailem). Technika (...) zbudowana jest (...) z serii kwestionariuszy wysyłanych do wybranej grupy ekspertów. Kwestionariusze są zaprojektowane tak, aby uzyskać indywidualne odpowiedzi na zadane tematy, a następnie umożliwić ekspertom redefiniowanie ich poglądów w miarę rozwoju pracy grupy"⁴. W przypadku gdy zakładamy udział ekspertów z różnych rejonów geograficznych i językowych (np. w ponadnarodowych badaniach kwestii horyzontalnych, badaniach sektorowych czy przy ewaluacjach programów współpracy transgranicznej), technika CAWI doskonale się sprawdza. Potrafi zapewnić stosunkowo krótki czas realizacji takiego badania (*field time*) i istotnie obniżyć jego koszty (w przeciwieństwie, gdy trzeba byłoby wszystkich członków panelu umówić na spotkanie w jednym wybranym miejscu i czasie). Niejednokrotnie forma internetowa w ogóle przesądza o możliwości udziału w panelu zagranicznych ekspertów wąskich specjalności. Często podkreślana w różnych opracowaniach dotyczących badań internetowych, ich mocna strona, jaką jest „niezależność geograficzna”, także w badaniach ewaluacyjnych stwarza duże możliwości.

Drugi rodzaj badań CAWI – popularne ankiety *pop-up*, realizowane w wyskakujących okienkach, należą do stosunkowo dobrze rozpowszechnionych technik także w badaniach ewaluacyjnych. Badania w formie *pop-up*, mogą być wykorzystywane na przykład w ewaluacjach projektów polegających na budowie portalu tematycznego (np. Portal Funduszy Europejskich – www.funduszeuropejskie.gov.pl; Portal Wspieramy e-biznes – www.web.gov.pl; Portal Innowacji – www.pi.gov.pl), ocenie systemów promocji określonych programów publicznych, realizowanych m.in. poprzez strony WWW (np. www.poig.gov.pl) lub fora internetowe adresowane do osób/podmiotów potencjalnie zainteresowanych programem. Krótkie ankiety ewaluacyjne w formie *pop-up*, mogą obejmować losowe osoby, aktualnie korzystające z informacji zamieszczonych na stronie internetowej, jak i węższe grupy użytkowników (odwiedzających określony dział /zakładkę; ściągających wybrany zakres opublikowanych materiałów; wyrażających opinie). Analiza nawet kilku prostych cech społeczno-demograficznych użytkowników portalu pozwala ocenić skuteczność komunikatów i jakość zamieszczanych informacji dla określonej grupy docelowej.

Stosowanie badań CAWI tego rodzaju bardzo często wiąże się z tym, że nie istnieją bazy danych zakładanych respondentów albo skorzystanie z baz istniejących nie wyczerpuje potrzeb informacyjnych w danym zakresie. Jednak ankiety *pop-up* (podobnie jak inne techniki internetowe) powinny stanowić jedynie element ewaluacji, dodatkowe źródło wiedzy w danym obszarze w ramach całego procesu. Nie można przecież dokonać rzetelnej, przeprowadzonej „zgodnie ze sztuką” oceny podejmowanego problemu badawczego, korzystając tylko z jednego źródła informacji (zwłaszcza tak trudno weryfikowalnego pod względem doboru respondentów, jak *pop-up survey*). Warto mieć na uwadze, że internetową technikę tego rodzaju cechuje stosunkowo dużo wadliwości, które istotnie utrudniają prowadzenie badań. Tak jak większość użytkowników Internetu nie lubi „samoczynnie” wyskakujących okienek, zakłócających bieżące skupienie czy poszukiwanie ważnej informacji (dlatego niektórzy użytkownicy w swoich przeglądarkach wyłączają funkcję *pop-up*), tak też powszechnie „nie lubiane” są wszelkiego rodzaju wyskakujące ankiety. Stąd w przypadku takich badań należy się spodziewać dużej skali w ogóle nierozpoczętych ankiet oraz porzuconych w trakcie wypełniania. Dlatego tak bardzo duże znaczenie w zwiększaniu wskaźników *response rate* mają tu wszelkie techniki związane z zachętą do rozpoczęcia udziału w badaniu i podnoszeniem motywacji do jego ukończenia przez respondenta (podkreślanie społecznej użyteczności wyników ewaluacji, roli, jaką ma do spełnienia respondent, zapewnienie dodatkowych form gratyfikacji za udział itp.).

⁴ „Ewaluacja Narodowego Planu Rozwoju i programów operacyjnych w Polsce. Poradnik”, Krajowa Jednostka Oceny, Departament Koordynacji Polityki Strukturalnej, MGiP, Warszawa 2005.

Siłą rzeczy, ankiety ewaluacyjne w wyskakujących okienkach, nie mogą być długie. Ponadto, podobnie jak przy innych badaniach typu CAWI, obowiązuje bardzo duża dyscyplina w zakresie funkcjonalności aplikacji, atrakcyjności wizualnej, sprawności na różnego rodzaju przeglądarkach, systemach operacyjnych czy wolniejszych łączach. Na pewno duże znaczenie przy zachęcaniu do udziału ma autorytet badacza lub patrona badania (renomowana instytucja, której logo widoczne jest na stronie badania, sponsor programu, poważana osoba albo instytut badawczy). Niemniej jednak, biorąc pod uwagę wielość programów, funduszy czy projektów, które są w Polsce aktualnie realizowane (w dużej mierze za sprawą Unii Europejskiej), trudno znaleźć osobę lub organizację, która w związku z udziałem w dofinansowanych szkoleniach czy w różnego rodzaju programach wsparcia nie spotkała się z ankietami (w tym internetowymi) – pod kątem rekrutacji do programu czy oceny po programie. Dlatego też coraz trudniej w Polsce realizuje się badania ankietowe przy okazji ewaluacji, a z całą pewnością trudność ta będzie dotyczyć badań CAWI realizowanych poprzez ankiety *pop-up*. Choć można się też zastanowić czy w przypadku skomplikowanych narzędzi CAWI, stosowanych w celach ewaluacyjnych, sprawozdawczych albo ewidencyjnych, z którymi muszą radzić sobie respondenci (wnioskodawcy, beneficjenci, potencjalni adresaci interwencji), nie warto byłoby dodatkowo wykorzystać okienek *pop-up*, aby zweryfikować w oparciu o opinie grupy docelowej, jakości przygotowanych narzędzi zbierania danych.

Trzecia, najważniejsza z punktu widzenia ewaluacji, technika badań internetowych – CAWI z wykorzystaniem baz danych respondentów – od kilku lat z mniejszym lub większym powodzeniem jest stosowana przez zewnętrznych wykonawców ewaluacji (instytuty badawcze), jak również przez komórki ewaluacyjne powołane w strukturach publicznych (rządowych, samorządowych, pozarządowych), odpowiedzialnych za poszczególne programy. Niemniej jednak należy podkreślić, że jej popularność w ewaluacji wzrasta coraz bardziej (podobnie jak w całym sektorze badań opinii i rynku). Z jednej strony ze względu na dostępność CAWI w ofercie coraz większej liczby agencji badawczych, z uwagi na niskie koszty i krótki czas procesu zbierania danych, ale chyba przede wszystkim dlatego, że w wielu przypadkach jest to najbardziej efektywna forma prowadzenia badań ankietowych wśród wnioskodawców i beneficjentów programów – osób fizycznych, a zwłaszcza podmiotów instytucjonalnych.

Możliwość korzystania z techniki CAWI (realizowanej z wykorzystaniem baz danych) w ewaluacji, w największym stopniu determinuje profil beneficjentów (lub respondentów innego rodzaju), którzy dysponują wiedzą w zakresie programu poddanego ocenie i mogą się nią podzielić. Programy wsparcia sektorów rolnictwa i rybołówstwa czy różnego typu programy przeciwdziałania marginalizacji (ekskluzji) społecznej oraz, z podobnych względów, wiele innych programów, z całą pewnością nie będą odpowiednią materią badawczą, przy której będzie można skutecznie zastosować jakiejkolwiek techniki internetowe. Można oczywiście wskazać specyficzne wyjątki, ale kluczową kwestią pozostaje penetracja Internetem wśród naszych jednostek obserwacji. Być może w niektórych przypadkach skuteczne mogą się okazać sposoby mieszane w dotarciu i zebraniu danych od respondentów (np. przekazane pocztą tradycyjną, zaproszenie respondenta do wypełnienia ankiety, w zależności od jego preferencji – 1) bezpośrednio na stronie internetowej pod wskazanym adresem WWW lub – 2) w formie papierowej, z prośbą o odesłanie pocztą do instytutu). Choć stosując tzw. *mixed-mode survey data collection* trzeba mieć na uwadze, że badania prowadzone w tej formie mogą „dostarczać” różnego rodzaju dodatkowych błędów, które w miarę możliwości należy minimalizować⁵. Niektórzy badawcze w ogóle nie akceptują takiego, podejścia, chociaż z całą pewnością w wielu przypadkach jego zastosowanie mogłoby być funkcjonalne

⁵ Por. H. Domański, K. Lutyńska, A. Rostocki (red.), *Spojrzenie na metodę. Studia z metodologii badań socjologicznych*, IFiS PAN, Warszawa 1999.

(np. przyczyniając się ograniczenia liczby braków danych w ankiecie ewaluacyjnej czy podnosząc poziom zwrotów).

Drugą, istotną do rozważania kwestią, przy stosowaniu tego rodzaju techniki CAWI w ewaluacji, pozostaje baza danych, która będzie służyła jako operat losowania. Na etapie procesu zbierania danych, respondent nie jest bowiem anonimowy. Są kierowane do niego indywidualnie zaadresowane zaproszenia (opatrzone takimi danymi jak imię i nazwisko lub nazwa, adres kontaktowy, numer projektu itp.), a jednocześnie w samym liście przewodnim lub ankiecie badany jest zapewniany o przyszłej „anonimizacji” wyników i analizach zbiorczych, bez możliwości identyfikacji informatora. Do procedur doboru próby, wysyłki zaproszeń, kontroli realizacji czy późniejszych analiz przy uwzględnieniu istotnych przekrojów (np. ze względu na wielkość dofinansowania, lokalizację czy typ projektu) potrzebna jest zasobna baza danych. W przypadku programów publicznych (zwłaszcza programów unijnych, począwszy od tych przedakcesyjnych), najczęściej bazy zawierają szereg zmiennych określających profil wnioskodawcy / beneficjenta oraz charakterystykę jego projektu(ów). Dzięki funkcjonującym od pewnego czasu generatorom wniosków dane te są zbierane od razu w formie elektronicznej⁶. Przy czym, w przypadku badań CAWI szczególnie istotną zmienną pozostaje adres *e-mail* beneficjenta. Gdy dysponujemy adresem *e-mail*, zaproszenia z linkiem do ankiety, listy monitorujące, kontakt z działem realizacji czy podziękowania za udział w badaniu, odbywać się mogą pocztą elektroniczną, a koszty badania dzięki temu automatycznie spadają kilkakrotnie⁷. Warto jednak podkreślić, że adresy *e-mail*, zwłaszcza osób prywatnych, ulegają dezaktualizacji dużo szybciej niż pozostałe dane teleadresowe. Biorąc pod uwagę, że badania ewaluacyjne dotyczące efektów wsparcia, są realizowane najczęściej, co najmniej po 6 miesiącach od zakończenia realizacji projektu (udziału w projekcie), a w przypadku dużych projektów inwestycyjnych po 24 miesiącach lub jeszcze później, dane dotyczące adresów mailowych (podawanych na etapie aplikacji we wniosku o dofinansowanie), mogą być nieaktualne nawet w przypadku kilkunastu procent beneficjentów. W trakcie realizacji projektu i w okresie jego trwałości, beneficjenci (ale już niestety nie pozostali wnioskodawcy, będący także cennym źródłem „wiedzy ewaluacyjnej”) są zobowiązani do informowania instytucji wdrażającej program, o zmianach wszelkich danych identyfikacyjnych. Jednak w przypadku adresów pocztowych, nazw organizacji czy danych przedstawicieli prawnych podmiotu, takie zmiany są odnotowywane, natomiast nie zawsze rejestrowane są zmiany adresów *e-mail* (a tym bardziej w elektronicznej bazie danych). Dostępność „pewnych” adresów *e-mail* respondentów stanowić może zatem istotny problem w badaniach ewaluacyjnych, prowadzonych w oparciu o technikę CAWI. Może on dotyczyć przede wszystkim ewaluacji programów skierowanych do osób fizycznych oraz małych przedsiębiorstw i organizacji, rzadziej występuje w programach, gdzie bezpośrednimi beneficjentami są instytucje, które posiadają dłuższą historię i skalę działalności (co w ich przypadku, w praktyce oznacza również względną niezmienną adresów poczty elektronicznej).

Ponadto, z bazami danych (zarówno z wtórnymi, jak i „wyprodukowanymi” w trakcie realizacji badania internetowego) wiążą się też ściśle sformalizowane procedury rejestracji i ochrony ich zawartości (w tym zapewnienia bezpieczeństwa przechowywania na serwerach i przetwarzania), w szczególności, gdy w grę wchodzi dane osobowe lub inne dane prawnie chronione. Procedury te nie różnią się jednak istotnie od tych stosowanych w przypadku pozostałych technik badawczych. Dodatkową rzeczą, która sprawdza się w badaniach CAWI jako zabezpieczenie przesyłu danych i jednocześnie swoista gwarancja

⁶ Co, jak pokazały doświadczenia starszych programów, eliminuje wiele błędów, wypływających niestety najczęściej dopiero na etapie oceny efektów projektów, tj. przy ewaluacji *ex post*.

⁷ Choć w przypadku programów perspektywy 2007–2013, skierowanych m.in. do przedsiębiorców, w trakcie konsultacji wzorów wniosku o dofinansowanie, w niektórych działaniach była rozważana propozycja, żeby zmienna „adres poczty elektronicznej” była podawana przez wnioskodawców fakultatywnie (co uniemożliwiłoby dotarcie do wielu beneficjentów poprzez *e-mail*).

dla respondentów, to protokoły połączenia szyfrowanego. Certyfikat SSL (ang. *Secure Socket Layer*) wystawiony dla domeny, na której realizowane jest badanie, gwarantuje, że przepływ informacji (transfer danych) następuje tylko i wyłącznie między dwoma punktami – komputerem respondenta i serwerem z bazą badania. Mimo wciąż powszechnie występujących w Polsce obaw internautów co do podawania jakichkolwiek ważnych (osobistych, firmowych, służbowych itp.) informacji w Internecie, dzięki komunikatom o bezpiecznym połączeniu, adresie strony rozpoczynającym się od „https” i widocznej na ekranie „kłódeczce”, respondenci w badaniach CAWI (zwłaszcza gdy są posiadaczami rachunku bankowego z dostępem *on-line*, korzystają ze sklepów lub aukcji internetowych) powinni poczuć się zdecydowanie pewniej. Jeśli zakładamy zbieranie, w oparciu o badanie CAWI, danych „wrażliwych” z punktu widzenia respondentów, z pewnością połączenie szyfrowane ze stroną badania, powinno zmniejszyć odsetek braków danych (*item-non response*).

Przyjmując minimum wymagań w odniesieniu do beneficjentów programu (aktywni użytkownicy Internetu, którzy w danych teleadresowych na potrzeby programu udostępnili swój adres *e-mail*), warto się dalej zastanowić, w jakiego rodzaju badaniach ewaluacyjnych można do nich dotrzeć poprzez sieć. Wydaje się, że trafne zastosowania techniki CAWI z wykorzystaniem baz danych programu, można wskazać w bardzo szerokim obszarze różnego typu badań ewaluacyjnych – związanych ze zbieraniem informacji w zakresie podejmowanego problemu badawczego (w tym ocen, opinii czy innego rodzaju danych o charakterze ilościowym) od beneficjentów, wnioskodawców oraz innych podmiotów zaangażowanych lub zainteresowanych⁸ programem. W szczególności będą to różnego rodzaju śródkresowe ewaluacje tematyczne, dotyczące trafności i użyteczności zaprojektowanej oferty wsparcia dla grup docelowych czy oceny funkcjonalności systemu udzielania pomocy na różnych etapach tego procesu. Podobnie, z powodzeniem można wykorzystać CAWI w badaniach ewaluacyjnych realizowanych *ad-hoc*, w odniesieniu do „palących” potrzeb informacyjnych służb publicznych w zakresie danego programu, tj. wynikających z dostrzeganego problemu proceduralnego, nagłej zmiany sytuacji społeczno-gospodarczej czy planowanej do wdrożenia kontrowersyjnej decyzji politycznej. W tego typu przypadkach szybkie i tanie badania internetowe doskonale się sprawdzają. Doświadczone komórki ewaluacyjne są w stanie samodzielnie takie badania zaprojektować i przeprowadzić (np. wśród uczestników programu), zostawiając z boku czasochłonne procedury zamówień publicznych. Biorąc pod uwagę możliwości, jakie stwarzają półprofesjonalne oprogramowania (bardzo często dostępne w formie *open source*) do ankiet *on-line* czy bardziej zaawansowane komercyjne platformy badań internetowych, nie dziwi fakt, że takie rozwiązania zaczynają być coraz chętniej wykorzystywane przez jednostki publiczne (samorządy, agendy rządowe) w prowadzonych badaniach (także o charakterze ewaluacyjnym).

Doskonale sprawdzają się także badania CAWI w przypadku zorientowanych na pomiar i ocenę efektów, ewaluacji *on-going* i *ex-post*. Bieżący pomiar i ocena osiągniętych wskaźników rezultatu czy oddziaływania, w ramach realizowanych instrumentów wsparcia, stanowi nieocenione narzędzie do obserwacji (monitorowania) zarówno postępów, jak i weryfikacji (ewaluacji) ich skuteczności i efektywności w zakładanym zakresie. W przypadku badania efektów, najczęściej horyzont pomiaru wskaźnika ewaluacyjnego (rezultatu czy bezpośredniego wpływu) powinien być sprzężony z momentem zakończenia projektu przez beneficjenta i zamknięciem odpowiedniego dla tego wskaźnika okresu sprawozdawczego. Techniki badań wspomaganych komputerowo i „internetowo” są w stanie w odpowiednim momencie wyselekcjonować respondenta, spełniającego zakładane warunki udziału w badaniu (kryteria doboru) i zaprosić *via e-mail* do wypełnienia ankiety (np. po 12 miesiącach od zakończenia udziału w programie).

⁸ W przypadku niektórych badań ewaluacyjnych, duże znaczenie będą miały bazy *newsletter*, zawierające dane potencjalnych zainteresowanych ofertą programu (w tym m.in. ich adresy poczty elektronicznej).

Zagregowane wyniki zebrane w taki sposób stanowią dane dużo lepszej jakości niż wyniki uzyskane w ramach punktowego badania *ex-post*, które w jednym momencie jest adresowane do wszystkich beneficjentów, niezależnie od czasu, jaki upłynął od ich udziału w projekcie (zakończenia przedsięwzięcia). Koordynacja tradycyjnych badań terenowych, realizowanych z udziałem ankietera (np. techniki PAPI, CAPI), przy kontrolowaniu odpowiednich dat dotyczących projektu i optymalnych momentów pomiaru wskaźnika, kilkakrotnie przewyższyłaby koszty ewentualnego zastosowania w tym wypadku techniki CAWI. W sytuacji, gdy ewaluacje bieżące (*on-going*) nie są realizowane wcale lub w odpowiedni sposób w trakcie wdrażania programu, przy ewaluacji *ex-post* siłą rzeczy próbuje się to nadrobić, co niestety odbija się na kosztach i jakości zebranego materiału empirycznego. Na tym etapie problemy w uzyskaniu odpowiedniego poziomu zwrotów także zwiększają się istotnie. Beneficjenci, którzy przeszli kilkakrotne kontrole i monitoringi czy brali udział w różnego typu śródkresowych ewaluacjach programu, w coraz mniejszym stopniu godzą się na współpracę z ewaluatorem przy ocenie *ex-post*.

Generalnie jednak w ewaluacjach *ex-post*, z dużym powodzeniem można także stosować technikę CAWI, mając jednak na uwadze mniejszą motywację respondentów, większy stopień nieaktualnych adresów *e-mail* (niedostępność wartościowych informatorów) czy zawodzącą każdego pamięć, po upływie dłuższego czasu od interesującego nas zdarzenia. Również, mimo że technika CAWI daje możliwość przeprowadzenia badania wyczerpującego (na całej populacji), przy nie dużo większych kosztach i czasie realizacji, warto raczej stosować reprezentacyjne badania na próbach. Szkoda bowiem „wyczerpywać” operat, który może stanowić dużą wartość w przyszłych badaniach o charakterze *ex-ante*⁹. Doświadczenie dotychczasowych beneficjentów może być niejednokrotnie bardzo cenne przy projektowaniu nowej generacji programów wsparcia adresowanych do analogicznej grupy docelowej.

Na etapie ewaluacji *ex post* bardzo często próbuje się także dotrzeć do ostatecznych odbiorców pomocy (beneficjentów ostatecznych). W przypadku wielu instrumentów wsparcia, bezpośrednimi grantobiorcami są beneficjenci instytucjonalni (np. instytucje otoczenia biznesu, samorządy, firmy szkoleniowe, organizacje pożytku publicznego, itp.), a ostatecznym odbiorcą pomocy transferowanej przez te podmioty jest ktoś inny. Mogą być to zarówno odbiorcy tzw. pomocy *de minimis* (np. przedsiębiorcy-klienci parków naukowo-technologicznych czy członkowie powiązań kooperacyjnych), osoby fizyczne (np. uczestnicy dofinansowanych szkoleń), ale również pośredni użytkownicy lub odbiorcy wytworzonych dóbr i usług (np. społeczność samorządowa, korzystająca z wytworzonej infrastruktury społeczno-gospodarczej). Są to najczęściej bardzo liczne populacje, w miarę odpowiednich potrzeb informacyjnych, doskonale do badania poprzez techniki ilościowe. Podobnie istnieją w tym przypadku stosunkowo duże możliwości wykorzystania ankietowania w formie CAWI. Wszystko zależy od profilu tych respondentów i dostępności odpowiedniej bazy danych do mailingu. Niestety na poziomie ostatecznych odbiorców pomocy, bardzo często ewaluatorzy spotykają się z niedostępnością danych w formie elektronicznej. Redukcje na tym poziomie elektronicznych zasobów informacyjnych, dotyczących programu, niekiedy wynikają z negatywnych doświadczeń instytucji z centralnymi systemami informatycznymi, „niekontrolowaną decentralizacją” systemów bazodanowych czy obawą przed przetwarzaniem danych osobowych. Niejednokrotnie w takich przypadkach eliminuje to możliwość wykorzystania w ewaluacji badań CAWI lub ze względu na koszty w ogóle wyklucza objęcie badaniami ewaluacyjnymi grup docelowych (ostatecznych odbiorców wsparcia).

⁹ Por. Ewaluacja pn. „Plany i potrzeby przedsiębiorców sektora MSP w zakresie ochrony własności przemysłowej, w kontekście uruchomienia działania 5.4 PO IG Zarządzanie własnością intelektualną”, IBC GROUP Central Europe Holding Ltd. na zlecenie PARP, Warszawa 2008 r. oraz ewaluacja pn. „Plany i potrzeby przedsiębiorców sektora MSP w zakresie rozwoju eksportu i powiązań z zagranicznymi rynkami i partnerami handlowymi, w kontekście uruchomienia działania 6.1 PO IG Paszport do eksportu”, Policy & Action Group Uniconsult Sp. z o.o. na zlecenie PARP, Warszawa 2009 r.

Specyficzna relacja badacz – respondent – beneficjent

Stosunkowo liczne już obecnie prace teoretyczne z zakresu stosowania badań internetowych – zarówno rodzimych, jak i obcojęzycznych autorów – niejednokrotnie podkreślają, że CAWI (z wykorzystaniem baz danych respondentów) jest bardzo skutecznym narzędziem badawczym wszędzie tam, gdzie grupa docelowa ma charakter zamknięty i jednoznacznie określony. Ponadto CAWI sprawdza się zwłaszcza, gdy członkowie grupy mają świadomość przynależności do tej grupy (np. beneficjenci – w wielu sytuacjach informowani, że biorą udział w programie współfinansowanym ze środków Unii Europejskiej). Z całą pewnością, m.in. te charakterystyki beneficjentów czy także osób nieskutecznie aplikujących o uczestnictwo w programie oraz cechy przywoływane wcześniej, uprawniają do sięgania po internetowe techniki badawcze w ramach ewaluacji programów.

Doświadczenia z badań *face to face* potwierdzają, że kontakt ankietera z beneficjentem w sprawie jego uczestnictwa w programie nie jest dla tego drugiego czymś zaskakującym. Beneficjenci, jeszcze wiele miesięcy przed tym, kiedy dochodzi do spotkania w ramach ewaluacji, sami deklarowali zgodę na udział w badaniach tego typu (czyli udzielania wywiadów, wypełniania ankiet ewaluacyjnych itp.). Ich nastawienie do prowadzonych badań ewaluacyjnych jest więc zupełnie inne niż nastawienie innych respondentów badań społecznych (nie-beneficjentów). Można powiedzieć, że beneficjenci spodziewają się, że przyjdzie taki moment, gdy ktoś (np. ankieter reprezentujący instytut badań lub przedstawiciel instytucji odpowiedzialnej za program) zapyta ich o efekty programu. Mimo że w przypadku CAWI nie występuje osoba ankietera, podobnie zaproszenia do udziału w badaniach związanych z ewaluacją programu, są traktowane przez respondentów z perspektywy diametralnie odmiennej motywacji, niż w przypadku zaproszeń dotyczących udziału w innych badaniach społecznych czy rynkowych. Warto zauważyć, że w zapisach umów o dofinansowanie projektu, zawieranych pomiędzy beneficjentem a instytucją wdrażającą, prawie zawsze jest mowa również o udzielaniu informacji na potrzeby ewaluacji programu. W umowach stosowanych m. in. w Sektorowym Programie Operacyjnym Wzrost Konkurencyjności Przedsiębiorstw, 2004–2006, w warunkach ogólnych można znaleźć następujące sformułowania „Ostateczny Odbiorca (beneficjent) zobowiązuje się (...) udzielać Instytucji Zarządzającej, Beneficjentowi Końcowemu (instytucji wdrażającej) lub wskazanym przez nie podmiotom informacji na temat efektów Projektu przez okres 5 lat od dnia zakończenia realizacji Projektu”¹⁰. Ten zapis bez wątpienia ułatwia zbieranie danych i realizację badań ankietowych, nawet w oparciu o techniki internetowe, jeśli tylko dla danego beneficjenta jest to adekwatny kanał komunikacji. Nawet niezależnie od zapisów umowy (których z pewnością duża część uczestników programu dokładnie nie pamięta), ale być może właśnie ze względu na odczuwane podświadomie zobowiązanie wobec donatora środków finansowych, generalnie beneficjenci są bardziej skłonni do udziału w badaniach ewaluacyjnych, niż jako „statystyczni Polacy” do udziału w badaniach innego typu¹¹. W przypadku badań ewaluacyjnych można więc mówić o występowaniu specyficznej relacji pomiędzy badaczem a respondentem (beneficjentem), która sprzyja podnoszeniu wskaźników realizacji próby (niezależnie od stosowanej techniki). Przy czym w przypadku istotnych ograniczeń metod internetowych w zakresie poziomu *response rate*, występująca w badaniach ewaluacyjnych korzystna relacja z respondentem, niewątpliwie zwiększa potencjalną skuteczność techniki CAWI i uzasadnia jej zastosowanie właśnie w badaniach realizowanych na potrzeby ewaluacji. Dodatkowo zastosowanie szeregu rozwiązań funkcjonalnych i merytorycznych opisanych w literaturze, które ofertuje technika CAWI,

¹⁰ Por. <http://www.parp.gov.pl/index/index/197> z dnia 11 października 2009 r.

¹¹ W zależności od wielu czynników (np. liczby badań ewaluacyjnych, w których beneficjent uczestniczył, wielkości uzyskanego dofinansowania, czasu, jaki upłynął od zakończenia udziału w programie lub rozliczenia projektu) skłonność ta będzie się zwiększać lub zmniejszać.

pozwala w badaniach ewaluacyjnych osiągnąć, jak na ankietę „samowypełnialną”, wysoce satysfakcjonujące wskaźniki realizacji. Rozwiązania te dotyczą etapu aranżacji „wywiadu” CAWI w listach przewodnich (mailowych i tradycyjnych pocztowych), pracy respondenta z narzędziem (np. jakość kwestionariusza ankiety i funkcjonalność aplikacji CAWI; dostęp osób do kontaktu po stronie działu realizacji, zakładki FAQ, pomoc podręczna, itp.), formuły monitorowania (np. przypomnienia mailowe lub telefoniczne) oraz gratyfikacji za udział w badaniu.

W badaniach ewaluacyjnych realizowanych *on-line* również mniejszy problem stanowią przypadki przerywania wypełniania ankiet w trakcie (*dropout*). W dużej mierze motywacja, która skłoniła respondenta (beneficjenta) do udziału w badaniu, jest na tyle silna, że dyscyplinuje go do wypełnienia ankiety do samego końca (zwłaszcza, gdy narzędzie CAWI zawiera panel logowania i możliwość wielokrotnego zapisywania wprowadzonych danych, tj. umożliwia respondentowi wypełnianie ankiety „na raty”)¹². Oczywiście są pewne wyjątki. Ze względu na to, że badania przez Internet są stosunkowo łatwe do przeprowadzenia i nie generują dużych kosztów, zdarza się, że niektóre projekty CAWI są uruchamiane w formie niedopracowanej (np. niewłaściwy dobór próby¹³; zbieranie danych, które beneficjenci już podawali na etapie wniosku lub sprawozdania, brak logiki ankiety i błędy w regułach przejść itp.). Wtedy mimo szczerych chęci respondentów, istotnie spada poziom realizacji ankiet *on-line* i wzrasta wskaźnik *dropout*.

Ze względu na zmniejszenie wpływu omawianej relacji pomiędzy badaczem a respondentem, niższą skuteczność będą miały badania CAWI na grupie beneficjentów, która nie jest związana tak silną więzią (zobowiązaniem wynikającym z otrzymanego wsparcia czy zapisów umowy) z przedstawicielem sponsora programu. Chodzi tu szczególnie o podmioty aplikujące, które ostatecznie nie uzyskały dofinansowania (np. ich projekty nie uzyskały odpowiedniej liczby punktów na etapie selekcji) albo beneficjentów ostatecznych, korzystających nie w pełni świadomie z „dobrodzieństw programu”. Jedni, ze względu na dostępność danych teleadresowych i charakterystyk społeczno-gospodarczych, mogą doskonale służyć za operat do konstruowania próby kontrolnej w porównawczych metodach oceny. Drudzy z kolei stanowią podstawową bazę informatorów w zakresie efektów na poziomie oddziaływania programu. Jednak zarówno pierwsi, jak i drudzy nie będą tak „chętnie” jak bezpośredni grantobiorcy wypełniać ankiet ewaluacyjnych (tym bardziej udostępnionych *on-line*).

Jednak mimo tej specyficznej relacji występującej w badaniach ewaluacyjnych pomiędzy badaczem lub zlecającym badanie a respondentem, nie przestają obowiązywać „standardy” badań społecznych. Podkreślanie obligatoryjnego charakteru udziału w badaniu dla beneficjentów czy straszenie ich niewypełnieniem zapisów umowy przy ewentualnej odmowie (co teoretycznie oznacza zwrot dotacji), z całą pewnością prowadzić będzie do skrzywienia wyników i całkowicie podważa sens ich realizacji. Tylko dobrowolny udział zagwarantuje zebranie szczerych opinii i rzetelnych ocen. Zapis w umowie o dofinansowanie ma przede wszystkim charakter informacyjny, trudno sobie wyobrazić, żeby w przypadku ewaluacji podlegał egzekucji (co innego natomiast, jeśli chodzi o niepoddanie się beneficjenta kontroli lub monitoringowi). Warto jednak podkreślić, że beneficjenci nierzadko mieszają różne rodzaje badań dotyczących programu, z którego korzystali, nie odróżniając czy służą one celom sprawozdawczym, kon-

¹² Rozwiązanie to sprawdza się w badaniach ewaluacyjnych realizowanych techniką CAWI, gdzie wraz z ankietą wykorzystywane są formularze wskaźnikowe. Najczęściej, aby respondent mógł „wywiązać się” z wypełnionej ankiety, powinien sięgnąć po odpowiednie dane. Ankieta CAWI, adresowana na przykład do przedsiębiorstw (beneficjentów dotacji), wcale nie musi być wypełniana przez jedną osobę w firmie. Nierzadko może wymagać zaangażowania kilku osób (np. prezesa zarządu, pełnomocnika ds. realizacji projektu, głównego księgowego), będących najlepszymi informatorami w danym zakresie projektu.

¹³ Bazy do badania CAWI, zaczerpnięte z centralnych systemów informatycznych dotyczących programów unijnych (SIMIK, KSI), w stosunkowo małym stopniu pozwalają na dokonanie jakiegokolwiek wstępnej analizy na etapie doboru próby, dlatego niekiedy czynność ta bezrefleksyjnie zostaje „przerzucona” na pytania filtrujące w ankiecie CAWI. Wtedy okazuje się, że przypadków, które są specyficzne, jest więcej niż zakładano, a badanie zostało skierowane do bardzo zróżnicowanych grup, z których tylko nieliczne potrafią odnieść się do abstrakcyjnych z ich punktu widzenia (a niekiedy nawet absurdalnych) pytań czy poleceń w ankiecie.

trolnym czy ewaluacyjnym. Z punktu widzenia poziomu realizacji ankiet w badaniach ewaluacyjnych, może być to korzystne, niestety z punktu widzenia jakości zebranego materiału (niezależności opinii, rzetelności wystawionych ocen) może działać to na niekorzyść. Najważniejszym motywatorem dla beneficjentów (respondentów) powinno być poczucie społecznej użyteczności wyników prowadzonych badań, choć nie da się ukryć, że duże znaczenie motywujące mają także inne czynniki.

Reasumując, bardzo często przytaczany zarzut wobec techniki CAWI, dotyczący niskiego poziomu realizacji zakładanej próby, przy spełnieniu minimum wymagań metodologicznych (w tym m.in. w obszarze kryteriów doboru respondentów, konstrukcji narzędzia, organizacji procesu zbierania danych), w dużej mierze nie powinien dotyczyć badań ewaluacyjnych. Oczywiście, w miarę „mnożenia się” ewaluacji różnego typu, które dotyczą nakładających się na siebie programów i obejmujących kilkakrotnie tych samych respondentów, wskaźniki realizacji stopniowo się obniżają. Z całą pewnością jednak w wielu badaniach ewaluacyjnych szereg zdefiniowanych ograniczeń techniki CAWI można z łatwością zredukować do minimum, jednocześnie wykorzystując jej mocne strony.

Doświadczenia PARP z techniką CAWI

Interesujący przykład wykorzystania techniki CAWI w badaniach ewaluacyjnych można znaleźć w Polskiej Agencji Rozwoju Przedsiębiorczości¹⁴. Wprowadzając w kontekst, na początku warto przywołać kilka podstawowych faktów dotyczących genezy projektów ewaluacyjnych w Agencji, zwłaszcza projektów wykorzystujących technikę CAWI.

Od początku działalności PARP prowadziła badania ewaluacyjne dotyczące realizowanych instrumentów wsparcia. Dynamiczny rozwój działań ewaluacyjnych w PARP, w dużej mierze za sprawą powołanej w 2004 r. komórki ewaluacyjnej, zbiegł się z początkiem ewaluacji w polskiej administracji publicznej (w podobnym okresie równolegle powstawały np. Jednostka Ewaluacyjna Narodowego Koordynatora Pomocy dla Programów *Phare* i *Transition Facility* w Urzędzie Komitetu Integracji Europejskiej oraz Krajowa Jednostka Oceny NPR/PWW 2004–2006 w Ministerstwie Gospodarki i Pracy)¹⁵. Przedmiotem ewaluacji prowadzonych przez PARP były programy przedakcesyjne (tu głównie ewaluacje *ex-post*), wybrane działania lub grupy działań programów operacyjnych okresu 2004–2006 (*on-going*, tematyczne uzupełniające, *ex-post*), a obecnie również perspektywy 2007–2013 (tematyczne, *on-going*, *ad-hoc*). Doświadczenia, które udało się zgromadzić Agencji na polu ewaluacji, przejawiają się zarówno w unikalnym podejściu do realizacji tego typu projektów (np. badanie efektów netto z wykorzystaniem techniki *Propensity Score Matching*, systematyczne badanie beneficjentów, projekty grup eksperckich do planowanych ewaluacji *on-going* w POIG, POKL i PORPW), jak i intensywnej działalności związanej z budowaniem potencjału ewaluacyjnego w Polsce (np. od 2005 r. organizacja kolejnych dorocznych Konferencji Ewaluacyjnych,

¹⁴ Polska Agencja Rozwoju Przedsiębiorczości jest agencją rządową podlegającą Ministrowi właściwemu ds. gospodarki, utworzoną na mocy ustawy z 9 listopada 2000 roku. Zadaniem Agencji jest zarządzanie funduszami pochodzącymi z budżetu państwa oraz Unii Europejskiej i realizacja programów rozwoju gospodarki, wspierających: przedsiębiorczość, działalność innowacyjną i badawczą (zwłaszcza małych i średnich przedsiębiorstw), rozwój regionalny, wzrost eksportu, rozwój zasobów ludzkich czy wykorzystanie nowych technologii. W perspektywie finansowej 2007–2013 Agencja jest odpowiedzialna za wdrażanie działań w ramach trzech programów operacyjnych: Innowacyjna Gospodarka (POIG), Kapitał Ludzki (POKL) i Rozwój Polski Wschodniej (PORPW). W przypadku okresu programowania 2004–2006, PARP pełniła funkcję Instytucji Wdrażającej m.in. dla działań realizowanych w ramach programów: Rozwój Zasobów Ludzkich (SPO RZL) i Wzrost Konkurencyjności Przedsiębiorstw (SPO WKP) (por. <http://bip.parp.gov.pl/index/index/505> z dnia 11 października 2009 r.).

¹⁵ Por. J. Pokorski, *System ewaluacji wykorzystania funduszy strukturalnych w Polsce. Ramy instytucjonalne i prawne* (praca dyplomowa), Uniwersytet Kardynała Stefana Wyszyńskiego, Warszawa 2008.

publikacje tematyczne, szkolenia z ewaluacji, bieżąca współpraca z jednostkami ewaluacyjnymi, światem nauki i ewaluatorami)¹⁶.

Większość badań ewaluacyjnych przeprowadzanych przez PARP z wykorzystaniem techniki CAWI miała dotychczas charakter ewaluacji wewnętrznych, tj. badania były prowadzone bez udziału zewnętrznych instytucji badań opinii i rynku, lecz w oparciu o wewnętrzne zasoby ludzkie, organizacyjne, techniczne i merytoryczne Agencji (przy ewentualnym wsparciu zewnętrznych programistów lub analityków). Najczęściej narzędzia badawcze (aplikacje CAWI) w ewaluacjach realizowanych przez PARP były projektowane indywidualnie do potrzeb danego badania (głównie w języku PHP) i uruchamiane pod adresem parp.gov.pl.

Pierwsze badanie CAWI skierowane do beneficjentów, które „zawisło” na serwerze PARP, było projektowane jeszcze w 2004 r. i dotyczyło ewaluacji *ex-post* Programu „Innowacje i technologie dla rozwoju przedsiębiorstw” Phare 2000 (PL0003.07 Krajowy program rozwoju MSP). Było to stosunkowo proste narzędzie badawcze¹⁷ skierowane do przedsiębiorców, realizujących niewielkie, innowacyjne jak na tamte czasy projekty. Od tego czasu udało się w Agencji zrealizować wiele tego typu badań ewaluacyjnych, obejmujących wnioskodawców i beneficjentów wdrażanych przez PARP programów. W badaniach brały udział tysiące przedsiębiorstw – korzystających z dotacji (Phare, SPO WKP) oraz wysyłających pracowników na współfinansowane szkolenia (SPO RZL). Ankiety *on-line* wypełniali także nauczyciele biorący udział w projektach skierowanych dla szkół, uczestnicy konferencji czy przedstawiciele wspieranych przez PARP instytucji otoczenia biznesu.

W podobnym okresie, co pierwsze ewaluacyjne badanie techniką CAWI, PARP uruchomiła badanie oparte o technikę ankiety *e-mail* (wspomagana ankietą pocztową). Było to badanie cykliczne o charakterze *ex-post*, realizowane w ramach corocznych edycji przez okres 5 lat na beneficjentach projektów infrastrukturalnych programu Phare 2000 SSG (w badaniu brały udział m.in. władze samorządowe, jednostki GDDKiA, organizacje pozarządowe)¹⁸. W innych ewaluacjach programów przedakcesyjnych PARP także wykorzystywała techniki internetowe – zarówno w ramach własnych projektów badawczych, jak również w projektach zleconych (np. zewnętrzna ewaluacja *ex-post* Phare 2003 komponent Rozwój Zasobów Ludzkich – badanie CAWI wśród uczestniczących w programie nauczycieli).

W początkowej fazie wdrażania przez Agencję Programów Operacyjnych na lata 2004–2006, na potrzeby ewaluacji były realizowane kolejne badania CAWI ze stron ewaluacja.parp.gov.pl lub badania.parp.gov.pl (m.in. badania ocen i opinii wnioskodawców działania 2.1 SPO WKP realizowane w 2006 r. we współpracy z komórką audytową). Jednak przełomowe znacznie dla rozwoju technik badań internetowych wykorzystywanych w ewaluacjach PARP, miało uruchomienie w roku 2007 dwóch projektów cy-

¹⁶ J. Pokorski, *Specyfika zarządzania w projektach ewaluacyjnych, na przykładzie Oceny Rezultatów wybranych działań Sektorowego Programu Operacyjnego Wzrost Konkurencyjności Przedsiębiorstw* (praca dyplomowa), Szkoła Główna Handlowa, Warszawa 2009.

¹⁷ Wykorzystano podstawowe funkcjonalności techniki CAWI (m.in. automatyczne filtry oraz rotacje i inwersje kafeterii odpowiedzi w pytaniach ankiety *on-line*, zaproszenie z linkiem do strony badania przesyłane pocztą elektroniczną do respondentów).

¹⁸ Załączana ankietą (a raczej formularz zawierający pakiet wskaźników danego beneficjenta oraz dodatkową część jakościową) była aplikacją przygotowaną w popularnym arkuszu kalkulacyjnym. Zawierała dwa panele do wprowadzania danych przez respondenta oraz ukryty arkusz analityczny. Być może, gdyby w przypadku oceny efektów tego programu zastosowano ankietę CAWI z opcją logowania dla beneficjentów i zagnieźdżonymi formularzami wskaźników poszczególnych beneficjentów, znacznie usprawniłoby to proces realizacji. Chociaż i tak było to rozwiązanie korzystniejsze od techniki projektowanej jeszcze w fazie realizacji projektów (lata 2002–2003) – papierowych raportów wpływu wysyłanych pocztą (w najlepszym razie wypełnionych w edytorze tekstu i odsyłanych pocztą elektroniczną). Początkowo forma mailowa była rozwiązaniem bardzo funkcjonalnym i zapewniała bardzo dobrą jakość danych (wykorzystano duże możliwości walidacji podawanych w Excelu danych wskaźnikowych – ograniczenie formatu danych, dopuszczalne przedziały wartości, kontekstowe komunikaty do respondentów; w późniejszych latach wprowadzono także automatyczne graficzne trendów w zakresie osiąganych wartości wskaźników beneficjenta), z czasem w toku rozwijania przez PARP techniki CAWI, dostrzegano ograniczenia ankiety *e-mail*. Ze względu na stosunkowo niewielki zakres projektu, fakt, że respondenci przyzwyczaili się do narzędzia (stosowanego w tym przypadku jak w badaniach *trackingowych* – ang. *tracking analysis*) i ryzyka obciążenie badania błędami związanymi ze zmianą techniki, nie warto było przechodzić na formę *Web-survey*.

klucznych badań kwartalnych, dotyczących beneficjentów SPO RZL (działanie 2.1, schemat a i b) oraz SPO WKP (działania 2.1 i 2.3). W badaniach CAWI realizowanych w ramach ewaluacji *on-going* wymienionych działań brali udział przedsiębiorcy, którzy dzięki programowi realizowali projekty doradcze, inwestycyjne lub których pracownicy korzystali z dofinansowanych szkoleń/studiów podyplomowych. Było to projekty wieloletnie, realizowane na dużo większą niż dotychczas skalę. Przykład „Oceny rezultatów działania 2.1 i 2.3 SPO WKP po 18 miesiącach” szerzej zostanie scharakteryzowany w dalszej części artykułu.

Wielość badań internetowych w ramach ewaluacji prowadzonych w PARP i stałe starania o wprowadzanie usprawnień, rozwijanie dotychczasowych podejść i stosowanych technologii CAWI, zainicjowały w 2008 r. projekt związany z wdrożeniem zaawansowanej platformy badań internetowych¹⁹.

W kolejnych latach nowej perspektywy finansowej 2007–2013, doskonale sprawdzająca się w badaniach typu B2B wśród beneficjentów PARP, technika CAWI nadal nie traci na znaczeniu. Obecnie jest wykorzystywana w badaniach ewaluacyjnych *ex-post* dotyczących poprzedniego okresu programowania (m.in. zlecona ocena efektów w zakresie innowacyjności poddziałania 2.2.1 SPO WKP²⁰), jak również planowana do zastosowania na większą skalę w ramach ewaluacji działań POIG, POKL i PORPW wdrażanych przez PARP. Na obecnym etapie finalizowane są działania przygotowawcze, związane z pracą grup eksperckich ds. ewaluacji przy wskaźnikach i narzędziach ewaluacyjnych. W toku postępu procesu wdrażania, w odpowiednich działaniach programowych, planowane jest uruchomienie ewaluacji o charakterze bieżącym (*on-going*). Mając na uwadze dotychczasową skuteczność (i efektywność) technik internetowych w poprzednich projektach badawczych PARP, w wielu projektowanych ewaluacjach *on-going* nowego okresu (m.in. w trzynastu działaniach POIG i instrumentach skierowanych do przedsiębiorstw w POKL), CAWI z całą pewnością będzie mogło być wykorzystywane przede wszystkim.

Ocena rezultatów działania 2.1 i 2.3 SPO WKP po 18 miesiącach (studium przypadku)²¹

Dobrym przykładem wykorzystania w badaniach ewaluacyjnych szerokich możliwości, jakie oferuje technika CAWI, jest realizowana przez PARP, ocena rezultatów dwóch działań dotacyjnych skierowanych do przedsiębiorców – działania 2.1 i 2.3 SPO WKP. Badanie zostało uruchomione w 2007 roku i docelowo (do połowy 2010 r.) obejmie ponad 4 000 wspartych małych i średnich przedsiębiorstw (MSP). Wybór techniki badawczej wynikał w pierwszej kolejności z dotychczasowych dobrych doświadczeń Agencji z realizacji badań *on-line* wśród przedsiębiorców (m.in. beneficjentów programów przedakcesyjnych Phare), ale również z szeroko zakrojonych celów i funkcji (potrzeb informacyjnych), jakie badanie miało spełnić dla poszczególnych odbiorców wyników, na które technika CAWI w pełni odpowiadała. Projekt „Ocena rezultatów działania 2.1 i 2.3 SPO WKP po 18 miesiącach” jest ewaluacją wewnętrzną o charakterze

¹⁹ Oprogramowanie, stosowane m. in. przez wyspecjalizowane w badaniach internetowych podmioty, pozwalające projektować narzędzia CAWI w oparciu o przyjazny interfejs użytkownika (np. praca z oknami dialogowymi i kreatorami graficznymi, zamiast konieczności „tłumaczenia” projektów kwestionariuszy na „hermetyczny” język programowania – PHP, SQL). (por. <http://bip.parp.gov.pl/index/more/3002> oraz <http://bip.parp.gov.pl/index/more/4065> z 11 października 2009 r.).

²⁰ Por. <http://bip.parp.gov.pl/index/more/11319> z 10 października 2009 r., w szczególności Opis wymagań funkcjonalnych Zamawiającego dotyczących narzędzia CAWI oraz Przebieg fazy realizacji badania CAWI (Załącznik 1 do Specyfikacji Istotnych Warunków Zamówienia).

²¹ W niniejszym podrozdziale zostały wykorzystane fragmenty pracy dyplomowej autora pt. „Specyfika zarządzania w projektach ewaluacyjnych, na przykładzie Oceny Rezultatów wybranych działań Sektorowego Programu Operacyjnego Wzrost Konkurencyjności Przedsiębiorstw” napisanej w ramach Podyplomowego Studium Zarządzania Projektami (*Project Management*) Szkoły Głównej Handlowej w Warszawie.

*on-going*²², która zakładała systematyczne badanie beneficjentów obu działań²³, zorientowane przede wszystkim na pomiarze długofalowych efektów udzielonego wsparcia.

W kolejnym roku wdrażania SPO WKP pojawiła się „paląca potrzeba” poznania rezultatów dotychczas zrealizowanych projektów w ramach tych dwóch kluczowych instrumentów programu. Potrzeba ta z jednej strony wynikała z wyzwań podejmowanych przez administrację różnych szczebli, związanych z programowaniem szczegółowych rozwiązań proceduralnych i wytycznych dla nowych programów wsparcia dla przedsiębiorców w ramach perspektywy 2007–2013 (POIG, 16 Regionalnych Programów Operacyjnych). Ustalenia te nie mogły odbywać się bez wiedzy na temat skuteczności dotychczas podejmowanych działań o podobnym charakterze. Z drugiej strony, obowiązki sprawozdawcze ciążyące na instytucji wdrażającej wymuszały podjęcie systematycznych badań, związanych z pomiarem określonych wskaźników wykonania celów danego działania czy poszczególnych projektów.

Fakty, które stały u podstaw przyjętych założeń projektowych, były niekiedy sprzeczne i wymagające kompromisu w podejmowanych decyzjach czy działaniach. Przykładowo dla wykazania kilku kluczowych wskaźników rezultatu i oddziaływania²⁴ – ogólnych, określonych w Uzupełnieniu SPO WKP na poziomie całego działania – można byłoby przeprowadzić badania terenowe na reprezentatywnej grupie beneficjentów w określonym horyzoncie czasowym. Tu moment pomiaru projektodawcy wyznaczyli na 18 miesięcy od zakończenia realizacji danego projektu (nie działania, co warto podkreślić). Natomiast na poziomie projektów łącznie było kilkadziesiąt tysięcy różnych wskaźników dotyczących kilku tysięcy podpisanych umów o dofinansowanie. Miały one roczną częstotliwość pomiaru w perspektywie kilkuletniego okresu przyjętej trwałości projektu. Co więcej wskaźniki projektowe nieagregowały się na poziomie działania w postaci wskaźników ogólnych (wyższego rzędu) – każdy projekt miał unikalny portfel wskaźników, który z założenia weryfikował efekty rzeczowe na poziomie pojedynczego projektu, nie pozwalał natomiast na pokazanie tych samych efektów w ujęciu ogółu dofinansowanych przedsięwzięć²⁵.

W przypadku wskaźników ogólnych (na poziomie działania) czy „miękkich” ocen i opinii beneficjentów, odnoszących się do efektów podjętych działań, badanie wszystkich projektów nie było konieczne (w przeciwieństwie do wskaźników szczegółowych z poszczególnych umów o dofinansowanie). W pewnym momencie nieproporcjonalnie zaważyłoby koszty badania w stosunku do niewielkich korzyści w postaci dokładności pomiaru. Z całą pewnością wystarczyłoby zastosowanie metod reprezentatywnych.

²² Z punktu widzenia projektów była to ewaluacja *ex-post* prowadzona po 18 miesiącach od ich zakończenia, natomiast z perspektywy działania, którego realizacja jeszcze się nie zakończyła (w 2007 i 2008 r.), była to ewaluacja śródkresowa o charakterze *on-going*.

²³ Beneficjentami przedmiotowych działań SPO WKP były przedsiębiorstwa sektora MSP, które zrealizowały projekty doradcze (działanie 2.1) lub inwestycyjne (działanie 2.3) i uzyskały z programu refundację części poniesionych kosztów (tj. zakupu usług doradczych – 2.1 czy zakupu środków trwałych – 2.3). Ogólnie rzecz biorąc przedsięwzięcia te miały wzmocnić konkurencyjność firm-beneficjentów. W przypadku działania 2.1 były to głównie projekty zakładające doradztwo w zakresie jakości, innowacji i nowych technologii, zasad prowadzenia firmy na terenie UE lub pozyskania zewnętrznego finansowania na rozwój. Działanie 2.3 natomiast wspierało inwestycje MSP, związane z modernizacją (prowadzącą do znaczącej zmiany produktu lub procesu produkcji), wdrażaniem i komercjalizacją technologii i produktów innowacyjnych czy zastosowaniem technologii IT/ITC w zarządzaniu.

²⁴ Były to tzw. wskaźniki ewaluacyjne, które wykraczają poza sprawozdanie końcowe z realizacji projektów i wymagały badań, a nie tylko agregacji danych zastanych.

²⁵ Przykładowo, wskaźniki projektowe dotyczące zatrudnienia były sformułowane w różnorodny sposób w poszczególnych projektach np. utworzenie X nowych miejsc pracy; zwiększenie zatrudnienia o X% w stosunku do roku bazowego, utrzymanie dotychczasowego stanu zatrudnienia; zwiększenie liczby pracowników itd. Znając wartość jednego ze wskaźników dotyczących zatrudnienia we wszystkich projektach, w żaden sposób nie można było obliczyć (zagregować) „nowych miejsc pracy – w podziale na kobiety i mężczyzn – utworzonych we wspartych przedsiębiorstwach”. Tym samym wiedza o skuteczności projektów w tym obszarze, nie pozwalała na potwierdzenie tej skuteczności w odniesieniu do wskaźników działania.

W toku podejmowanych ustaleń pomiędzy stronami udało się sformułować wspólne cele²⁶ inicjowanego projektu.

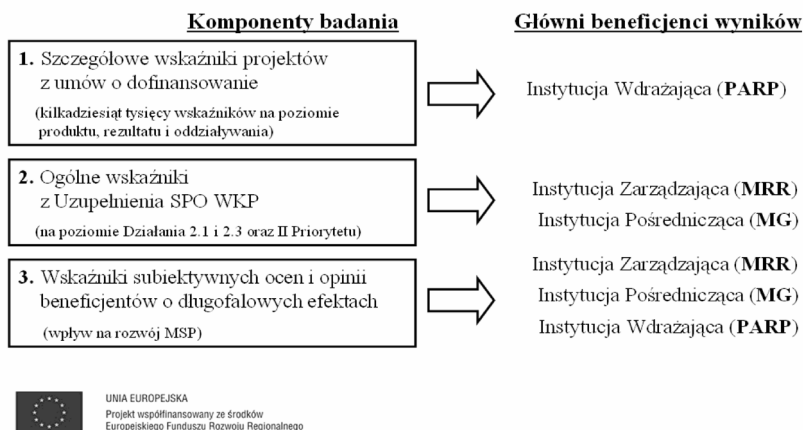


POLSKA AGENCJA ROZWOJU PRZEDSIĘBIORCZOŚCI
POLISH AGENCY FOR ENTERPRISE DEVELOPMENT



UNIA DLA PRZEDSIĘBIORCZYCH
PROGRAM KONKURENCYJNOŚĆ

STRUKTURA BADANIA



UNIA EUROPEJSKA
Projekt współfinansowany ze środków Europejskiego Funduszu Rozwoju Regionalnego

Rys. 1. Struktura badania

Źródło: Haber A., Pokorski J., „Ocena rezultatów SPO WKP. Systemowe podejście do mierzenia długofalowych efektów programu” Prelekcja na III Konferencji Ewaluacyjnej pt. „Rola Ewaluacji w kreowaniu polityki spójności” (Warszawa, 25 października 2007 r.), organizowanej przez Ministerstwo Rozwoju Regionalnego i Polską Agencję Rozwoju Przedsiębiorczości (por. www.konferencja-ewaluacja.pl).

Ogólna koncepcja zakładała badanie techniką CAWI docelowo wszystkich rozliczonych projektów. Badanie miało mieć charakter cykliczny (falowy) realizowany w ramach rund kwartalnych. Do danej rundy badania mieli kwalifikować się beneficjenci, w przypadku których od rozliczenia projektu upłynął okres między 16,5 a 19,5 miesiąca²⁷.

W przypadku zasobów technicznych, istotne znaczenie miało wykorzystanie wewnętrznych serwerów Agencji do prowadzenia badań oraz zastosowanie certyfikatu SSL do zabezpieczenia wprowadzanych danych przez respondentów. Powodzenie w tej kwestii z całą pewnością zwiększyło zaufanie do badania wśród przedsiębiorców, którzy niejednokrotnie musieli podawać dane wrażliwe z punktu widzenia tajemnicy firmy (np. dane finansowe). Wiarygodność badania zwiększało zapewne korzystanie z serwerów Agencji i tym samym adresów rządowych – parp.gov.pl.

²⁶ Celem ogólnym projektu była ocena rezultatów projektów działania 2.1 i 2.3 SPO WKP po 18 miesiącach od zakończenia ich realizacji. Cele szczegółowe natomiast odpowiadały wyżej opisanym potrzebom i sformułowano je następująco: 1) zmierzenie efektów Działania 2.1 i 2.3 SPO WKP w odniesieniu do wskaźników z Uzupełnienia Programu (miejsca pracy k/m, przychody, poziom eksportu, nakłady inwestycyjne, certyfikaty, innowacje, dokumenty strategiczne...); 2) wykazanie efektów projektów w zakresie wskaźników założonych w umowie o dofinansowanie (badanie indywidualnych, nieagregowalnych wskaźników projektu, które posłużą na potrzeby procesów monitorowania); 3) poznanie ocen i opinii beneficjentów dotyczących długofalowych efektów projektów (przede wszystkim w odniesieniu do kryteriów skuteczności i efektywności wsparcia).

²⁷ Dla zwiększanie efektywności fazy realizacji, przyjęto kompromis związany z horyzontem pomiaru. Zamiast punktowego podejścia po 18 miesiącach dla każdego projektu, zaproszenie do udziału w badaniu wysyłało co kwartał, do beneficjentów wpadających w przedział czasowy 16,5–19,5 miesiąca od rozliczenia projektów. Z kolejnych rund wyłączone były projekty już przebadane. Przeciętnie każda runda badania w danym działaniu liczyła ok. 200 przypadków. Dane zebrane z kolejnych kwartałów były łączone i analizowane zarówno w rozbiu na rundy, jak i łącznie. Wykorzystywano również istotne zmienne niezależne, charakteryzujące profil beneficjentów czy typ projektów w celu wyodrębnienia interesujących przekrojów analitycznych.

Przyjęto, że narzędzie będzie składało się z 3 głównych elementów: formularza wskaźników szczegółowych²⁸, formularza wskaźników ogólnych²⁹, ankiety oceniającej.

Układ formularza wskaźników szczegółowych w narzędziu CAWI odpowiadał układowi wskaźników w Generatorze Wniosków dla działań II Priorytetu SPO WKP, w którym przedsiębiorcy deklarowali wskaźniki realizacji projektów i ich wartości. Miało to niewątpliwą korzyść – beneficjenci mogli swobodnie poruszać się po narzędziu ze znajomym interfejsem. Beneficjenci w formularzu szczegółowym mogli zobaczyć swoje wskaźniki, które deklarowali we wniosku o dofinansowanie, w rozbiu na kolejne lata zakładanej trwałości projektu.

WSKAŹNIKI PRODUKTÓW

Jeśli nazwy bądź wartości założone dla wskaźników są nieprawidłowe, prosimy kliknąć [tutaj](#).

Nazwa wskaźnika: Liczba dni usługi doradczej z zakresu zarządzania jakością ?

	2004	2005	2006	2007	2008	2009	2010
Wartość założona (na podstawie umowy)	0,00	17,00	0,00	0,00	0,00	0,00	0,00
Wartość zrealizowana	0	16	0	0	-	-	-

Jednostka wskaźnika: szt

Nazwa wskaźnika: Liczba dni na przeprowadzenie auditu certyfikującego ?

	2004	2005	2006	2007	2008	2009	2010
Wartość założona (na podstawie umowy)	0,00	0,00	4,00	0,00	0,00	0,00	0,00
Wartość zrealizowana					-	-	-

Jednostka wskaźnika: szt

Nazwa wskaźnika została zaczerpnięta z Bazy Beneficjentów administrowanej przez PARP. Nazwa ta została zadeklarowana przez Państwa we Wniosku o dofinansowanie.

WSKAŹNIKI REZULTATÓW

WSKAŹNIKI ODDZIAŁYWANIA

WSKAŹNIKI PROJEKTU

Rys. 2. Formularz szczegółowy

Źródło: Ocena rezultatów działania 2.1 i 2.3 SPO WKP, po 18 miesiącach, PARP.

Formularz ogólny wyświetlał tabele ze wskaźnikami, które zawierały miejsce na wprowadzenie wartości osiągniętej w danym okresie (wartość wskaźnika musiała być podana zgodnie z sugerowaną jednostką miary i zgodnie z wskazanym przedziałem czasowym – stan na dany rok lub na moment badania) oraz pytania zamknięte i otwarte, w przypadku których należało wybrać lub wpisać odpowiedź. W przy-

²⁸ Każdy przedsiębiorca składając wniosek o dofinansowanie, zakładał osiągnięcie określonego zestawu wskaźników dotyczących projektu. Formularz wskaźników szczegółowych miał zawierać listę wskaźników monitorowania, zadeklarowanych przez Beneficjenta dla danego projektu i potwierdzonych ostatecznie w umowie o dofinansowanie. Formularz był przypisany indywidualnie do każdego z projektów zrealizowanych przez przedsiębiorstwo. W przypadku gdy przedsiębiorca zrealizował więcej niż jeden projekt – jeśli od jego rozliczenia upłynął już okres około półtora roku – miał wypełniać niezależne formularze wskaźników szczegółowych dla każdego z zrealizowanych projektów. Jeśli przedsiębiorca stosunkowo niedawno zrealizował jeszcze inne projekty w ramach któregoś działania, w sytuacji gdy upłynął okres 16,5–19,5 miesiąca od ich rozliczenia, miał otrzymać informację z prośbą o wypełnienie formularza dla kolejnego projektu.

²⁹ Formularz wskaźników ogólnych miał służyć zgromadzeniu informacji statystycznych o efektach wszystkich projektów (zrealizowanych w ramach działania 2.1 i 2.3 SPO WKP) w obszarze działalności przedsiębiorstw. Beneficjenci zostali poinformowani, że na podstawie Uzupelnienia SPO WKP, 2004–2006, PARP jest zobowiązana do zgromadzenia informacji o łącznych efektach udzielonych dotacji, w postaci wskaźnikowej. Na co nie pozwalał pomiar unikalnych wskaźników projektowych, które nie dawały się agregować na poziomie działań. Do tego celu został ustalony jednolity zestaw wskaźników, dotyczących działalności wspartych przedsiębiorstw (beneficjentów), który pozwolił miał na wykazanie łącznych efektów udzielonych dotacji (ze środków Unii Europejskiej i polskich środków publicznych) w ramach SPO WKP.

padku wąskiego katalogu wskaźników przyjętego w formularzu ogólnym, beneficjenci mogli korzystać z pomocy podręcznej (ukrytej pod znakiem zapytania obok wskaźnika – por. rys. 3). Zawierała ona m.in. definicje wskaźników, sposób obliczania ich wartości oraz inne pomocne informacje dla poprawnego wypełnienia formularza.

ZATRUDNIENIE (pyt. 1-7)

nazwa wskaźnika	pomoc	przedział czasowy	wartość wskaźnika
1. Poziom zatrudnienia ogółem w ostatnim okresie sprawozdawczym	?	Średnia za rok 2007	Przeciętna liczba zatrudnionych na podstawie umowy o pracę w roku 2007, w przeliczeniu na pełny wymiar czasu pracy (pełny etat)
2. Poziom zatrudnienia ogółem w ostatnim okresie sprawozdawczym			Pełny wymiar czasu pracy (pełny etat) to liczba wynikająca z sumy pełnych etatów i części (z dwoma miejscami po przecinku) np.: trzy osoby zatrudnione na 1/4 = 0,75 etatu. Przeciętne zatrudnienie należy obliczyć jako sumę przeciętnego zatrudnienia w poszczególnych miesiącach podzieloną przez 12 (bez względu na to, czy przedsiębiorstwo funkcjonowało przez cały rok, czy też nie). Przeciętne zatrudnienie w miesiącu należy obliczyć na podstawie sumy dwóch stanów dziennych (w pierwszym i ostatnim dniu miesiąca) podzielonej przez dwa.
			12 etat(-y/ów)

Rys. 3. Formularz ogólny – pomoc podręczna

Źródło: Ocena rezultatów działania 2.1 i 2.3 SPO WKP, po 18 miesiącach, PARP.

Ankieta miała służyć ocenie efektów działania i zawierała pytania dotyczące wpływu dotacji na działalność wspartych firm (w ocenie ich właścicieli lub przedstawicieli zarządu) oraz pytania metryczkowe. Były to tradycyjne pytania ankietowe: otwarte (respondent miał za zadanie wpisać w puste miejsce zwięzłą odpowiedź) i zamknięte (zaznaczyć właściwą lub kilka właściwych odpowiedzi we wskazane pola lub określić odpowiedzi na skali lub w tabeli). Ponadto projektowana ankieta *on-line* miała być dynamiczna w czasie jej wypełniania – jej kształt zależał od odpowiedzi udzielanych przez respondenta. Na początku widoczne było tylko pierwsze pytanie, a po wypełnieniu wyświetlały się kolejne. To rozwiązanie zapewniało wiele korzyści – np. eliminowało pomijanie pytań przez respondenta, ankieta była przejrzysta, a o zaawansowaniu procesu respondenta informował pasek postępu.

Ponadto narzędzie musiało mieć zaprojektowanych szereg dodatkowych elementów, które zapewniały funkcjonalność i jakość. Były to m.in. zakładki z informacją o badaniu, pełna instrukcja obsługi, tzw. FAQ cyklicznie uzupełniany przez zespół projektowy, kontakt do badacza (pracownika komórki ewaluacyjnej PARP). Standardowo pojawiły się funkcje związane z pracą narzędzia w sieci: panel logowania, klawisze nawigacji i zapisu, formularz zgłoszenia błędu, połączony z kontem poczty *e-mail* pracownika komórki ewaluacyjnej, okna podglądu wydruku i wiele innych. Na stronę badania, w oparciu o przekazany identyfikator i hasło, mogły się logować różne osoby w firmie (np. dane finansowe wypełniała księgowa, wskaźniki szczegółowe – koordynator projektu, ankietę – właściciel firmy). To dodatkowo miało ułatwić przedsiębiorcom pracę z narzędziem i jednocześnie podnieść jakość zebranych danych.

Realizacja celów badania wymagała opracowania bardzo złożonego i zaawansowanego narzędzia, w stosunku do standardowych CAWI stosowanych przez większość firm badania rynku i opinii. Pojawiły się np. zagnieżdżone formularze, połączenie szyfrowane w oparciu o certyfikat SSL wystawiony dla domeny Agencji, format zapisu informacji w formularzach szczegółowych, zintegrowany z systemem baz danych stosowanym przez PARP.

Mimo że ankieta była stosunkowo trudna i sprawiała beneficjentom sporo wysiłku, liczba zapytań mailowych czy telefonicznych nie była wyjątkowa w porównaniu z innymi badaniami tego typu, realizowanymi na podobnej grupie. Ale na przykład okazało się, że bazy zrealizowanych projektów i beneficjentów zawierają sporo błędów i do tego bardzo się zdezaktualizowały od etapu wnioskowania o dotację. Jedną rzecz to, że nie uaktualniano w bazach informacji wnoszonych na mocy aneksu do umowy (m.in. w za-

krześle zakładanych wskaźników), jednak nawet bez tego, w okresie 18 miesięcy od rozliczenia projektów bardzo wiele firm zmieniało nazwę, siedzibę, dane kontaktowe (*e-mail*, telefon) czy nawet przekształcało spółkę (np. fuzje lub przejęcia niejednokrotnie powodowały, że zmieniał się NIP beneficjenta – podstawowa zmienna, stanowiąca identyfikator respondenta w badaniu). To wszystko generowało dodatkową pracę w projekcie (wymagało uaktualniania bazy i poprawiania profilu respondenta na stronie badania). Ponadto powodowało, że pisma przewodnie wysyłane pocztą niejednokrotnie nie trafiały pod właściwy adres lub nieskuteczne okazywały się monity (mailing przypominający). Dodatkowo wykorzystywano indywidualne kanały komunikacji z respondentami (skany listów przewodnich, wysyłane na nowe adresy mailowe czy bezpośredni kontakt telefoniczny). Doświadczenia pierwszych rund badania CAWI pozwoliły jednak na wypracowanie modelowych rozwiązań poszczególnych problemów, kontrolę przebiegu realizacji kolejnych rund badania i wypracowanie mechanizmu automatyki procesu, powtarzalności określonych czynności, które zapewniały wysoką jakość i stabilny przebieg fazy realizacji.

Nie sposób zliczyć czy wymienić wszystkie czynniki sukcesu tego projektu. Dotyczyły one sfery projektowej, wykonawczej i zarządczej. Z całą pewnością sukces projektu budowali przede wszystkim ludzie zaangażowani w jego realizację. Popyt na „produkt” i „skala sprzedaży”, którą potwierdzają liczne prezentacje, konferencje czy opracowania wyników jest niepodważalnym faktem. Doceniono przede wszystkim rzetelność prowadzonych badań i ich interesujące wyniki, oparte na twardych danych wskaźnikowych. Projekt stanowi niewątpliwą wartość dodaną w działalności Agencji, wniósł istotny wkład w rozwój ewaluacji *on-going* w Polsce oraz zastosowanie techniki CAWI w badaniach ewaluacyjnych. W trakcie realizacji kolejnych rund „Oceny rezultatów” w Agencji były realizowane inne projekty ewaluacyjne, wykorzystujące technikę CAWI. Szereg doświadczeń z badań kwartalnych 2.1 i 2.3 SPO WKP wykorzystywano przy wprowadzaniu usprawnień w nowych projektach³⁰.

Ponadto rozwijanie doświadczeń z realizacji projektu zostało zaplanowane w ewaluacjach *on-going* dedykowanych nowym działaniom realizowanym przez Agencję (m.in. w ramach Programu Operacyjnego Innowacyjna Gospodarka, Kapitał Ludzki czy Rozwój Polski Wschodniej, na lata 2007–2013). Ponadto, w porozumieniu z Grupą Sterującą Ewaluacją POIG, na 2009 r. zaplanowano uzupełnienie prowadzonej „Oceny rezultatów” o badania odpowiednio dobranej grupy kontrolnej celem przeprowadzenia analizy efektów netto (oszacowania skali zmian we wspartych przedsiębiorstwach, wywołanych wyłącznie przez interwencję, po oddzieleniu wpływu czynników zewnętrznych, takich jak np. zmiany koniunktury, efekty innych dotacji itp.).

Z całą pewnością można wyodrębnić w projekcie również szereg ograniczeń czy słabszych stron, których nie udało się przezwyciężyć (np. lepsza koordynacja w fazie realizacji pomiędzy oboma działaniami, odciążenie respondentów w zależności od typu projektu, zwiększenie automatyki fazy realizacji i tym samym odciążenie zespołu badawczego, być może również lepsze wskaźniki i narzędzia pomiaru czy optymalizacja horyzontu pomiaru wskaźników). Stanowią one wyznawania w nowej fali ewaluacji *on-going*, w przyszłości realizowanych przez PARP w dużej mierze w oparciu o technikę CAWI.

³⁰ Przykładowo, w badaniu efektów działania 1.2 SPO WKP (którego przedmiotem było m.in. dokapitalizowanie funduszy mikropożyczkowych i funduszy poręczeń kredytowych), dla uelastycznienia pracy badaczy w fazie realizacji, przygotowano specjalny panel administracyjny (rozwiązanie określane jako system zarządzania treścią typu CMS – ang. *Content Management System*), który pozwalał, bez użycia języka programisty, na generowanie statystyk z realizacji, wysyłanie monitów drogą mailową, eksport danych, aktualizację FAQ, edycję pytań ankiety i innych informacji widocznych dla respondentów na stronie badania itp.).

Zakończenie

Szeroko pojęty rozwój Internetu, upowszechnianie dostępu do sieci WWW i wzrost aktywności – społecznej, gospodarczej i kulturowej – zarówno osób prywatnych, jak i różnego typu organizacji, instytucji czy ich sieci, dokonujący się aktualnie „na naszych oczach”, nie może pozostać bez wpływu na sektor badań (w tym badań ewaluacyjnych). Poprzez ciągły, dynamiczny wzrost liczby użytkowników Internetu wzrasta populacja potencjalnych respondentów badań realizowanych poprzez sieć. Dotykają one coraz większej liczby ważnych społecznie problemów, zjawisk, rzeczy. Wraz z rozwojem Internetu wyznaczone są nowe kierunki rozwoju i zastosowań dla techniki CAWI. Badania ewaluacyjne z całą pewnością mogą być doskonałą ku temu materią – z korzyścią dla doskonalenia samej techniki, a przede wszystkim dla jakości dostarczanych w oparciu o nią wyników badań ewaluacyjnych. Już na dzień dzisiejszy trudno wskazać te rodzaje programów poddawanych ewaluacji, typy respondentów (beneficjentów lub innych zaangażowanych w obszarze programu potencjalnych informatorów) albo obszary interwencji, gdzie nie mogłyby mieć zastosowania techniki badań internetowych. Bez wątpienia, jak pokazały wieloletnie doświadczenia m.in. Polskiej Agencji Rozwoju Przedsiębiorczości, CAWI doskonale znajduje zastosowanie w badaniach realizowanych wśród przedsiębiorców (przede wszystkim wśród firm małych, średnich i dużych oraz mikroprzedsiębiorstw opartych o nowoczesne technologie czy działających w sektorze e-usług). Liczne prace teoretyczne zmierzające do ugruntowania nowej techniki i praktyka badań komercyjnych, wskazują coraz szersze zastosowania CAWI. Co do potrzeby wdrożenia w większym (i lepszym) zakresie techniki CAWI w badaniach ewaluacyjnych, można powtórzyć zdanie Taylora i Terhaniana wypowiedziane w kontekście badań *on-line* – „jest to rozpędzony pociąg, którego nie można zatrzymać i przyspiesza on coraz bardziej. Ci, którzy nie wsiadają do niego, ryzykują, że pozostaną daleko w tyle”³¹.

Jacek Pokorski – absolwent Instytutu Politologii Uniwersytetu Kardynała Stefana Wyszyńskiego oraz Katedry Zarządzania Projektami Szkoły Głównej Handlowej w Warszawie. Od ponad pięciu lat związany z Sekcją Ewaluacji Polskiej Agencji Rozwoju Przedsiębiorczości. Specjalizuje się w ewaluacji programów wsparcia konkurencyjności i innowacyjności gospodarki, w szczególności skierowanych do sektora MŚP oraz otoczenia biznesu. Kierownik i współautor wielu projektów badawczych (w tym badań ewaluacyjnych realizowanych techniką CAWI). Doświadczenie w ewaluacji programów współfinansowanych z funduszy przedakcesyjnych i strukturalnych Unii Europejskiej (w szczególności Phare SSG 2000–2003, program Wzrost Konkurencyjności Przedsiębiorstw 2004–2006, program Innowacyjna Gospodarka na lata 2007–2013). Członek Grupy Sterującej Ewaluacją Programu Operacyjnego Innowacyjna Gospodarka oraz Grupy roboczej ds. ewaluacji przy Komitecie Monitorującym PO IG. Obecnie zaangażowany w przygotowywanie kompleksowego systemu pomiaru efektów 13 działań PO IG wdrażanych przez PARP. Autor i współredaktor opracowań i publikacji poświęconych ewaluacji oraz efektom projektów współfinansowanych ze środków UE. Od kilku lat wykłada na studiach podyplomowych i prowadzi szkolenia z ewaluacji dla przedstawicieli instytucji publicznych oraz organizuje konferencje poświęcone tej tematyce.

³¹ H. Taylor, G. Terhanian, *Heady days are here again: Online rolling is rapidly doming of Age*, Public Perspective, nr 10, 1999.

Bibliografia

- Babbie E. (2003), *Badania społeczne w praktyce*, PWN, Warszawa.
- Batorski D., *Korzystanie z technologii informacyjno komunikacyjnych*, [w:] Czapiński J., Panek T., *Diagnoza społeczna 2009. Warunki i jakość życia Polaków*.
- Batorski D., Olcoń M., *Prowadzenie badań przez Internet: Podstawowe zagadnienia metodologiczne*, „Studia Socjologiczne”, 3(182).
- Ewaluacja Narodowego Planu Rozwoju i programów operacyjnych w Polsce. Poradnik* (2005), Krajowa Jednostka Oceny, Departament Koordynacji Polityki Strukturalnej, MGiP, Warszawa.
- Domański H., Lutyńska K., Rostocki A (red.) (1999), *Spojrzenie na metodę. Studia z metodologii badań socjologicznych*, IFiS PAN, Warszawa.
- Haber A., Pokorski J. (2007), *Ocena rezultatów SPO WKP. Systemowe podejście do mierzenia długofalowych efektów programu*, Prelekcja na III Konferencji Ewaluacyjnej pt. „Rola Ewaluacji w kreowaniu polityki spójności” (Warszawa, 25 października 2007 r.), organizowanej przez Ministerstwo Rozwoju Regionalnego i Polską Agencję Rozwoju Przedsiębiorczości.
- Hewson, C., Yule, P., Laureat, D., Kogel, C. (2003). *Internet Research Methods. A practical guide for social and behavioral sciences*, London: Sage Publications.
- Pokorski J. (2009), *Specyfika zarządzania w projektach ewaluacyjnych, na przykładzie Oceny Rezultatów wybranych działań Sektorowego Programu Operacyjnego Wzrost Konkurencyjności Przedsiębiorstw* (praca dyplomowa), Szkoła Główna Handlowa, Warszawa.
- Pokorski J. (2008), *System ewaluacji wykorzystania funduszy strukturalnych w Polsce. Ramy instytucjonalne i prawne* (praca dyplomowa), Uniwersytet Kardynała Stefana Wyszyńskiego, Warszawa.
- Raport pt. „Badanie zapotrzebowania na działania wspierające rozwój usług świadczonych elektronicznie (e-usług) przez przedsiębiorstwa mikro i małe”* (2009), Pentor RI na zlecenie PARP, Warszawa.
- Polskie Towarzystwo Badaczy Rynku i Opinii, *Katalog PTBRiO 2008. Rynek badań. Badacze. Firmy Badawcze*. XIII edycja.
- Rossi P., Freeman H., Lipsey M. (1999), *Evaluation – a systematic approach*, 6th edition, Sage Publication Inc., Thousand Oaks.
- Taylor H., Terhanian G. (1999), *Heady days are here again: Online rolling is rapidly doming of Age*, Public Perspective, nr 10.
- The GUIDE – *Evaluation of socio-economic development* (2004), Komisja Europejska, (www.evaled.info).
- Zajac J. (2007), *Materiały do kursu „Wykorzystanie Internetu w badaniach społecznych”* Uniwersytet Warszawski.
- Zajac J. M., Batorski D. (2007), *Jak skłonić do udziału w badaniach internetowych: Zwiększanie realizacji próby*. „Psychologia Społeczna”, nr 3.

Metody zwiększania zwrotności w badaniach internetowych

Wprowadzenie

W ostatnich latach coraz popularniejsze staje się prowadzenie badań, w tym również badań ewaluacyjnych, za pomocą ankiet internetowych (CAWI – *Computer-Assisted Web Interview*). Ten sposób prowadzenia badań ma swoje olbrzymie zalety, jak choćby możliwość szybkiej realizacji badań na dużych próbach przy małych kosztach czy możliwość dotarcia do bardzo specyficznych populacji, często trudno dostępnych w inny sposób. Zarazem nie jest jednak pozbawiony ograniczeń, spośród których wiele wiąże się z niską realizacją próby, ograniczającą możliwości generalizacji wyników.

Niektóre rodzaje i typy badań w większym stopniu niż inne nadają się do prowadzenia przez Internet. Jak będziemy starali się pokazać w niniejszym rozdziale, specyfika badań ewaluacyjnych sprawia, że w bardzo wielu sytuacjach znacznie lepiej nadają się one do realizacji przez Internet niż inne badania. Jednocześnie badania takie mogą być w pełni wartościowe – nie gorsze, a często nawet lepsze niż te prowadzone tradycyjnymi metodami, o ile tylko przestrzegane są odpowiednie standardy metodologiczne¹.

Jednym z najistotniejszych problemów związanych z realizacją badań ankietowych przez Internet jest zapewnienie wystarczająco wysokiego stopnia realizacji próby (*response rate*), czyli odsetku zrealizowanych ankiet w stosunku do wielkości próby². Jeśli stopień realizacji próby jest niski, a motywacja do udziału w badaniu istotnie związana z tematyką badania, to wyniki badania mogą być zniekształcone lub w ogóle bezwartościowe. Wynika to z występowania błędu wynikającego z braku odpowiedzi (*non-response error*), czyli skutku tego, że nie wszystkie wylosowane w próbie jednostki mogą bądź chcą uczestniczyć w badaniu. Odpowiedzi osób, które nie wzięły udziału w badaniu, mogą istotnie różnić się od odpowiedzi osób, które wzięły udział. Wówczas nieuprawniona staje się generalizacja wyników z próby na populację³.

Nie oznacza to jednak, że realizacja badań przez Internet zawsze wiąże się z niższym *response rate*. W literaturze i raportach można znaleźć przykłady badań CAWI, także ewaluacyjnych, z realizacją próby rzędu 80–90%, czyli wyższą niż w standardowych badaniach prowadzonych poza siecią. Sposób przeprowadzenia badania może w bardzo dużym stopniu zwiększyć bądź obniżyć chęć do udziału w nim.

¹ D. Batorski, M. Olcoń, *Prowadzenie badań przez internet: Podstawowe zagadnienia metodologiczne*, „Studia Socjologiczne”, 3(182), 2006, s. 99–132; Zajęc, J.M., Batorski, D., *Jak skłonić do udziału w badaniach internetowych: Zwiększanie realizacji próby*, „Psychologia Społeczna”, nr 3, 2007, s. 234–247.

² Warto zauważyć, że w literaturze często pojawia się alternatywna definicja *response rate*, uwzględniająca wszystkie rozpoczęte ankiety, niezależnie od tego, czy zostały one wypełnione do końca bądź przynajmniej w stopniu, który umożliwia uwzględnienie ich wyników. W badaniach internetowych, ze względu na dość częste zjawisko rezygnacji w trakcie badania (*drop out*) może występować istotna różnica między odsetkiem respondentów rozpoczynających i kończących badanie.

³ Oczywiście pod warunkiem, że badania robione są na próbach losowych, bo tylko wtedy możliwe jest stosowanie metod wnioskowania statystycznego. W praktyce nie wszystkie badania ankietowe realizowane w Internecie są przeprowadzane na próbach losowych.

W tym artykule zamieściliśmy konkretne wskazówki służące zwiększaniu stopnia realizacji próby w badaniach CAWI na różnych etapach procesu badawczego. Przedstawione rekomendacje dotyczą badań z kwestionariuszami zamieszczanymi na stronach internetowych, niezależnie od sposobu zapraszania. Poruszamy tu zarówno problematykę zwiększania odsetka respondentów rozpoczynających badanie (*survey starting rate*), jak i zmniejszania odsetka osób rezygnujących w trakcie badania (*dropout*). Innymi słowy, omawiamy zarówno kwestie skłonienia badanego do rozpoczęcia badania, jak i utrzymania jego zainteresowania i zmotywowania go do pozostania w badaniu do końca.

Wpisanie badań ewaluacyjnych metodą CAWI w projekt i jego strukturę

Mogłoby się wydawać, że metoda CAWI pozwala na bardzo szybkie i łatwe przeprowadzenie badań ewaluacyjnych, jednak odpowiednie przygotowanie badań ma kluczowe znaczenie dla jakości uzyskanych wyników. Dlatego też przygotowania badań ewaluacyjnych należy zacząć na możliwie wczesnym etapie realizowania projektu, który podlega ewaluacji lub nawet przed jego rozpoczęciem. Nie chodzi tu wyłącznie o decyzję o wykorzystaniu CAWI, ale przede wszystkim o podjęcie działań, które przyczynią się do większej realizacji próby badawczej, a w przypadku słabej realizacji i niefortunnych efektów błędu braku odpowiedzi pozwolą na poprawę możliwości uogólniania wyników badania za pomocą wag analitycznych (piszemy o tym szerzej w ostatniej części niniejszego rozdziału).

W pierwszej kolejności należy zapewnić środki nacisku na osoby i instytucje biorące udział w programie objętym ewaluacją, aby skłonić ich do udziału w badaniach ewaluacyjnych. Jest to szczególnie ważne w przypadku badań rozciągniętych w czasie. Obowiązek udziału w badaniach może być wręcz jednym z warunków uczestnictwa w programie. Istotne jest również zbieranie aktualnych i dokładnych danych kontaktowych do potencjalnych badanych. Nie mniej ważne jest także zebranie informacji o uczestnikach ewaluowanych działań, które będą mogły posłużyć na przykład do korygowania otrzymanych wyników. Największe możliwości daje zebranie najważniejszych informacji o wszystkich uczestnikach (por. podrozdział *Realizacja próby po zakończeniu badania*).

Dobór próby a zwiększanie *response rate*

Już na etapie doboru badanej próby można wprowadzić rozwiązania, które wpłyną na *response rate*. Przede wszystkim należy zastanowić się, czy ta metoda jest w ogóle odpowiednia dla danej populacji. Do badań metodą CAWI nie nadają się populacje, gdzie korzystanie z Internetu jest rzadkie lub też skłonność do korzystania z Internetu wśród ich członków jest bardzo zróżnicowana. To drugie wynika z faktu, że osoby częściej i bardziej aktywnie korzystające z Internetu będą znacznie bardziej skłonne do udziału w badaniu. Niedoreprezentowanie pewnych grup, do których szczególnie ciężko jest dotrzeć w badaniu internetowym, jak na przykład seniorzy czy ludzie o niskim wykształceniu, może znacząco zniekształcać wyniki badań. W niektórych przypadkach stosuje się mieszane schematy zbierania danych (*mixed-mode data collection*), umożliwiając wypełnienie ankiet także przez telefon lub na papierze. Zwiększa to szanse udziału respondentów, którzy nie wypełniliby ankiety w Internecie. Trzeba jednak pamiętać, że te same badania przeprowadzone w różnych kanałach mogą nie być równoważne. Badanie internetowe to ankieta samowypełniana, podczas gdy badania telefoniczne (CATI) prowadzone są jako wywiad ankieterski. Różnice te mogą mieć wpływ na udzielane odpowiedzi nie tylko ze względu na zjawisko „wpływu

ankieta”, ale także dlatego, że niektóre pytania – jak chociażby wielokrotnego wyboru z większą liczbą odpowiedzi są inaczej odbierane przez respondenta, gdy widzi je przed oczami, a inaczej gdy są tylko odczytywane przez ankietę.

Jeśli organizator badania ma pewną dowolność w wyborze badanej populacji, to zalecanym rozwiązaniem jest wybór mniejszej, dobrze określonej populacji, zainteresowanej tematem badania lub przekonanej, że badanie bezpośrednio jej dotyczy. Nie ma sensu badać populacji, które są przytłoczone niechcianymi wiadomościami i informacjami reklamowymi (tzw. spam), gdzie stopień realizacji próby będzie szczególnie niski⁴. Warto w tym miejscu zauważyć, że w przypadku badań ewaluacyjnych znacznie częściej niż w innych typach badań mamy do czynienia z dobrze określonymi populacjami – znacznie częściej wiadomo dokładnie, jakie jednostki do populacji należą, a jakie nie. Dodatkowo respondenci w badaniach ewaluacyjnych częściej będą też przekonani o tym, że badanie ich dotyczy (może z wyjątkiem badań dotyczących bardzo szerokich konsekwencji ewaluowanych działań).

Duże znaczenie ma też jakość informacji kontaktowych – jeśli będą niskiej jakości, np. wiele nieaktualnych adresów lub też są to adresy, gdzie przychodzi dużo spamu, to realizacja próby będzie niższa. Również w tym przypadku w dużej części badań ewaluacyjnych powinniśmy nie mieć tego problemu⁵. Warunkiem tego jest jednak zebranie danych kontaktowych (przede wszystkim aktualnych adresów poczty elektronicznej) osób i instytucji biorących udział w programach podlegających ewaluacji. Jest to często robione. Warto jednak mając na uwadze późniejsze prowadzenie ewaluacji przykładać większą wagę do zbierania tych informacji kontaktowych.

W przypadku badań ewaluacyjnych stosunkowo często jednostkami doboru próby są nie pojedyncze osoby, lecz organizacje i instytucje. W przeciwieństwie do osób indywidualnych, wśród których dopiero 51% korzysta z Internetu⁶, prawie wszystkie firmy i organizacje posiadają dostęp do Internetu i adres mailowy. Według badań GUS z 2008 roku, 93% firm zatrudniających przynajmniej 10 osób ma dostęp do Internetu⁷. W mniejszych firmach jest nieco gorzej, jednak badania *Diagnozy Społecznej* z marca 2009 pokazują, iż ponad 80% prywatnych przedsiębiorców (w większości są to po prostu osoby prowadzące działalność gospodarczą) korzysta z Internetu. Tym samym prowadzenie badań CAWI na takich populacjach ma znacznie większy sens.

W badaniach internetowych koszty krańcowe dotarcia do kolejnych respondentów są często znikome. Z tego powodu organizatorzy często decydują się na badanie całej populacji, zamiast doboru próby. Takie działania nie są jednak zalecane, gdyż dobrze dobrana próba daje możliwość otrzymania wyników nieróżniących się znacząco od wyników dla całej populacji. Jednocześnie w przypadku prowadzenia większej liczby badań na tej samej populacji zaangażowanie w badania mniejszej liczby osób sprawia, że członkowie populacji będą się czuli mniej znużeni ciągłymi prośbami o udział w badaniach.

Narzędzia badawcze a realizacja próby

Cechy narzędzia badawczego mają wpływ przede wszystkim na skłonność do wypełniania ankiety przez osoby, które ją rozpoczęły (tzw. *retention rate*, czyli odsetek respondentów, którzy dokończyli

⁴ K. Rakocy, J.M. Zając, Czy wierzysz w słowa te? Wiarygodność odpowiedzi uzyskiwanych w ankietach internetowych. Poster przedstawiony na Ogólnopolskiej Konferencji Naukowej Społeczne Aspekty Internetu SAI'06, Warszawa, 8–10.10.2006.

⁵ W praktyce lepsza sytuacja niż w badaniach ewaluacyjnych może występować jedynie w badaniach organizacyjnych, szczególnie w organizacjach i instytucjach, w których wszyscy pracownicy dysponują skrzynką poczty elektronicznej.

⁶ D. Batorski, *Korzystanie z technologii informacyjno-komunikacyjnych*, [w:] J. Czapiński, T. Panek, *Diagnoza społeczna 2009. Warunki i jakość życia Polaków*, s. 279–307. Warszawa 2009.

⁷ *Wykorzystanie technologii informacyjno-telekomunikacyjnych w przedsiębiorstwach, gospodarstwach domowych i przez osoby prywatne w 2008 r.*, GUS 2008.

badanie w stosunku do tych, którzy je rozpoczęli; przerywanie badania w jego trakcie to, jak już wspominaliśmy, *dropout*). W mniejszym stopniu oddziałują na decyzję o rozpoczęciu badania, zwłaszcza że najczęściej badany nie może obejrzeć narzędzia przed podjęciem decyzji o udziale. Cechy narzędzia badawczego mają ponadto duże znaczenie ze względu na błąd pomiaru, ale to zagadnienie wykracza poza ramy niniejszego artykułu.

Kontrola dostępu do badania

Istotnym elementem badań internetowych jest kontrola dostępu do badania. Stosuje się ją przede wszystkim w celu zapewnienia, że jedna osoba nie weźmie udziału w badaniu więcej niż jeden raz, a także dla kontroli, kto z próby wziął już udział w badaniu, a kto jeszcze nie. Jeśli zaproszenie do badania wysyłane jest e-mailem, prostym sposobem kontroli dostępu jest przydzielenie każdemu badanemu indywidualnego linka do badania, przez niektórych autorów określane jako automatyczne logowanie⁸. Zaletą tej metody jest łatwość dotarcia badanego do ankiety, jako że musi on jedynie kliknąć w link. Przydzielanie indywidualnych linków jest coraz częściej jedną z funkcji programów do obsługi badań internetowych, dzięki czemu taką wysyłkę można przeprowadzić automatycznie. Ten sposób kontroli dostępu wydaje się obecnie najczęstszy w praktyce badawczej.

Innym rozwiązaniem jest wymaganie ręcznego logowania na stronę z badaniem, przy użyciu oddzielnych haseł dostępu i ewentualnie także loginów dla każdego badanego. Dla badanych oznacza to więcej zachodu w związku z wpisywaniem kodów dostępu, ale zarazem od początku wymusza się nieco większe zaangażowanie i daje się poczucie poufności. Wyniki badań pokazują, że faktycznie w porównaniu z automatycznym logowaniem stopień realizacji próby jest niższy. Zarazem uzyskuje się dane o wyższej jakości, o czym świadczy mniejszy odsetek rezygnacji z udziału w trakcie badania, dłuższy namysł nad pytaniami i dłuższe odpowiedzi na pytania otwarte⁹. Podobne rozwiązanie polega na uzupełnieniu automatycznego logowania na stronę z użyciem indywidualnego URL ręcznym podawaniem prostego hasła, np. czterocyfrowego numeru PIN¹⁰. Taki sposób kontroli dostępu dał w cytowanym badaniu najwyższy stopień realizacji próby, a jednocześnie niewiele niższą jakość danych niż przy logowaniu ręcznym. Warto jednak zachować ostrożność w stosowaniu haseł i kodów dostępu. Mogą one nastręczać trudności dla mało doświadczonych użytkowników Internetu.

Ręczne, półautomatyczne czy zwłaszcza automatyczne logowanie do badania to skuteczne sposoby kontroli dostępu do badania, które jednocześnie nie powinny znacząco obniżać realizacji próby. Kontrola dostępu do badania przez pliki *cookies* czy tym bardziej numery IP jest niezalecana ze względu na mniejszą dokładność. Dodatkowo może rodzić dalsze problemy, a w skrajnych przypadkach prowadzić do skrzywienia wyników.

Długość ankiety

Jedną z najważniejszych właściwości ankiety internetowej jest jej długość. Ma ona znaczenie dla realizacji próby, przede wszystkim dla przerywania udziału w badaniu. Pośrednio może wpływać również na skłonność do udziału w badaniu, jeżeli w treści zaproszenia podajemy orientacyjny czas potrzebny na

⁸ S.D. Crawford, M.P. Couper, M.J. Lamias, *Web Surveys: Perceptions of Burden*, „Social Science Computer Review”, 19, 2001, s. 146–162; D. Heerwegh, G. Loosveldt, *An evaluation of the semiautomatic login procedure to control web survey access*, „Social Science Computer Review”, 21 (2), 2003, s. 223–234.

⁹ S.D. Crawford, M.P. Couper, M.J. Lamias, *ibidem*; D. Heerwegh, G. Loosveldt, *Web surveys: The effect of controlling survey access using PIN numbers*, „Social Science Computer Review”, 20, 2002, s. 10–21; D. Heerwegh, G. Loosveldt, *An evaluation of the semiautomatic login procedure to control web survey access*, „Social Science Computer Review”, 21 (2), 2003, s. 223–234.

¹⁰ D. Heerwegh, G. Loosveldt, *An evaluation of the semiautomatic...*, *ibidem*.

udzielenie wszystkich odpowiedzi¹¹. Generalnie badania prowadzone w Internecie powinny być krótkie, odstępstwa od tej zasady są możliwe, gdy badania skierowane są do specyficznej, dobrze zmotywowanej populacji. Przy długich ankietach wskazane jest umożliwienie zapisania badania i powrotu do wypełniania ankiety później. W takim przypadku warto też wysłać przypomnienia osobom, które nie dokończyły ankiety. Innym rozwiązaniem jest podział długich badań na kilka części i wysyłanie oddzielnych zaproszeń do nich w kolejnych dniach¹². W tym ostatnim przypadku ważne jest oczywiście zachowanie możliwości identyfikacji respondenta pomiędzy badaniami.

Na znacznie dłuższe kwestionariusze można sobie pozwolić w badaniach instytucji czy organizacji. Jednak również w przypadku, gdy jednostkami w badaniu ewaluacyjnym są firmy czy organizacje, a nie osoby indywidualne, warto zamiast jednej długiej ankiety dla całej instytucji przygotować kilka krótszych ankiet, które dodatkowo mogą być wypełniane przez różnych członków organizacji lepiej zorientowanych w poszczególnych zagadnieniach objętych badaniem.

Specyfika badań CAWI i duża wrażliwość respondentów na długość ankiety sprawiają, że lepiej sprawdzają się badania krótkie. Dlatego też lepiej jest badać jedno zagadnienie w jednym badaniu, organizując tym samym kilka oddzielnych badań, niż badać wiele różnych zagadnień w ramach jednego badania. Organizacja kilku osobnych badań może być nieco bardziej skomplikowana, a z drugiej strony koszty kontaktów z badanymi czy tworzenia kilku wersji narzędzia badawczego mają w badaniach CAWI bardzo małe znaczenie.

Czynnikiem wpływającym na wrażenie dłuższej ankiety jest podział ankiety na wiele stron. W przypadku krótkich i prostych ankiet wskazane jest umieszczenie ich na jednej stronie, ponieważ brak widoczności wszystkich pytań od razu sprzyja przerywaniu udziału w badaniu¹³. Zarazem w wielu sytuacjach dzielenie kwestionariusza jest jednak niezbędne: przy bardzo długich ankietach; przy wykorzystywaniu pytań filtrujących (choć niektóre narzędzia umożliwiają filtrowanie w obrębie jednej strony); gdy nie chcemy, aby badani widzieli kolejne pytania przed odpowiedzią na poprzednie; gdy chcemy dokładnie szacować, w którym momencie następuje *dropout*.

W przypadku dłuższych ankiet i podziału na podstrony wskazanym rozwiązaniem jest wskaźnik postępu (*progress indicator*)¹⁴. Może on przybierać różne formy, od najprostszych wskaźników typu „Strona 3 z 7”, przez wskaźniki procentowe i graficzne, aż po wskaźniki dynamiczne, dostosowujące się do szybkości odpowiedzi badanego i odpowiedzi na pytania filtrujące. Można też stosować wskaźniki postępu, które na początku ankiety sugerują badanemu, że skończy ją znacznie szybciej niż w rzeczywistości. Jak jednak pokazują wyniki Kaczmirka¹⁵, powoduje to tak naprawdę większą realizację w trakcie badania niż rzetelne wskaźniki postępu.

Istotne znaczenie, zwłaszcza przy dłuższych badaniach, ma kolejność pytań. Warto zacząć badanie od pytań łatwych, a zarazem interesujących. Pytania metryczkowe, które zazwyczaj są dla respondentów najmniej interesujące, warto umieścić na końcu kwestionariusza, z wyjątkiem sytuacji, kiedy właśnie cechy socjodemograficzne są w badaniu najważniejsze¹⁶, bo umieszczanie ich na końcu wiąże się z częst-

¹¹ S.D. Crawford, M.P. Couper, M.J. Lamias, *op. cit.*

¹² I.S. Choi, *Online research in Korea: Reliability and validity issues*. Wystąpienie na konferencji The 1st International Workshop on Internet Survey, Deajeon, Korea, 9–10.09.2009.

¹³ D. Batorski, *Sieci społeczne a Internet: Sieciowy indywidualizm czy nowe wspólnoty*. (Raport KBN 1-H02E-029-27). Warszawa: Uniwersytet Warszawski, Instytut Studiów Społecznych, 2005.

¹⁴ S.D. Crawford, M.P. Couper, M.J. Lamias, *op. cit.*; L. Kaczmirek, *Human-Survey Interaction: Usability and Nonresponse in Online Surveys*. Cologne: Herbert von Halem Verlag (w druku).

¹⁵ Ibidem.

¹⁶ D.A. Dillman, R.D. Tortora, D. Bowker, *Principles for Constructing Web Surveys*. (SESRC Technical Report 98-50). Pullman, Washington, 1998.

szym ich pomijaniem przez respondentów. Z drugiej strony, jak pokazały badania Reipsa¹⁷, zadanie pytań metryczkowych na początku ankiety zwiększa motywację badanych – odpowiadając na pytania o płeć, wiek itp. badani zmniejszają swoje poczucie anonimowości, a będąc mniej anonimowi czują się bardziej zobowiązani do udziału w badaniu i rzadziej przerywają udział przed skończeniem ankiety. Pozornie wyniki tych różnych badań mogą wydawać się sprzeczne. Należy się jednak spodziewać, że inne są efekty długich i nudnych pytań metryczkowych, które mogą zniechęcać do udziału, jeśli zostaną umieszczone na początku kwestionariusza, a inne 3–4 krótkich pytań o najogólniejsze dane.

Format odpowiedzi

Pewne znaczenie dla udziału respondentów w badaniu ma format odpowiedzi. Niektóre wyniki badań¹⁸ dowodzą, że respondenci preferują widoczne listy (*radio buttons*) od rozwijanych menu (*drop-down menu*). Taki format jest po prostu wygodniejszy, zwłaszcza w przypadku krótkich list (kilka elementów). W przypadku populacji aktywnych użytkowników Internetu, które na co dzień stykają się z tymi formatami przy wypełnianiu różnych formularzy, format może mieć mniejsze znaczenie. Problematiczne dla badanych mogą być też pytania otwarte, zwłaszcza tak zwane narracyjne. Odpowiedź na nie wymaga więcej wysiłku niż zaznaczenie jednej z podanych odpowiedzi w pytaniach zamkniętych. Respondenci mogą wówczas rezygnować z badania w ogóle lub nie odpowiadać na te pytanie (*item non-response*).

W ankietach internetowych stosunkowo proste jest uniemożliwienie przejścia do kolejnego pytania bez udzielenia odpowiedzi na pytania poprzednie. Takie rozwiązanie zapewnia, że wszystkie skończone ankiety będą miały komplet odpowiedzi. Niemniej jednak często wpływa to równocześnie na zwiększenie *dropout'u*, zwłaszcza przy pytaniach drażliwych, na przykład o dochody. Warto w takich sytuacjach po prostu nie zmuszać do udzielenia odpowiedzi lub dodać odpowiedź „nie wiem / odmowa odpowiedzi”.

Aspekty techniczne kwestionariusza

Należy unikać rozwiązań technicznych, które mogą uniemożliwić bądź utrudnić części badanych udział w badaniu. Przy projektowaniu narzędzia badawczego warto zastanowić się nad sprzętem i oprogramowaniem używanym przez badaną próbę¹⁹. Generalnie najlepiej sprawdzają się rozwiązania najprostsze. Nawet tak popularne technologie jak JavaScript, Java czy Flash nie są używane przez wszystkich internautów. Większe problemy mogą wystąpić w przypadku odtwarzania materiałów wideo. Wydaje się, że Flash jest znacznie bardziej popularny niż jakikolwiek player wideo (np. Windows Media Player czy RealPlayer). Warto także sprawdzić, czy narzędzie badawcze dobrze wyświetla się pod różnymi przeglądarkami (przynajmniej Internet Explorer i Firefox) oraz przy różnych rozdzielczościach ekranu. Statystyki popularności różnych rozwiązań tego rodzaju można znaleźć na przykład na stronie www.ranking.pl.

Problemów technicznych nie rozwiązuje spytanie przed rozpoczęciem badania o używany sprzęt i oprogramowanie, gdyż część badanych może tego nie wiedzieć, ani umożliwienie ściągnięcia potrzebnego oprogramowania, gdyż część respondentów może nie umieć go zainstalować lub nie mieć niezbędnych uprawnień. Możliwe jest natomiast, aby strona internetowa z ankietą sprawdzała pewne parametry techniczne, jak chociażby używana przeglądarka, fakt zainstalowania Flash lub Java. Tym samym możliwe jest uzależnienie wyświetlania różnych wersji strony od niektórych parametrów technicznych

¹⁷ U.-D. Reips, *Standards for Internet-based experimenting*, „Experimental Psychology”, 49 (4), 2002, s. 243–256.

¹⁸ Np. D.A. Dillman, D.K. Bowker, *The Web questionnaire challenge to survey methodologists*, [w:] U.-D. Reips i M. Bosnjak (red.), *Dimensions of Internet science*, (s. 159–178). Lengerich: Pabst Science Publishing, 2001..

¹⁹ L. Kaczmirek, ibidem.

komputera czy przeglądarki respondenta. Z drugiej strony, w niektórych badaniach podaje się po prostu minimalne wymagania, jak na przykład zainstalowanie jakiegoś dodatkowego oprogramowania, podając link, pod którym użytkownicy mogą dane oprogramowanie pobrać.

Pewne znaczenie ma także szybkość ładowania się stron²⁰. Gdy strony badania ładują się powoli, badany może zniechęcać się i rezygnować z udziału. Wprawdzie wraz z rozwojem technologii i upowszechnianiem szerokopasmowego dostępu do Internetu te cechy tracą na znaczeniu, niemniej jednak warto unikać elementów, które są zbyt „ciężkie” oraz upewnić się, czy duża liczba jednoczesnych wejść na stronę nie spowolni serwera.

Narzędzie do zarządzania badaniem

Warto w tym miejscu zwrócić uwagę również na narzędzia zarządzania badaniem, które mogą znacznie ułatwić badaczowi pracę. Narzędzie powinno umożliwiać uproszczoną wysyłkę zaproszeń i przypomnień do respondentów, najlepiej z możliwością automatycznego personalizowania treści korespondencji, w tym również indywidualnych linków umożliwiających automatyczne logowanie do badania. Korzystny jest także monitoring realizacji badania w jego trakcie, śledzenie *dropout'u* na poszczególnych pytaniach, gdyż zdarza się, że pytania wymuszające odpowiedź mogą generować duży wzrost przerywania badania. Powinno być możliwe również sprawdzenie tego, kto z badanych jeszcze nie wziął udziału w badaniu, a kto zaczął, ale nie dokończył wypełniania ankiety, oraz kierowanie do tych osób ponownych zaproszeń i przypomnień. Wygodna jest również możliwość generowania prostych raportów w trakcie badania.

Czas przeprowadzania badania i wysłania zaproszeń

Podczas planowania badania należy uwzględnić rytmy aktywności badanej populacji. Wysłanie zaproszenia do badania w nieodpowiednim momencie zmniejsza szanse, że respondent zareaguje na nie, lub w ogóle je otrzyma. Jeśli tylko istnieje taka możliwość, badanie powinno być przeprowadzone w momencie, gdy będą największe szanse, że zaproszenie dotrze do respondentów w najbardziej sprzyjającym momencie. Mniej efektywne są badania przeprowadzane w sezonie urlopowym, w czasie ferii, wakacji itp., gdyż potencjalni respondenci rzadziej lub w ogóle nie korzystają wówczas z Internetu²¹, zaś w kontekście organizacyjnym – w okresie szczególnie intensywnej pracy.

Co więcej, badania różnych populacji potwierdzają, że gros respondentów bierze udział w badaniach internetowych wkrótce po odebraniu zaproszenia²². Analogicznie, można się spodziewać krótkotrwałych wzrostów w liczbie wypełnień wkrótce po wysłaniu przypomnień²³. Należy to brać pod uwagę, gdyż w przypadku zastosowania w badaniu niedoskonałych technologii zbyt duża liczba jednoczesnych wejść na strony badania może spowodować problemy techniczne, w tym również uniemożliwić realizację badania.

²⁰ D.A. Dillman, R.D. Tortora, D. Bowker, *ibidem*; U-D. Reips, *ibidem*.

²¹ S.D. Crawford, M.P. Couper, M.J. Lamias, *ibidem*; S.J. Sills, Ch. Song, *Innovations in survey research: An application of Web-based surveys*. „Social Science Computer Review”, 20, 2002, s. 22–30.

²² P. Mazurek, J.M. Zając, K. Rakocy, *Online digital surveillance: What they know about you depends on what you know*. Wystąpienie na konferencji General Online Research 07, Lipsk, 26–28.03.2007; J.M. Zając, *Nieważne jak pytasz: Cechy zaproszenia a skłonność do udziału w ankiecie internetowej*, [w:] M. Sokołowski (red.), *Oblicza internetu: Internet w przestrzeni komunikacyjnej XXI wieku*, Wydawnictwo PWSZ w Elblągu, 2006, s. 167–178.

²³ S.D. Crawford, M.P. Couper, M.J. Lamias, *ibidem*.

Warto wziąć pod uwagę, że wielu użytkowników Internetu jest przeciążonych zalewem napływających do nich informacji. Jeśli nie zareagują szybko po otrzymaniu zaproszenia, dość prawdopodobne jest, że nie będą o nim pamiętać i nie zareagują w ogóle. Goeritz i Steiger²⁴ sugerują, że w badaniach na zarejestrowanych uczestnikach panelu badawczego w ciągu pierwszych 3 dni badania wpływa 70–90% wszystkich odpowiedzi. Hamilton²⁵ w swojej metaanalizie dowodzi, że ponad połowa respondentów brała udział w ankietach w ciągu 24 godzin. Podkreśla też, że najbardziej efektywne były zaproszenia wysyłane pomiędzy 6 a 9 rano w dni robocze. W innym badaniu²⁶ najbardziej efektywne okazały się zaproszenia wysyłane rano (między 7.30 a 8.30) i pod koniec dnia roboczego (15.30–16.30). Te wyniki wydają się istotne, gdy badani biorą udział w badaniu w pracy. Jeśli jednak spodziewamy się, że uczynią to w czasie wolnym, skuteczniejsze mogą być zaproszenia wysyłane w godzinach popołudniowych, tak aby badani przeczytali je w momencie, gdy siadają przed komputerem po powrocie z pracy czy szkoły.

Długość czasu zbierania danych (*field time*) również ma znaczenie. Zazwyczaj wystarczy zapewne stosunkowo krótki okres (do 2 tygodni), choć zarazem ze względu na różne tygodniowe rytmy aktywności warto, żeby badanie trwało co najmniej tydzień. Powodem wydłużenia tego okresu może być ograniczona dostępność badanej populacji lub jej części w okresie badania lub też chęć zastosowania innych kanałów dotarcia do badanych w ramach schematów mieszanych, zwłaszcza poczty. Jeśli chodzi o podawanie lub nie terminu zakończenia badania w zaproszeniu oraz o to, czy termin ten powinien być krótki czy relatywnie dłuższy, wyniki dotychczasowych badań nie są konkluzywne²⁷. Wydaje się jednak, że informację o dacie zakończenia zbierania danych warto umieścić ze względów organizacyjnych.

Skłonność do szybkiej reakcji na zaproszenie bywa także wykorzystywana w sytuacji bardzo szybkich badań, gdy wyniki są potrzebne w ciągu 2–3 dni. Zakłada się wówczas, że znaczne skrócenie czasu zbierania danych spowoduje niewielki spadek realizacji próby, gdyż i tak gros odpowiedzi zebrano by wkrótce po rozpoczęciu. Oczywiście takie rozwiązanie zmniejsza reprezentatywność oraz uniemożliwia stosowanie niektórych technik zwiększania realizacji próby.

Zaproszenie do udziału w badaniu

Znaczenie zaproszenia do badania dla realizacji próby jest od dawna znane metodologom. Odpowiedni sposób kontaktu z respondentami przyciąga ich uwagę i motywuje do udziału. W ankietach internetowych zaproszenie jest szczególnie ważne, gdyż w odróżnieniu od na przykład sondaży pocztowych badani zazwyczaj nie oglądają narzędzia badawczego przed podjęciem decyzji o udziale²⁸. Ma to znaczenie zwłaszcza w sytuacji, gdy kolejne części kwestionariusza są umieszczone na różnych stronach internetowych oraz w przypadku kontroli dostępu do ankiety (np. hasłem). Co więcej, w Internecie zaproszenia do badania powinny być krótkie i zwięzłe. Optymalizacja zaproszenia wydaje się stosunkowo prostą metodą zwiększania motywacji badanych w badaniach internetowych, które dodatkowo nie wymaga od badacza szczególnych nakładów pieniężnych, jak na przykład gratyfikacja.

²⁴ A.S. Goeritz, S. Steiger, *The impact of the field time on response, retention, and response completeness in list-based Web surveys*, „International Journal of Human Computer Studies”, 67, 2009, s. 342–348.

²⁵ M.B. Hamilton, *Online Survey Response Rates and Times: Background and Guidance for Industry*. SuperSurvey White Paper, 2003.

²⁶ Granello i Wheaton, 2004, za: V. M. Sue i L. A. Ritter, *Conducting online surveys*. Sage, 2007.

²⁷ A.S. Goeritz, S. Steiger, *The impact of the field time on response, retention, and response completeness in list-based Web surveys*, „International Journal of Human Computer Studies”, 67, 2009, s. 342–348; S.R. Porter, M.E. Whitcomb, *The impact of contact type on Web survey response rates*, „Public Opinion Quarterly”, 67, 2003a, s. 579–588.

²⁸ S.D. Crawford, M.P. Couper, M.J. Lamias, *ibidem*.

Kontakty poprzedzające badanie

Kontakty z respondentami przed rozpoczęciem właściwego badania (*pre-notification*) przygotowują do udziału i ułatwiają późniejszą komunikację. Mają przede wszystkim funkcję informacyjną, jednak mogą się również przyczyniać do zwiększenia realizacji próby. Na tym etapie można sprawdzić jakość listy mailingowej czy innego stosowanego sposobu kontaktu. Jeśli badanie dotyczy organizacji, można umożliwić jej przygotowanie się do badania, na przykład przez wyznaczenie odpowiednich osób czy zebranie potrzebnych informacji. Można także na tym etapie próbować techniki wpływu społecznego zwanej „stopa w drzwiach”, czyli w tym przypadku poprzedzania właściwego badania drobną prośbą²⁹. Zgodnie z mechanizmem wpływu opisanym przez Cialdiniego³⁰ badani angażujący się w spełnienie małej prośby są później bardziej skłonni do spełnienia także większej. Należy jednak uważać, aby takie działania nie obciążały wyników właściwego badania.

Sposób dostarczenia zaproszenia

Wybór formy komunikacji wykorzystanej do wysłania zaproszenia zależy w dużym stopniu od dostępnych informacji na temat badanej próby. Obecnie najczęściej stosuje się rozesłanie zaproszeń przez pocztę elektroniczną, która od wielu lat jest zdecydowanie najpopularniejszą formą komunikacji w sieci. Według badania Diagnoza Społeczna 2009 korzysta z niej około 90% polskich internautów powyżej 16 lat, czyli mniej więcej 45% dorosłych osób. W analogiczny sposób można wysłać zaproszenia przez komunikator internetowy³¹.

Jeśli organizator badania nie zna tego rodzaju danych kontaktowych lub też chce przeprowadzić badanie na próbie ochotników, można umieścić zaproszenia do badania na stronach WWW odwiedzanych przez badaną populację. Zaproszenia można dodatkowo wyróżnić graficznie, warto jednak, aby nie przypominały za bardzo form reklamowych. Zaproszenie przez WWW jest często stosowane, gdy organizator badania nie ma danych umożliwiających kontakt z respondentami poprzez e-mail czy poza Internetem. Niestety to rozwiązanie nie daje możliwości kontroli nad respondentami i precyzyjnego określenia stopnia realizacji próby.

Innym rozwiązaniem jest wyświetlanie zaproszeń w wyskakujących okienkach (tzw. *pop-up*), wyświetlających się podczas wchodzenia na daną stronę. Umożliwia to losowy dobór respondentów spośród użytkowników strony na podstawie plików *cookies*, w sytuacji gdy nie ma innej listy bazowej. Ta metoda może mieć zastosowanie zwłaszcza wtedy, gdy istnieje możliwość wyświetlania zaproszeń na stronach sprofilowanych tematycznie. Zarazem jednak ma istotne ograniczenia. Wyskakujące okienka przypominają nie ulubiane reklamy, a część użytkowników używa blokad form tego rodzaju w przeglądarkach, stąd też maleją szanse na zwrócenie uwagi na zaproszenie. Co więcej, zamknięcie takiego okienka powoduje, że powrót do badania w bardziej dogodnym momencie staje się niemożliwy. Dlatego też, o ile jest to możliwe, należy starać się unikać umieszczania ankiet internetowych lub zaproszeń do nich w wyskakujących okienkach.

Do zapraszania do udziału i przypominania o badaniu można również wykorzystywać pozainternetowe środki komunikacji – pocztę, telefon czy spotkania bezpośrednie. Zwłaszcza przypomnienia telefoniczne mogą być skuteczne, gdyż równocześnie można na bieżąco pomóc respondentowi np. w przypadku problemów technicznych. Warto odpowiednio wcześniej zaplanować sposób kontaktu

²⁹ N. Guéguen, *Foot-in-the-door technique and computer-mediated communication*, „Computers in Human Behavior”, 18, 2002, s. 11–15.

³⁰ R. Cialdini, *Wwiercanie wpływu na ludzi*. Gdańsk: GWP, 1996.

³¹ D. Batorski, *ibidem*.

z badanymi. Dużą zaletą badań ewaluacyjnych jest to, że wiele programów podlegających ewaluacji pozwala na zbieranie w trakcie ich trwania danych kontaktowych uczestników.

Nadawca zaproszenia

W przypadku badań realizowanych poza Internetem informacje o organizatorach, i czasem też sponsorach, badania wykorzystuje się do uwiarygodnienia badania i skłonienia respondentów do udziału³². Nadawca, który jest bardziej wiarygodny i cieszy się autorytetem, powinien być bardziej przekonujący dla respondentów. W przypadku badań internetowych takie techniki nie zawsze są skuteczne. Można wskazać kilka powodów: stosowanie podobnych sposobów uwiarygodnienia i przyciągnięcia uwagi przez wysyłających niechcianą pocztę (*spam*), przez co wielu internautów nie ufa takim manipulacjom³³; specyfikę zachowania i poczucie anonimowości w Internecie³⁴, powodujące, że internauci mogą być mniej skłonni do posłuszeństwa wobec autorytetów; wreszcie fakt, że internauci zazwyczaj nie czytają dokładnie otrzymywanych komunikatów, a tylko je przeglądają.

Wbrew pozorom, nie zawsze nadawca o wyższym statusie społecznym i autorytecie wpłynie na wyższą realizację próby. Wprawdzie w eksperymencie Guéguena i Jacob³⁵ profesor uniwersytetu uzyskał znacznie wyższy *response rate* niż student, ale w badaniach Portera i Whitcomba³⁶ oraz Zająca³⁷ manipulacja różnymi cechami nadawcy nie miała żadnego wpływu na skłonność respondentów do udziału w badaniu. Wydaje się natomiast, że skuteczne jest odwołanie się do władzy organizatora i sponsora badania nad badanymi³⁸. Wynikałoby z tego, że skuteczniejsze będą zaproszenia firmowane na przykład przez szefów firmy czy dyrektorów programu obejmującego wiele przedsiębiorstw. Można również rozważyć możliwość wcześniejszego zobowiązania uczestników takich programów do późniejszego udziału w badaniach ewaluacyjnych.

Skuteczność zaproszenia mogą zwiększać dodatkowe informacje o nadawcy, np. zdjęcie oraz manipulacje tożsamością nadawcy, ale możliwość stosowania tych technik wydaje się ograniczona. W badaniach potwierdzono większy stopień realizacji próby po dodaniu zdjęcia do zaproszenia w e-mailu³⁹ i w wyskakującym okienku⁴⁰. Wydaje się jednak, że duże znaczenie mogą mieć atrakcyjność osoby na zdjęciu, jej płeć i płeć respondenta oraz interakcje między tymi czynnikami, co jednak wymaga dalszych badań. W innym interesującym badaniu przeprowadzonym przez Agnieszkę Janiszewską z firmy Gemius⁴¹ zauważono, że nowe twarze w zaproszeniu typu *pop-up* przyciągają uwagę bardziej niż twarz wykorzystywana w tym celu wielokrotnie wcześniej i z tego powodu dobrze znana respondentom. Wresz-

³² R.M. Groves, F.L. Fowler jr., M.P. Couper, J.M. Lepkowski, E. Singer, R. Tourangeau, *Survey methodology*, Wiley, 2004.

³³ S.R. Porter, M.E. Whitcomb, *E-mail subject lines and their effect on web survey viewing and response*, „Social Science Computer Review”, 23 (3), 2005, s. 380–387.

³⁴ A.N. Joinson, *Internet behaviour and the design of virtual methods*, [w:] C. Hine (red.), *Virtual Methods: issues in social research on the Internet*, Oxford: Berg, 2005.

³⁵ N. Guéguen, C. Jacob, *Solicitation by e-mail and solicitor's status: A field study of social influence on the Web*, „CyberPsychology & Behavior”, 5, 2002, s. 377–383.

³⁶ S.R. Porter, M.E. Whitcomb, *The impact of contact type on Web survey response rates*, „Public Opinion Quarterly”, 67, (2003a), s. 579–588.

³⁷ J.M. Zając, *Nieważne jak pytasz: Cechy zaproszenia a skłonność do udziału w ankiecie internetowej*, [w:] M. Sokołowski (red.), *Oblicza internetu: Internet w przestrzeni komunikacyjnej XXI wieku*, Wydawnictwo PWSZ w Elblągu, 2006, s. 167–178.

³⁸ A.N. Joinson, U-D. Reips, *Personalized salutation, power of sender and online survey response rates to web-based surveys*, „Computers in Human Behavior”, 23, 2007, s. 1372–1383; A.N. Joinson, A. Woodley, U-D. Reips, *Personalization, authentication and self-disclosure in self-administered Internet surveys*, „Computers in Human Behavior”, 23, 2007, s. 275–285.

³⁹ N. Guéguen, C. Jacob, *Social presence reinforcement and computer-mediated communication: The effect of the solicitor's photography on compliance to a survey request made by e-mail*, „CyberPsychology & Behavior”, 5, 2002a, s. 139–142.

⁴⁰ J. Przewłocka, P. Janczewski, *Jake-Kowalski uczył nas stawiać pytania, czyli rola formatu i wyglądu pytań w badaniach online*. Wystąpienie na VIII Kongresie PTBRIO, Warszawa 18–19.10.2007.

⁴¹ A. Janiszewska, *Działania prowadzone przez Gemius na rzecz podnoszenia response rate w badaniach online*. Research paper. Warszawa: Gemius SA, 2009.

cie, Althoff, Greif, Griel i Batinic⁴² zasugerowali, że skuteczniejsze są zaproszenia wysłane przez nadawcę płci żeńskiej (tzw. efekt Anity), zastrzegając, że dotyczy to sytuacji, gdy próba zdominowana jest przez mężczyzn, a badanie prowadzone jest na grupie osób intensywnie używających poczty elektronicznej, które trzeba przekonać do poświęcenia czasu na udział w badaniu. Użyteczność powyższych wskazówek jest niewielka, przede wszystkim ze względu na ich silne uzależnienie od kontekstu badania. Ponadto podawanie nieprawdziwych informacji na temat organizatora i wykonawcy badania będzie w większości sytuacji nieetyczne.

Sklonność do udziału w badaniu jest wyraźnie większa w sytuacji, gdy osoba przeprowadzająca badanie lub nadawca zaproszenia jest członkiem badanej zbiorowości, zwłaszcza jeżeli ta zbiorowość ma wspólną tożsamość. Fakt ten wynikać może z większej skłonności do działania na rzecz członków wspólnoty. Jeśli nawet badacz sam nie należy do niej, może mu pomóc zaangażowanie osoby, która dla tej wspólnoty jest autorytetem.

Treść zaproszenia

Odpowiednie sformułowanie komunikatu informującego o badaniu może pomóc w zwiększeniu realizacji próby. Działające mechanizmy można wyjaśniać odwołując się do procesów wpływu społecznego. Trzeba jednak uważać, aby nie przesadzić z technikami tego rodzaju, zwłaszcza jeśli badanie dotyczy tematów poufnych i drażliwych, gdyż wówczas sytuacja badania będzie miała w zbyt dużym stopniu charakter społeczny⁴³.

W większości sytuacji sposobem zwiększenia realizacji próby bez szkody dla jakości danych może być personalizacja zaproszenia. Najczęściej stosowanym sposobem jest zwracanie się do respondenta imieniem bądź imieniem i nazwiskiem (oczywiście badacz musi znać te dane w momencie wysyłki zaproszeń!). Większość badań potwierdza skuteczność tej metody⁴⁴, wyjątek stanowią badania Portera i Whitcomba⁴⁵. Joinson i Reips⁴⁶ na podstawie wyników serii eksperymentów zasugerowali, że brak wpływu spersonalizowanego przywitania na stopień realizacji próby w pewnych przypadkach wynika z jednoczesnego zmniejszania się wrażenia anonimowości respondenta.

Prostym sposobem personalizacji zaproszenia jest również wysyłka indywidualnych wiadomości do każdego badanego zamiast e-maili wysyłanych jednocześnie do całej grupy (nawet na zasadzie ukrytej kopii), co sprawia na badanych wrażenie prośby bardziej indywidualnej⁴⁷. Można także próbować innych sposobów odwoływania się do wyjątkowości badania i reguły niedostępności⁴⁸, ale ich skuteczność może być ograniczona. W badaniu Portera i Whitcomba⁴⁹ informacja typu „jesteś jedną z niewielu osób uczestniczących w tym badaniu” nie miała wpływu na realizację próby.

⁴² Althoff S., Greif, V., Griel, B., Batinic, B. (2006). *Determinants of response rates of online surveys - the Anita effect: Results of a joint project*. Wystąpienie na konferencji General Online Research 06, Bielefeld.

⁴³ A.N. Joinson, *Internet behaviour and the design of virtual methods*, [w:] C. Hine (red.), *Virtual Methods: issues in social research on the Internet*. Oxford: Berg, 2005.

⁴⁴ D. Heerwegh, *Effects of personal salutations in e-mail invitations to participate in a Web survey*, „Public Opinion Quarterly”, 69 (4), 2005, s. 588-598; D. Heerwegh, T. Vanhove, K. Matthijs, G. Loosveldt, *The effect of personalization on response rates and data quality in web surveys*, „International Journal of Social Research Methodology: Theory and Practice”, 8, 2005, s. 85-99.

⁴⁵ S.R. Porter, M.E. Whitcomb, *The impact of contact type on Web survey response rates*, „Public Opinion Quarterly”, 67, 2003a, s. 579-588.

⁴⁶ A.N. Joinson, U-D. Reips, *Personalized salutation, power of sender and online survey response rates to web-based surveys*, „Computers in Human Behavior”, 23, 2007, s. 1372-1383; A.N. Joinson, A. Woodley, U-D. Reips, *Personalization, authentication and self-disclosure in self-administered Internet surveys*, „Computers in Human Behavior”, 23, 2007, s. 275-285.

⁴⁷ D. Heerwegh, G. Loosveldt, *Web surveys: The effect of controlling survey access using PIN numbers*, „Social Science Computer Review”, 20, 2002, s. 10-21.

⁴⁸ Por. R. Cialdini, *Wywieranie wpływu na ludzi*. Gdańsk: GWP, 1996.

⁴⁹ S.R. Porter, M. E. Whitcomb, *The impact of contact type on Web survey response rates*, „Public Opinion Quarterly”, 67, 2003a, s. 579-588.

Z punktu widzenia technik wpływu społecznego nie dziwi, że pomocne jest sformułowanie zaproszenia jako prośby⁵⁰. Ważne jest także umieszczanie od początku informacji o organizatorach badania i możliwości kontaktu z badaczami. Niezależnie od względów realizacji próby jest to zresztą wymagane przez standardy prowadzenia badań. Badani mają również prawo do informacji o dostępności wyników badania i sposobie ich wykorzystania. Co więcej, jak to zostanie dalej omówione, taka dostępność i społeczna użyteczność badań mogą być postrzegane jako forma gratyfikacji, a tym samym mogą mieć pozytywny wpływ na motywację do udziału w badaniu.

Bardzo istotne jest, aby podać w zaproszeniu, ile czasu zajmie wypełnienie ankiety. Krótsze badanie wymaga od badanego mniej wysiłku, stąd też podanie niewielkich wartości zwiększa skłonność do rozpoczęcia badania. Nie warto jednak wprowadzać w błąd, gdyż jeśli w trakcie udziału respondenci orientują się, że badanie jest dłuższe niż zapowiedziano, są bardziej skłonni, aby zrezygnować z wypełniania w trakcie⁵¹.

Trudno wyrokować na temat znaczenia tematu zaproszenia. W badaniu Portera i Whitcomba⁵² nie miało to wpływu na skuteczność zaproszeń. Z pewnością warto jednak zastanowić się nad zatytułowaniem e-maila z zaproszeniem w sposób adekwatny, a zarazem nie odstraszaający odbiorców i nie upodabniający zaproszenia do typowych tytułów spamu.

Wygląd zaproszenia

Wraz z rozwojem technologii internetowych i przepustowości łącz pojawiły się możliwości nadania zaproszeniom szczególnego wyglądu, na przykład przez użycie elementów graficznych i w ogóle nietypowego szablonu. Takie rozwiązania mogą jednak przynosić negatywne efekty, prowadząc do zmniejszenia realizacji próby. Respondenci będą mniej chętni, aby odpowiadać na takie zaproszenia. Może to wynikać z większej objętości i co za tym idzie dłuższego czasu ładowania takich wiadomości, z poczucia przeładowania napływającymi bodźcami, a przede wszystkim z tego, że takie metody kojarzą się przede wszystkim z niechcianą pocztą. Takie wnioski znajdują oparcie w badaniu Whitcomba i Portera⁵³, którzy udowodnili negatywny wpływ kolorowego tła i złożonej grafiki zaproszenia e-mailowego na odsetek zwrotów. Należy mieć również na uwadze, że elementy graficzne mogą się źle wyświetlać, szczególnie że użytkownicy korzystają z różnych klientów poczty, a bardzo dużo osób odbiera pocztę jedynie przez strony WWW.

W przypadku zaproszeń wyświetlanych jako wyskakujące okienka wyniki badań polskich badaczy sugerują, że dodanie do tekstu nieskomplikowanych elementów graficznych może zwiększyć jego skuteczność, ale najważniejsze pozostają prosty i estetyczny wygląd⁵⁴. Rekomendujemy zatem powściągliwość i przestrzeganie zasady prostoty graficznej zaproszeń oraz unikanie długich tekstów⁵⁵.

⁵⁰ Por. A.R. Trouteaud, *How do you ask counts: A test of Internet-related components of response rates to a Web-based survey*, „Social Science Computer Review”, 24, 2004, s. 385–393.

⁵¹ S.D. Crawford, M.P. Couper, M.J. Lamias, *Web Surveys: Perceptions of Burden*, „Social Science Computer Review”, 19, 2004, s. 146–62.

⁵² S.R. Porter, M.E. Whitcomb, *E-mail subject lines and their effect on web survey viewing and response*, „Social Science Computer Review”, 23 (3), 2005, s. 380–387.

⁵³ M.E. Whitcomb, S.R. Porter, *Email contacts: A test of complex graphical designs in survey research*, „Social Science Computer Review”, 22 (3), 2004, 370–376.

⁵⁴ J. Przewłocka, P. Janczewski, *Jak e-Kowalski uczył nas stawiać pytania, czyli rola formatu i wyglądu pytań w badaniach online*. Wystąpienie na VIII Kongresie PTBRIo, Warszawa 18–19.10.2007.

⁵⁵ J.M. Zając, D. Batorski, *Jak skłonić do udziału w badaniach internetowych: Zwiększanie realizacji próby*, „Psychologia Społeczna”, nr 3, 2007, s. 234–247.

Gratyfikacja

Dosyć oczywistym sposobem oddziaływania na motywację do udziału w badaniu jest wynagradzanie badanych. Gratyfikacja może przyjmować postać finansową, rzeczową lub symboliczną (podziękowanie, informacja zwrotna, ewentualnie możliwość zobaczenia swoich wyników na tle innych). Może być dostarczona zarówno po zakończeniu badania, jak i przed jego rozpoczęciem, co w przypadku płacenia za udział jest zazwyczaj skuteczniejsze.

W badaniach ewaluacyjnych szczególnie istotne wydaje się stosowanie gratyfikacji symbolicznej, co zresztą wynika też ze standardów prowadzenia badań – warto podziękować badanym i powiedzieć im, do czego potrzebny jest ich udział. W przypadku badania organizacji czy instytucji szczególnie cenny dla respondentów jest często dostęp do wyników i możliwość porównania się z innym. Nawet jeśli podawane są jedynie wyniki zbiorcze, i tak mogą to być ciekawe i potrzebne informacje.

Gratyfikacja finansowa i rzeczowa ma generalnie pozytywny wpływ na realizację próby i ogranicza rezygnację w trakcie⁵⁶. Efekty bywają jednak niewielkie i nie zawsze równoważą koszty i dodatkowy wysiłek związany na przykład z przekazaniem nagród. Mało wiadomo na temat efektów gratyfikacji w czasie, gdy na tej samej próbie realizujemy kilka badań⁵⁷. Większy skutek może przynieść zainwestowanie dodatkowych środków na przykład w personalizację zaproszeń czy stworzenie doskonalszych narzędzi badawczych. Co więcej, nagrody o większej wartości są zazwyczaj tylko trochę bardziej skuteczne.

Efekty gratyfikacji materialnej silnie zależą od kontekstu i specyfiki badanej populacji. Jeśli respondenci biorą udział z ciekawości i chęci pomocy, dostarczanie im bodźców materialnych nie wpływa istotnie na skłonność do udziału. Co więcej, dawanie nagród jest generalnie złą praktyką, gdyż niszczy wewnętrzną motywację. Gratyfikacja często przyciąga określony typ badanych, co może paradoksalnie obniżyć możliwości generalizacji wyników badania. Dodatkowo w przypadku wynagradzania za udział częstsze są próby wielokrotnego udziału⁵⁸ i nierzetelnego wypełniania ankiety, aby tylko skończyć i dostać nagrodę⁵⁹.

W przypadku nagród finansowych i rzeczowych możliwe jest zarówno nagradzanie wszystkich, którzy wzięli udział w badaniu, jak również organizowanie loterii, w której określona liczba osób wygra nagrodę o większej wartości. W przypadku loterii można wyrzucić wrażenie na badanych dużą sumą. Ponadto przyznanie większej nagrody jednej osobie zamiast wielu małych nagród ułatwia wysyłkę i dostarczenie nagród zmniejsza koszty. Brakuje natomiast pewności czy loterie są rzeczywiście bardziej skuteczne niż nagrody przyznawane każdemu⁶⁰ i czy rzeczywiście bardziej okazałe nagrody przy niewielkich szansach wygrania przyciągają bardziej niż mniejsze nagrody z większym prawdopodobieństwem ich zdobycia, jak sugerują Deutsdens i współpracownicy⁶¹. Jeśli jednak organizujemy loterię, warto, żeby jej wyniki były

⁵⁶ A.S. Goeritz, *Incentives in Web studies: Methodological issues and a review*, „International Journal of Internet Science”, 1(1), 2006, s. 58–70.

⁵⁷ A.S. Goeritz, D.G. Goldstein, *Individual payments as a longer-term incentive in online panels*, „Behavior Research Methods”, 40 (4), 2008, s. 1144–1149.

⁵⁸ D. Batorski (2005), *ibidem*; J.A. Konstan, B.R.S. Rosser, M.W. Ross, J. Stanton, W.M. Edwards, *The story of subject naught: A cautionary but optimistic tale of Internet survey research*, „Journal of Computer Mediated Communication”, 10 (2), 2005.

⁵⁹ A.S. Goeritz, *Incentives in Web studies: Methodological issues and a review*, „International Journal of Internet Science”, 1(1), 2006, s. 58–70.

⁶⁰ Por. D. Batorski, *Sieci społeczne a Internet: Sieciowy indywidualizm czy nowe wspólnoty*. (Raport KBN 1-H02E-029-27). Warszawa: Uniwersytet Warszawski, Instytut Studiów Społecznych, 2005; S.R. Porter, M.E. Whitcomb, *The impact of lottery incentives on student survey response rates*, „Research in Higher Education”, 44 (4), 2003b, s. 389–407. T.L. Tuten, M. Galesic, M. Bosnjak, *Effects of immediate versus delayed notification of prize draw results on response behavior in web surveys: An experiment*, „Social Science Computer Review”, 22 (3), 2004, s. 377–384.

⁶¹ E. Deutsdens, K. de Ruyter, M. Wetzels, P. Oosterveld, *Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study*, „Marketing Letters”, 15 (1), 2004, s. 21–36.

dostępne dla badanego od razu po zakończeniu przez niego badania⁶², gdyż respondenci mogą być nieufni co do uczciwości loterii.

Warto też wspomnieć, że obecnie wykorzystanie nagród rzeczowych nie musi ograniczać się do przesyłania nagród poza Internetem. Możliwe jest użycie gratyfikacji rzeczowych w Internecie, jak na przykład dostęp do płatnych serwisów, możliwość ściągnięcia materiałów, kupony zniżkowe czy punkty lojalnościowe w serwisach handlowych⁶³. Wyniki badania Birnholtza i współpracowników sugerują jednak, że skuteczniejsza jest gotówka, gdyż dla respondentów elektroniczne kupony mają obiektywnie niższą wartość niż ekwiwalent w gotówce.

Realizacja próby w trakcie badania

Z pewnością warto jest śledzić realizację próby i *dropout* w trakcie badania, gdyż można w ten sposób uzyskać wiele cennych informacji. Dla przykładu, wysoki odsetek przerywających badanie w jego trakcie, zwłaszcza jeśli dotyczy to jednego z elementów kwestionariusza, może świadczyć o złym sformułowaniu pytania lub o problemach technicznych. Jeśli realizacja próby jest bardzo niska, warto przedsięwziąć dodatkowe środki jak na przykład opisane dalej dodatkowe kontakty z badanymi, czy nawet zmiany w procedurze badania i na przykład kontakty z respondentami także poza Internetem czy umożliwienie odpowiedzi w inny sposób, na przykład przez telefon.

Dobrym sposobem zwiększania skłonności do udziału w badaniu jest wysyłanie przypomnień i mobilizacja osób, które nie wzięły udziału od razu⁶⁴. Można kontaktować się z badanymi także poza Internetem, na przykład pocztą, telefonicznie czy na spotkaniach. Aby nie męczyć respondentów, wskazane jest wysyłanie przypomnień wyłącznie do osób, które jeszcze nie wypełniły ankiety. Jeśli brakuje takich informacji, wyjściem jest wysłanie do całej badanej próby przypomnienia połączonego z podziękowaniem za udział dla tych, którzy to już zrobili i informacją o dacie zakończenia badania⁶⁵. Przypomnienia są wskazane także w sytuacji, gdy respondent może przerwać wypełnianie ankiety i później do niej wrócić.

Przypomnienie przynosi pożądany efekt, jeśli w momencie jego otrzymania sytuacja respondenta jest inna, niż gdy dostał pierwsze zaproszenie. W przypadku badań internetowych taka sytuacja może zmieniać się szybko. Na przykład, gdy badany dostał e-mailowe zaproszenie do ankiet w dniu, gdy otrzymał jednocześnie wiele innych, ważniejszych dla niego wiadomości, i przeoczył je bądź też odłożył na później, warto jest przypomnieć mu w bardziej sprzyjającym momencie. Stąd też wydaje się, że w badaniach internetowych przypomnienia mogą być wysyłane w małych odstępach, na przykład po 2 dniach⁶⁶, zwłaszcza jeśli badani intensywnie korzystają z poczty elektronicznej. Co więcej, po krótkim czasie badanym łatwiej jest skojarzyć ze sobą zaproszenie i przypomnienie, podczas gdy przy dłuższym odstępie mogą myśleć, że chodzi o nowe badanie. Zdarza się jednak, że przypomnienia wysyłane są rzadziej, jak

⁶² T.L. Tuten, M. Galesic, M. Bosnjak, *Effects of immediate versus delayed notification of prize draw results on response behavior in web surveys: An experiment*, „Social Science Computer Review”, 22 (3), 2004, s. 377–384.

⁶³ J.P. Birnholtz, D.B. Horn, T.A. Finholt, S.J. Bae, *The effects of cash, electronic, and paper gift certificates as respondent incentives for a web-based survey of technologically sophisticated respondents*, „Social Science Computer Review”, 22 (3), 2004, s. 355–362; A.S. Goeritz, *Incentives in Web studies: Methodological issues and a review*, „International Journal of Internet Science”, 1(1), 2006, s. 58–70; T.L. Tuten, M. Galesic, M. Bosnjak, *Effects of immediate versus delayed notification of prize draw results on response behavior in web surveys: An experiment*, „Social Science Computer Review”, 22 (3), 2004, s. 377–384.

⁶⁴ D.A. Dillman, *Mail and Internet surveys: The tailored design method*, New York: John Wiley, 2000.

⁶⁵ D. Heerwegh, G. Loosveldt, *Web surveys: The effect of controlling survey access using PIN numbers*, „Social Science Computer Review”, 20, 2002, s. 10–21.

⁶⁶ S.D. Crawford, M.P. Couper, M.J. Lamias, *Web Surveys: Perceptions of Burden*, „Social Science Computer Review”, 19, 2001, s. 146–162.

choćby w drugim, trzecim i czwartym tygodniu sześciotygodniowego badania⁶⁷. Warto uważać, aby nie przesadzać z liczbą przypomnień. Efekty krańcowe kolejnych komunikatów są malejące, poza tym mogą one nużyć i denerwować badanych⁶⁸.

Realizacja próby po zakończeniu badania

Po zakończeniu realizacji badania należy poddać ocenie otrzymany *response rate*. Nie chodzi tu wyłącznie o odsetek zaproszonych osób, które ukończyły badanie, ale również o strukturę uczestnictwa. Szczególnie istotne jest, by dysponować innymi źródłami informacji dotyczących badanej populacji, które pozwolą nam porównać strukturę zrealizowanej próby ze strukturą populacji lub, bardziej dokładnie, ze strukturą wylosowanej próby. Kluczowe jest czy są grupy, których udział w próbie jest wyraźnie różny od występowania w populacji. Jeżeli tak jest, a zmienne wyodrębniające grupy są znacząco skorelowane z najważniejszymi badanymi zmiennymi, oznacza to najprawdopodobniej istotne skrzywienie wyników badania.

Nawet po zakończeniu badania istnieją możliwości korygowania otrzymanych wyników i zwiększania reprezentatywności wyników. Polegają one na konstrukcji wag analitycznych. Wyniki są ważone tak, aby w większym stopniu uwzględnić jednostki z grup o mniejszym uczestnictwie, a mniejsze wagi przypisać osobom z grup o wyższym *response rate*. Istnieją różne metody konstrukcji takich wag, jednak należy podkreślić, że wymagają one posiadania informacji o całej populacji, co może wiązać się z koniecznością realizowanego poza Internetem badania referencyjnego o lepszych charakterystykach. Zbieranie takich, choćby podstawowych, danych typu płeć, wiek itp. warto planować jeszcze w trakcie działań objętych ewaluacją, szczególnie, jeśli możliwe jest uzyskanie takich informacji o wszystkich uczestnikach.

Do najczęściej obecnie stosowanych sposobów redukcji błędu braku odpowiedzi (*non-response error*) należy ważenie przy użyciu wag stratyfikacyjnych. Metoda ta wymaga jedynie informacji o rozkładach częstości najważniejszych zmiennych w populacji i polega na dopasowaniu struktury zrealizowanej próby do struktury całej populacji. Jednak w badaniach internetowych⁶⁹ coraz częściej stosuje się metodę *propensity score weighting*⁷⁰. Jest to o tyle trudniejsze, że metoda ta wymaga posiadania danych o wszystkich członkach populacji, ale w przypadku niektórych badań ewaluacyjnych jest to możliwe. Na podstawie tych danych możliwe jest określenie prawdopodobieństwa udziału w badaniu dla każdej z osób z populacji. Odwrotność tego prawdopodobieństwa jest później wykorzystywana przy konstruowaniu wag analitycznych, używanych do przeważania wyników. Efektem zastosowania *propensity score weighting* jest zmniejszenie błędów oszacowań badania internetowego⁷¹. Bez względu na metodę ważenia, szczególnie dobre efekty daje wykorzystanie do konstruowania wag zmiennych, które są mocno skorelowane ze zmiennymi interesującymi badaczy.

⁶⁷ J.P. Birnholtz, D.B. Horn, T.A. Finholt, S.J. Bae, *The effects of cash, electronic, and paper gift certificates as respondent incentives for a web-based survey of technologically sophisticated respondents*, „Social Science Computer Review”, 22 (3), 2004, s. 355–362.

⁶⁸ D.J. Solomon, *Conducting web-based surveys*, „Practical Assessment, Research & Evaluation”, 7 (19), 2001.

⁶⁹ S. Lee and R. Vallian, *Estimation for Volunteer Panel Web Surveys Using Propensity Score Adjustment and Calibration Adjustment*, „Sociological Methods and Research” 37(3), 2009, s. 319–343; M. Schonlau, A. van Soest, A. Kapteyn and M. Couper, *Selection Bias in Web Surveys and the Use of Propensity Scores*, „Sociological Methods and Research” 37(3), 2009, s. 291–318.

⁷⁰ P.R. Rosenbaum and D.B. Rubin, *Reducing Bias in Observational Studies Using Sub-classification on the Propensity Score*, JASA 79(387): 1984, 516–524; R.J.A. Little and D. B. Rubin (1987), *Statistical Analysis with Missing Data*. New York: John Wiley and Sons, 1987.

⁷¹ S. Lee and R. Vallian, *Estimation for Volunteer Panel Web Surveys Using Propensity Score Adjustment and Calibration Adjustment*, „Sociological Methods and Research” 37(3), 2009, s. 319–343.

Podsumowanie

Niechęć do udziału w badaniach i niska realizacja próby należą do największych problemów ankiet internetowych. Zarazem można wskazać wiele sposobów zaradzenia temu problemowi. Jak staraliśmy się pokazać, na każdym etapie realizacji badań można podjąć działania służące zwiększeniu skłonności respondentów do rozpoczęcia i dokończenia badania, począwszy od wyboru badanej populacji i doboru próby, poprzez wybór czasu badania, komunikację z respondentami, dostarczanie gratyfikacji, odpowiednie reakcje w trakcie badania, aż po wtórne metody zapewniania reprezentatywności. Jednocześnie należy pamiętać, że adekwatność i skuteczność poszczególnych działań silnie zależy od kontekstu danego badania.

Wpływając na skłonność do udziału w badaniu, warto się zastanowić, czy nasze zabiegi nie zwiększają nielosowych odchyłeń. Grozi to wtedy, gdy skuteczność zabiegów jest nierównomierna ze względu na cechy badanych. Ten problem jest szczególnie istotny, jeśli chcielibyśmy generalizować wyniki badania próby na populację. W takiej sytuacji zwiększenie realizacji próby wcale nie musi oznaczać zwiększenia reprezentatywności, lecz wręcz jej zmniejszenie. Ponadto niektóre ze sposobów wpływania na udział w badaniu, jak na przykład gratyfikacja, mogą jednocześnie obniżać jakość danych. Warto także pamiętać, że niektóre sposoby zwiększania skłonności do rozpoczęcia badania mogą jednocześnie wpływać na wzrost rezygnacji w trakcie badania. Nie należy zatem przesadzać w oddziaływaniu na badanych, zwłaszcza jeśli naszym celem nie jest generalizacja z próby na populację, lecz raczej poszukiwanie zależności między zmiennymi.

Jak pokazaliśmy, z wielu względów badania ewaluacyjne bardzo często nadają się do realizacji metodą CAWI. Sądzymy, że w przyszłości badania ewaluacyjne będą coraz częściej wykorzystywały technikę ankiet internetowych. Warto zatem korzystać z wciąż poszerzającej się wiedzy na temat organizacji badania i zachowań respondentów i starać się motywować osoby badane.

Jan M. Zając: Doktor psychologii (UW, 2008), od 2009 r. adiunkt na Wydziale Psychologii Uniwersytetu Warszawskiego. Magister psychologii (UW, 2004) i ekonomii (Szkoła Główna Handlowa, 2003). Przebywał na semestralnych stypendiach na Uniwersytecie Bocconi w Mediolanie (2002) i na Katholieke Universiteit w Leuven w Belgii (2003). Autor kilkunastu publikacji z zakresu psychospołecznych aspektów Internetu oraz metodologii badań online. Brał udział w wielu akademickich i komercyjnych projektach z zakresu badań ilościowych z wykorzystaniem Internetu. Organizator ogólnopolskich konferencji „Społeczne Aspekty Internetu w latach 2005–2006”. Członek International Review Board cyklicznych konferencji „General Online Research” (od 2008 r.).

Dominik Batorski: Doktor socjologii (UW, 2004). Obecnie pracuje jako adiunkt w Interdyscyplinarnym Centrum Modelowania Matematycznego i Komputerowego na Uniwersytecie Warszawskim. Naukowo zajmuje się przede wszystkim: problematyką relacji i sieci społecznych oraz dynamiką procesów społecznych zachodzących w sieciach, takich jak dyfuzja innowacji czy wpływ społeczny. Zajmuje się także metodami analizy sieci społecznych i metodologią prowadzenia badań psychologicznych i społecznych przez Internet. Jest także członkiem Rady Monitoringu Społecznego, przygotowującej cykl badań Diagnoza Społeczna. W badaniach tych odpowiada za moduł poświęcony uwarunkowaniom, sposobom i konsekwencjom korzystania z nowych technologii. W latach 2006–2007 był stypendystą Fundacji na rzecz Nauki Polskiej. Należy do kilku stowarzyszeń naukowych, między innymi International Network for Social Networks Analysis i Polskiego Towarzystwa Socjologicznego.

Bibliografia

- Althoff S., Greif V., Griel B., Batinic B. (2006), *Determinants of response rates of online surveys - the Anita effect: Results of a joint project*. Wystąpienie na konferencji General Online Research 06, Bielefeld.
- Batorski D. (2005), *Sieci społeczne a Internet: Sieciowy indywidualizm czy nowe wspólnoty*. (Raport KBN 1-H02E-029-27). Warszawa: Uniwersytet Warszawski, Instytut Studiów Społecznych.
- Batorski D. (2009), *Korzystanie z technologii informacyjno komunikacyjnych*, [w:] J. Czapiński i T. Panek, *Diagnoza społeczna 2009. Warunki i jakość życia Polaków*, Warszawa.
- Batorski D., Olcoń M. (2006), *Prowadzenie badań przez internet: Podstawowe zagadnienia metodologiczne*. „Studia Socjologiczne”, 3(182).
- Birnholtz J.P., Horn D.B., Finholt T.A., Bae S.J. (2004), *The effects of cash, electronic, and paper gift certificates as respondent incentives for a web-based survey of technologically sophisticated respondents*. „Social Science Computer Review”, 22 (3).
- Choi I. S. (2009), *Online research in Korea: Reliability and validity issues*. Wystąpienie na konferencji The 1st International Workshop on Internet Survey, Deajeon, Korea, 9-10.09.2009.
- Cialdini R. (1996), *Wywieranie wpływu na ludzi*. Gdańsk: GWP.
- Crawford S.D., Couper M.P., Lamias M.J. (2001), *Web Surveys: Perceptions of Burden*. „Social Science Computer Review”, 19.
- Deutskens E., de Ruyter K., Wetzles M., Oosterveld P. (2004), *Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study*. „Marketing Letters”, 15 (1).
- Dillman D.A. (2000), *Mail and Internet surveys: The tailored design method*. New York: John Wiley.
- Dillman D.A., Bowker D.K. (2001), *The Web questionnaire challenge to survey methodologists*, [w:] U.-D. Reips i M. Bosnjak (red.), *Dimensions of Internet science*. Lengerich: Pabst Science Publishing.
- Dillman D.A., Tortora R.D., Bowker D. (1998), *Principles for Constructing Web Surveys*. (SESRC Technical Report 98-50). Pullman, Washington. Pobrano 2 czerwca 2007 r. z <http://www.sesrc.wsu.edu/dillman/papers/websurveyppr.pdf>
- Goeritz A.S. (2006), *Incentives in Web studies: Methodological issues and a review*. „International Journal of Internet Science”, 1(1).
- Goeritz A.S., Goldstein D.G. (2008), *Individual payments as a longer-term incentive in online panels*. „Behavior Research Methods”, 40 (4).
- Goeritz A.S., Stieger S. (2009), *The impact of the field time on response, retention, and response completeness in list-based Web surveys*. „International Journal of Human Computer Studies”, 67.
- Groves R.M., Fowler F.L. jr., Couper M.P., Lepkowski J.M., Singer E., Tourangeau R. (2004), *Survey methodology*. Wiley.
- Guéguen N. (2002), *Foot-in-the-door technique and computer-mediated communication*. „Computers in Human Behavior”, 18.
- Guéguen N., Jacob C. (2002a), *Social presence reinforcement and computer-mediated communication: The effect of the solicitor's photography on compliance to a survey request made by e-mail*. „CyberPsychology & Behavior”, 5.
- Guéguen N., Jacob C. (2002b), *Solicitation by e-mail and solicitor's status: A field study of social influence on the Web*. „CyberPsychology & Behavior”, 5.
- GUS (2008), *Wykorzystanie technologii informacyjno-telekomunikacyjnych w przedsiębiorstwach, gospodarstwach domowych i przez osoby prywatne w 2008 r.*

- Hamilton M. B. (2003), *Online Survey Response Rates and Times: Background and Guidance for Industry*. SuperSurvey White Paper. Pobrano 20 sierpnia 2009 r. from http://www.supersurvey.com/papers/supersurvey_white_paper_response_rates.htm.
- Heerwegh D. (2005), *Effects of personal salutations in e-mail invitations to participate in a Web survey*. „Public Opinion Quarterly”, 69 (4).
- Heerwegh D., Loosveldt G. (2002), *Web surveys: The effect of controlling survey access using PIN numbers*. „Social Science Computer Review”, 20.
- Heerwegh D., Loosveldt G. (2003), *An evaluation of the semiautomatic login procedure to control web survey access*. „Social Science Computer Review”, 21 (2).
- Heerwegh D., Vanhove T., Matthijs K., Loosveldt G. (2005), *The effect of personalization on response rates and data quality in web surveys*. „International Journal of Social Research Methodology: Theory and Practice”, 8.
- Janiszewska A. (2009), *Działania prowadzone przez Gemius na rzecz podnoszenia response rate w badaniach online*. Research paper. Warszawa: Gemius SA.
- Joinson A.N. (2005), *Internet behaviour and the design of virtual methods*, [w:] C. Hine (red.), *Virtual Methods: issues in social research on the Internet*. Oxford: Berg.
- Joinson A.N., Reips U-D. (2007), *Personalized salutation, power of sender and online survey response rates to web-based surveys*. „Computers in Human Behavior”, 23.
- Joinson A.N., Woodley A., Reips U-D. (2007), *Personalization, authentication and self-disclosure in self-administered Internet surveys*. „Computers in Human Behavior”, 23.
- Kaczmirek L. (2009, w druku), *Human-Survey Interaction: Usability and Nonresponse in Online Surveys*. Cologne: Herbert von Halem Verlag.
- Konstan J.A., Rosser B.R.S., Ross M.W., Stanton J., Edwards W.M. (2005), *The story of subject naught: A cautionary but optimistic tale of Internet survey research*. „Journal of Computer Mediated Communication”, 10 (2). Pobrano 28 sierpnia 2009 r. z <http://jcmc.indiana.edu/vol10/issue2/konstan.html>
- Lee S. and Vallian R. (2009), *Estimation for Volunteer Panel Web Surveys Using Propensity Score Adjustment and Calibration Adjustment*. „Sociological Methods and Research” 37(3).
- Little R.J.A. and Rubin D.B. (1987), *Statistical Analysis with Missing Data*. New York: John Wiley and Sons
- Mazurek P., Zając J.M., Rakocy K., (2007), *Online digital surveillance: What they know about you depends on what you know*. Wystąpienie na konferencji General Online Research 07, Lipsk, 26-28.03.2007.
- Porter S.R., Whitcomb M.E. (2003a), *The impact of contact type on Web survey response rates*. „Public Opinion Quarterly”, 67.
- Porter S.R., Whitcomb M.E. (2003b), *The impact of lottery incentives on student survey response rates*. „Research in Higher Education”, 44 (4).
- Porter S.R., Whitcomb M.E. (2005), *E-mail subject lines and their effect on web survey viewing and response*. „Social Science Computer Review”, 23 (3).
- Przewłocka J., Janczewski P. (2007), *Jak e-Kowalski uczył nas stawiać pytania, czyli rola formatu i wyglądu pytań w badaniach online*. Wystąpienie w VIII Kongresie PTBRIQ, Warszawa 18-19.10.2007.
- Rakocy K., Zając, J. M., (2006), *Czy wierzysz w słowa te? Wiarygodność odpowiedzi uzyskiwanych w ankietach internetowych*. Poster przedstawiony na Ogólnopolskiej Konferencji Naukowej Społeczne Aspekty Internetu SAI'06, Warszawa, 8–10.10.2006.
- Reips U-D. (2002), *Standards for Internet-based experimenting*. „Experimental Psychology”, 49 (4).
- Rosenbaum P.R. and Rubin D.B. (1984), *Reducing Bias in Observational Studies Using Sub-classification on the Propensity Score*. JASA 79(387).

- Schonlau M., van Soest A., Kapteyn A. and Couper M. (2009), *Selection Bias in Web Surveys and the Use of Propensity Scores*. „Sociological Methods and Research” 37(3).
- Sills S.J., Song Ch. (2002), *Innovations in survey research: An application of Web-based surveys*. „Social Science Computer Review”, 20.
- Solomon, D.J. (2001), *Conducting web-based surveys*. Practical Assessment, Research & Evaluation, 7 (19).
- Sue, V. M. i Ritter, L.A. (2007). *Conducting online surveys*. Sage.
- Trouteaud, A.R. (2004), *How do you ask counts: A test of Internet-related components of response rates to a Web-based survey*. „Social Science Computer Review”, 24.
- Tuten T.L., Galesic M., Bosnjak M., *Effects of immediate versus delayed notification of prize draw results on response behavior in web surveys: An experiment*. „Social Science Computer Review”, 22 (3).
- Whitcomb M.E., Porter S.R. (2004), *Email contacts: A test of complex graphical designs in survey research*. „Social Science Computer Review”, 22 (3).
- Zajac J.M. (2006), *Nieważne jak pytasz: Cechy zaproszenia a skłonność do udziału w ankiecie internetowej*, [w:] M. Sokołowski (red.), *Oblicza internetu: Internet w przestrzeni komunikacyjnej XXI wieku*. Wydawnictwo PWSZ w Elblągu.
- Zajac J.M., Batorski D. (2007), *Jak skłonić do udziału w badaniach internetowych: Zwiększanie realizacji próby*. „Psychologia Społeczna”, nr 3.

Ankieta Internetowa (CAWI): czy rzeczywiście idealna technika?¹

Uwagi wstępne

Każde badanie jest procesem, nie zaś jednorazowym aktem. Proces ten obejmuje 3 podstawowe etapy:

- 1) przygotowanie badania, którego elementami są: przygotowanie koncepcji badawczej, przygotowanie narzędzia badawczego oraz kosztorysu i harmonogramu badania;
- 2) realizacja badania (zbieranie materiałów) oraz
- 3) ocena i opracowanie materiału sprowadzające się do: wstępnej selekcji i walidacji zebranego materiału, przygotowania zbioru danych i opracowania materiałów (analizy statystyczne i interpretacja wyników).

Na etapie przygotowania badania kluczową sprawą jest opracowanie koncepcji badawczej. Jest to ogół ustaleń dotyczących: przedmiotu (problematyki badania), badanej zbiorowości oraz metody badania. Przygotowując koncepcję, musimy zatem odpowiedzieć na trzy pytania: co badać?, kogo i jaką liczbę osób badać? oraz jak badać? W ramach pierwszego z nich należy określić główny problem (na jakie pytanie ma odpowiedzieć badanie) oraz problemy szczegółowe, które należy uwzględnić w badaniu (wskaźniki). W zależności od powyższych rozstrzygnięć musimy wskazać badaną zbiorowość (jednostki vs. instytucje) oraz charakter próby (próba wyczerpująca vs. próba niewyczerpująca, losowa vs. nielosowa). W zależności od problematyki badania (co badać?) i badanej zbiorowości (kogo badać?) wybieramy dopiero procedurę badawczą (jak badać?). Oznacza to, iż pytania te są zhierarchizowane i odpowiedź na wcześniejsze pytanie wyznacza odpowiedź na kolejne.

Zazwyczaj jednak jest tak, iż nasz problem badawczy, w przypadku interesującej nas zbiorowości, możemy rozstrzygnąć w oparciu o materiały zebrane różnymi technikami. W przypadku badań surveyowych (ilościowych), a więc takich, które ze względu na swoją masowość, losowy dobór próby oraz standaryzację umożliwiają ścisłe oszacowanie liczbowe badanych zjawisk oraz uogólnienie wniosków na całą badaną populację, technik takich jest bardzo wiele. Są to przede wszystkim techniki oparte na komunikacji ustnej, a więc techniki wywiadu osobistego z kwestionariuszem papierowym (PAPI: *Paper and Pencil Interviewing*), wywiadu osobistego ze wspomaganie komputerowym (CAPI: *Computer Assisted Personal Interviewing*) oraz wywiadu telefonicznego ze wspomaganie komputerowym (CATI: *Computer Assisted Telephone Interviewing*). Znacznie rzadziej myśli się o technikach ankietowych, w których Respondent samodzielnie wypełnia formularz ankiety. Chodzi o ankietę „papierową” (typu PSAQ: *Paper Self Administered Questionnaires*), a więc np.: ankietę pocztową, ankietę prasową, ankietę roznoszoną, ankietę pod kontrolą, ankietę audytoryjną oraz ankietę wypełniane samodzielnie przez Respondenta ze wspomaganie

¹ Tytuł jest trawestacją tytułu artykułu PB. Sztabińskiego, *Wywiad osobisty ze wspomaganie komputerowym (CAPI): czy rzeczywiście idealna technika?* ASK „Społeczeństwo, Badania, Metody”, 10; 2001, s. 65–89.

komputerowym (CASI: *CompuTer Assisted Self Interviewing*), ale także VRE (*Voice Recognition Entry*) i TDE (*Touchtone Data Entry*)².

Jak zatem wybrać właściwą, najlepszą czy optymalną technikę badania, jakimi kryteriami należy się kierować, biorąc pod uwagę nasze zapotrzebowanie badawcze i charakter próby? Kryterium takim może być adekwatność techniki do problematyki i badanej zbiorowości, tzn. adekwatność metodologiczna (a więc jej informatywność i wiarygodność) oraz adekwatność psychologiczno-społeczna (ze względu na cechy emocjonalne i intelektualne badanej zbiorowości oraz doświadczenia społeczne)³. Innym kryterium może być minimalizacja *Total Survey Error* (TSE). Mówiąc o TSE mam na myśli⁴:

- błąd próby (tzw. *sampling error*), który jest efektem tego, iż badanie realizowane jest wśród wylosowanych osób, nie zaś na całej zbiorowości;
- błąd pokrycia (tzw. *coverage error*), który polega na tym, że operat losowania nie obejmuje wszystkich jednostek, wśród których chcemy zrealizować badanie;
- błąd braku odpowiedzi (tzw. *nonresponse error*), zarówno jako odmowa udziału w badaniu (*unit nonresponse*), jak i odmowa odpowiedzi na dane pytanie (*item nonresponse*). Błąd ten nie ma charakteru losowego. Jest błędem systematycznym. Jednostki niedostępne różnią się w sposób istotny od pozostałych. Powoduje to ryzyko tendencyjności próby (ze względu na cechy społeczno-demograficzne, psychologiczne oraz opinie) i w rezultacie wypaczenie wniosków z badania⁵;
- błąd pomiaru (tzw. *measurement error*), który jest „sumą” błędów związanych ze sposobem uzyskiwania danych (techniką) i sformułowaniem pytań (tzw. błąd narzędzia) oraz błędów związanych z osobą ankietera i osobą Respondenta.

Minimalizacja TSE pociąga za sobą jednak określone koszty finansowe oraz wydłuża czas realizacji badania. Jest to swoisty „*trade off*” między jakością badania (minimalizacja błędów) a kosztem oraz czasem jego realizacji. Koszty badania są bowiem odwrotnie proporcjonalne do wielkości błędu badania: chcąc je zredukować, musimy ponosić większe koszty⁶.

Niezależnie od tego, które kryterium wybierzemy, musimy wziąć pod uwagę, że każda technika badawcza ma swoje zalety i ograniczenia. Przykładowo podstawowe zalety grupy technik wywiadu (PAPI, CAPI, CATI) w porównaniu z ankietą, to: bardzo szeroki zakres zastosowań, możliwość pomocy i dodatkowych wyjaśnień ze strony ankietera (w przypadku pytań trudnych, a także w przypadku znużenia, zniechęcenia Respondenta, co minimalizuje ryzyko artefaktów), standaryzacja sytuacji badania (miejsca i warunków), pewność, że w badaniu uczestniczy właściwa osoba oraz kompletność wypełnienia kwestionariusza. Jednocześnie wiadomo, iż w przypadku tej grupy technik istnieje ograniczone poczucie anonimowości badanych (w związku z tym istnieje ryzyko udzielania społecznie akceptowanych odpowiedzi), występuje wpływ ankietera, a koszty realizacji badania są bardzo wysokie. Zastosowanie technik

² VRE (*Voice Recognition Entry*) jest techniką, w której Respondent odsłuchuje nagranych pytań przez telefon i udziela odpowiedzi również przez telefon. Z kolei TDE (*Touchtone Data Entry*) jest techniką, w której Respondent odsłuchuje nagranych pytań przez telefon i udziela odpowiedzi używając klawiszy telefonu. Te ostatnie należy uznać raczej jako techniki wywiadu, nie zaś ankiety (jest to komunikacja „ustna z komputerem”).

³ F. Sztabiński, *Logika badacza i logika respondenta. Problem adekwatności narzędzia pomiaru*, ASK, „Społeczeństwo, Badania, Metody”, 12, 2003, s. 150–154.

⁴ Zob. R.M. Groves *Survey Errors and Survey Costs*, New York, John Wiley & Sons, 1989.

⁵ Obecnie błąd ten jest z jednym najpowszechniejszych i najpoważniejszych. W ciągu 16 lat (1992–2008) odsetek realizacji próby w projekcie Polski Generalny Sondaż Społeczny zmniejszył się o 30 punktów procentowych: z 82% do 52%, co oznacza, że w 2008 nie zrealizowano 48% wylosowanej próby. Za: B. Cichomski (kierownik programu), T. Jerzyński, M. Zieliński, *Polskie Generalne Sondaże Społeczne: struktura skumulowanych wyników badań 1992–2008*. Instytut Studiów Społecznych, Uniwersytet Warszawski, Warszawa, styczeń 2009. Nie jest to jednak zjawisko obserwowane tylko w Polsce: z pewnością dotyczy ono całej Europy. Zob.: I. Stoop, *The Hunt for the Last Respondent: Nonresponse in Sample Surveys*. The Hague: Social and Cultural Planning Office of the Netherlands, 2005.

⁶ Wydaje się, iż wpływ na rozmiar TSE, poza kosztami i czasem mają również aktualne warunki prowadzenia badań. Chodzi tutaj generalnie o „stosunek” społeczeństwa do badań surveyowych, ich akulturację, ale także wzrost poczucia podmiotowości badanych. Oczywiście przekłada się to pośrednio na błąd wynikający z odmowy udziału w badaniu (*unit nonresponse*).

ankietowych – standaryzowanych, opartych na komunikacji pisemnej (tzn. PSAQ i CSAQ) gwarantuje niższy koszt badania (w tym „niezależność” geograficzną, co nie pozostaje bez wpływu na zmniejszenie błędu próby i tym samym TSE) oraz poczucie anonimowości badanych. Posiadają jednak one te wszystkie wady, które są zaletami wywiadu (nie mamy pewności, kto wypełnia ankietę; ze względu na konsultację „wewnętrzdomową” odpowiedzi są zakotwiczone społecznie – uzgodnione z innymi osobami, nie zaś indywidualne; długi okres realizacji badania; relatywnie niska jakość materiałów – część ankiet wraca niewypełniona lub wypełniona niekompletnie; trudność realizacji badania na próbach innych niż imienne). Z kolei, podstawowymi zaletami grupy technik wspomaganych komputerowo (CASIC: *Computer Assisted Survey Information Collection*), a więc CATI, CAPI, CASI, w porównaniu z technikami „papierowymi” (PAPI, PSAQ) jest: standaryzacja zachowań ankierskich, ograniczenie błędów ankietera oraz możliwość monitorowania przebiegu wywiadu i skrócenie czasu realizacji badania, a w dłuższej perspektywie także obniżenie kosztów badania. Ich ograniczeniami są z kolei: redukcja wymiaru „społecznego”, ograniczenie lub eliminacja komunikacji niewerbalnej, ryzyko wypaczenia próby ze względu na poziom penetracji telefonów/Internetu.

Przy czym, oczywiście należy pamiętać o specyfice poszczególnych grup technik. I tak, wywiad PAPI, pod pewnymi względami, ma „przewagę” nad CAPI lub CATI, a pod innymi im „ustępuje”. Podobnie jest w przypadku poszczególnych technik ankiety (np. technik ankiety „papierowej” – PSAQ i ankiety wspomagananej komputerowo – CSAQ). Wszystko to przekłada się na uzyskiwane informacje: w zależności od tego, jaką zastosujemy technikę badania, możemy uzyskać różne informacje. Każda technika badawcza stwarza bowiem specyficzną sytuację badania. W rezultacie dane uzyskiwane za jej pomocą mogą być nieporównywalne z danymi uzyskiwanymi za pomocą innej techniki⁷. Przygotowując zatem badanie należy przede wszystkim mieć świadomość, jaki wpływ na uzyskane rezultaty może mieć wybrana technika.

Możemy sobie pomarzyć, że gdybyśmy mieli dużo pieniędzy i dużo czasu, to znaleźlibyśmy idealną technikę eksploracji naszego problemu. Ale zazwyczaj nie mamy nadmiaru ani jednego, ani drugiego. Bez wątpienia, techniki wspomagane komputerowo są relatywnie tańsze niż techniki tradycyjne („papierowe”). Co więcej, okres realizacji badania z ich wykorzystaniem jest krótszy. Generalnie można przyjąć, iż ze względu na koszty i długość okresu realizacji badania, na ostatnim miejscu plasuje się wywiad PAPI i CAPI. Przodują zaś techniki wspomagane komputerowo, administrowane samodzielnie przez Respondenta (CASI), w szczególności CAWI (*Computer Assisted Web Interviewing*). Technika ta jest najtańsza i w związku z tym winna pozwalać na minimalizację błędów, przy danych kosztach. Czy jednak rzeczywście techniki internetowe są optymalne i idealne?

Badania internetowe⁸

Początki badań internetowych datuje się na połowę lat 90. i wiążą się z możliwością powszechnego dostępu do sieciowego systemu wymiany informacji WWW (*World Wide Web*), aczkolwiek już w latach

⁷ Przykładowo zob.: F. Sztabiński, *Ankieta pocztowa i wywiad kwestionariuszowy*, Wydawnictwo IFIS PAN, Warszawa, 1997 s. 118–130; P.B. Sztabiński, *Wywiad telefoniczny ze wspomaganiem komputerowym (CATI) a wywiad kwestionariuszowy. Problem porównywalności wyników*, [w:] H. Domański, K. Lutyńska, A.W. Rostocki, (red.) *Spojrzenie na metodę. Studia z metodologii badań socjologicznych*, Wydawnictwo IFIS PAN, Warszawa 1999, s. 47–70. Efekt techniki (tzw. *Mode Effect*) występuje oczywiście z różnym natężeniem, w zależności od techniki, przedmiotu i struktury pytania. Technika wpływa również na występowanie efektu znużenia, skłonność do „zgadzania się”, czy efekt porządku dla skal semantycznych.

⁸ Używam terminu badania internetowe w odróżnieniu od badań Internetu, mając na myśli wykorzystanie Internetu jako medium komunikacji między badaczem a badanymi. W drugim przypadku (badania Internetu) chodzi bowiem o pomiar i opis aktywności użytkowników Internetu (tzn. ruchu internetowego). Por.: T. Zmijewska-Jędrzejczyk, *Badania internetowe*, [w:] P.B. Sztabiński, F. Sztabiński, Z. Sawiński (red.), *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Wydawnictwo IFIS PAN, Warszawa, 2004, s. 241–245.

80. w Stanach Zjednoczonych wykorzystywano technikę EMS (*Electronic Mail Survey*)⁹. Obecnie, mówiąc o badaniach internetowych ma się na myśli dwie podstawowe techniki internetowe: e-mailing oraz Web-survey. E-mailing jest najwcześniejszą formą badań internetowych. Polega on na tym, iż w e-mailu lub do e-maila (w formie załącznika) dołączana jest treść ankiety. Po udzieleniu odpowiedzi na pytania ankiety badani odsyłają ją na wskazany adres e-mailowy. W tym przypadku badani nie muszą być użytkownikami stron WWW, wystarczy, jeśli korzystają z programu pocztowego. Technika ta wykorzystuje Internet jedynie jako kanał dystrybucji. Z kolei Web-survey to wszelkiego rodzaju ankiety umieszczone w Internecie, edytowane na stronach WWW. Technika ta jest zatem adaptacją tradycyjnych metod gromadzenia informacji za pośrednictwem nowego medium, jakim jest Internet, jednocześnie jest alternatywnym sposobem komunikacji z respondentem w stosunku do komunikacji ustnej i pisemnej – „papierowej”. Znajduje ona zastosowanie zarówno w badaniach ilościowych (wspomniane: *e-mailing* i – *Web surveys*), jak i jakościowych (*On-line Focus Group Interview*, *e-mail groups*, *e-mail In-Depth Interviews*).

Czym jednak w istocie są techniki internetowe? Czy są to raczej techniki wspomagane komputerowo, czy też raczej techniki ankietowe? Bez wątplenia skupiają one ograniczenia i możliwości obu grup technik. Ich zaletami, jak wspomniano, są między innymi: krótki czas realizacji badania, możliwość dotarcia do Respondentów trudno dostępnych, w tym także „niezależność” geograficzna i ograniczenie kosztów badania (brak uczestnictwa ankieterów), swoboda w udzielaniu odpowiedzi (Respondent ma czas na sformułowanie odpowiedzi), większe poczucie prywatności oraz brak wpływu ankietera na Jego odpowiedzi. Istnieje też możliwość wzbogacenia narzędzia o elementy audio i wideo oraz indywidualizacji kwestionariusza oraz jego dostosowania do różnych kategorii Respondentów. Równocześnie jednak pamiętać należy o ograniczeniach tych technik. Nigdy nie mamy pewności, czy ankieta trafiła do adresata, czy została wypełniona samodzielnie i kto właściwie odpowiada na pytania ankiety. Brak jest również możliwości kontroli kompletności i poprawności wypełnianej kwestionariusza w przypadku skomplikowanych reguł przejścia (o ile nie zakładamy możliwości „blokowania” poszczególnych okien). Co więcej, ze względu na różne wersje oprogramowania, nigdy nie mamy pewności, że kwestionariusz ankiety dotrze do wszystkich Respondentów w tej samej formie. Może się zdarzyć, że „rozformatuje” się on i na przykład opisy skali przy pytaniu „przesuną się”. Nie bez znaczenia jest również fakt, iż przed rozpoczęciem wypełniania ankiety Respondent może zapoznać się z jej treścią, z poszczególnymi pytaniami (istnieje jednak możliwość „zablokowania” opcji przeglądania całego kwestionariusza).

I wreszcie ostatnie, ale nie najmniej ważne ograniczenie związane z możliwością wykorzystania tych technik w praktyce badawczej. Techniki internetowe – *ex definitione* – zakładają, iż badane przez nas osoby mają dostęp i posługują się Internetem. Tymczasem blisko 48% Polaków (w wieku 15 lat i powyżej) nie ma dostępu do Internetu lub z niego nie korzysta¹⁰. Co więcej, jeśli przyjmiemy, że za internautę uważa się osobę korzystającą z Internetu raz w miesiącu lub częściej, to tylko 50% badanych spełnia to kryterium¹¹. Dane te nie przekładają się oczywiście na penetrację dostępu do Internetu w gospodarstwach domowych. Według raportu Instytutu Łączności „Inwestycje przyszłością rozwoju telekomunikacji” opublikowanego w 2008 r. wynika, iż tylko 30% gospodarstw domowych w Polsce posiada dostęp do Internetu.

⁹ Zob.: M. Couper, W.L. Nicholls II, *The History and Development of Computer Assisted Survey Information Collection Methods*, [w:] M. Couper, R.P. Baker, J. Bethlehem, C.Z. Clark, J. Martin, W. L. Nicholls II, J. M. O'Reilly, *Computer Assisted Survey Information Collection*, New York, John Wiley Co; 1998, s. 13. Obecnie technika ta oraz generalnie e-mailing jest coraz rzadziej stosowany (aczkolwiek chyba jest to błędem – przyp. F. S.).

¹⁰ Na podstawie badania Europejski Sondaż Społeczny Runda 4, realizowanego na przełomie roku 2008/2009. Zob.: <http://ess.nsd.uib.no/ess/round4/>. Pytanie dotyczyło korzystania z Internetu, sieci WWW lub poczty elektronicznej w domu lub w pracy.

¹¹ Definicja internauty przyjęta za Internet Advertising Bureau. Zob.: <http://www.iabeurope.eu/media/6238/directory%20of%20definitions%20ed1ii.pdf>.

Plasuje to Polskę znacznie poniżej średniej z krajów starej piętnastki UE, która wynosi 46%¹². Co więcej, dostęp do Internetu jest skorelowany z wiekiem, wykształceniem i wielkością miejsca zamieszkania – jest on odwrotnie proporcjonalny do tych cech¹³. Z kolei penetracja dostępu do Internetu dla przedsiębiorstw zatrudniających więcej niż dziesięć osób sięga 53%, podczas gdy średnia z krajów piętnastki UE wynosi 82%. Wraz z Polską ranking zamykają Rumunia i Litwa¹⁴. Powyższe dane wskazują na największe ograniczenie technik internetowych, a mianowicie występowanie błędu pokrycia. Jak wspomniano, polega on na tym, że operat losowania nie obejmuje wszystkich jednostek, wśród których chcemy zrealizować badanie. Problem ten, co należy podkreślić, dotyczy nie tylko jednostek (gospodarstw domowych), ale i instytucji¹⁵. Ma to istotne konsekwencje dla wszelkiego rodzaju badań, w tym także badań ewaluacyjnych. Co więcej, na błąd pokrycia „nakłada się” błąd braku odpowiedzi. Szacuje się, iż Response Rate wynosi od 1% (przy samodoborze on-line w *Banner surveys* i *Pop-up surveys*) do 50% (przy pre-rekrutacji „off-line”: PAPI, CATI, CAPI); zaś w badaniach B2B około 70%¹⁶.

Enumeracja ograniczeń i zalet technik internetowych niewiele nam mówi o ich istocie i nie daje odpowiedzi na pytanie czy są to techniki zupełnie nowe, czy też, wykorzystywane od wielu lat, ankiety wypełniane samodzielnie przez Respondenta? Zarówno e-mailing, jak i Web-survey są technikami opartymi na komunikacji pisemnej, nie zaś ustnej. Respondent NIE SŁYSZY pytania, lecz WIDZI tekst pytania¹⁷. Z kolei, udzielając odpowiedzi na pytanie, Respondent NIE MÓWI, lecz PISZE. Oznacza to, że w technikach ankietowych źródło informacji jest tożsame z materiałem badawczym. Innymi słowy, wypowiedź Respondenta (pisemna) dociera do badacza w formie pierwotnej. Wszystko to, co chciał przekazać Respondent, dociera do badacza. W wywiadzie kwestionariuszowym jest inaczej. Źródło informacji nigdy nie jest tożsame z materiałem badawczym: wypowiedź Respondenta (ustna) dociera do badacza „prze-filtrowana” przez ankietera (jest to efekt dwuogniowości procesu komunikowania między badaczem i Respondentem¹⁸). Ale jednocześnie badany WIDZI i PISZE na KOMPUTERZE, nie zaś na papierze. I to jest sens badań internetowych.

Ponieważ w technikach internetowych Respondent odczytuje tekst pytania i zapisuje swoje odpowiedzi, w największym stopniu, biorąc pod uwagę sposób komunikowania, przypominają one technikę ankiety pocztowej. Internet jest jedynie specyficznym kanałem dystrybucji. Niemniej, właśnie ze względu na owe medium, nie możemy zapominać o „związках” tych technik z innymi technikami wspomagany-
mi komputerowo i o tym, że należą one do „rodziny” CADAC (*Computer Assisted Data Collection*)¹⁹.

¹² Za: <http://biuroprasowe.tp.pl/pr/107315/instytut-laczności-potrzebna-strategia-inwestycji-telekomunikacyjnych?changeLocale=PL>. Warto tutaj wspomnieć, iż w połowie roku 2008 średnia światowa wyposażenia gospodarstw domowych w Internet wynosiła 21,9% (Północna Ameryka – 73,6%, Afryka 5,3%). Dla Europy średnia wynosiła 48,1% (Holandia – 86%, Rumunia – 11%). Za: www.internet-worldstats.com.stats.htm

¹³ Zob.: Polski Internet 2008/2009, http://pliki.gemius.pl/Raporty/2009/02_2009_Polski_internet_2008_2009.pdf oraz E.D. de Leeuw, *Mixed Mode Surveys – 6th ESS Training Course*, 30. 04–01. 05., Ljubljana, 2009.

¹⁴ Za: <http://biuroprasowe.tp.pl/pr/107315/instytut-laczności-potrzebna-strategia-inwestycji-telekomunikacyjnych?changeLocale=PL>

¹⁵ Zaobserwować można tutaj różnice w zależności od wielkości instytucji, charakteru jej działalności (np.: relatywnie niska penetracja internetu dla firm produkcyjnych, z branż przetwórstwa spożywczego, czy usług komunalnych vs znacznie wyższa w przypadku firm z branży doradztwa prawnego lub finansowego), ale także grup respondentów w instytucji (np.: kadra zarządzająca, wysokiej klasy specjaliści).

¹⁶ Tym samym realizując badania internetowe na próbie ogólnej popełnić możemy taki sam błąd, jaki został popełniony w 1936 r. w Stanach Zjednoczonych w badaniu Literaty Digest. Zob. T.W. Mangione, *Mail Surveys. Improving the Quality*, Thousand Oaks, Sage, 1995, s. 3–4.

¹⁷ Pomijam tutaj specyficzną technikę, jaką jest ACASI (*Audio Computer Assisted Self-interviewing*). W przypadku tej techniki Respondent odsłuchuje treść pytań z komputera (mogą one jednocześnie być wyświetlane na ekranie) i zapisuje w komputerze swoje odpowiedzi. Za: M. Couper, W. L. Nicholls II, *The History and Development of Computer...*, s. 14.

¹⁸ Zob.: F. Sztabiński, *Ankieta pocztowa i wywiad...*, s. 21–47.

¹⁹ Zob.: W.E. Saris, *Computer-Assisted Interviewing*, Newbury Park. Sage, 1991. Przywołuję tutaj tę pozycję, ponieważ jest ona jedną z pierwszych, która w sposób całościowy podejmowała problematykę technik wspomaganych komputerowo.

Ankieta PSAQ i ankieta CSAQ

Czy ankieta „papierowa” różni się od ankiety „elektronicznej”, tzn. e-mailowej i Web-owej? Jest to zasadniczo pytanie retoryczne. Tak, różni się, ale zasadniczo nie powinna. Tak można streścić książkę Dona A. Dillmana zatytułowaną *„Mail and Internet Surveys: The Tailored Design Method”*²⁰. Jest to druga książka Dillmana poświęcona technikom ankietowym. Pierwsza z nich, wydana w 1978 r., nosiła tytuł: *„Mail and Telephone Surveys. The Total Design Method”*²¹. Wspominam o tym, ponieważ to dzięki niemu, a właściwie jego pierwszej książce nastąpił renesans ankiety pocztowej. Podstawową jej tezę było twierdzenie, iż badanie ankietowe jest formą wymiany społecznej, w przypadku której zachowanie jednostki jest funkcją przewidywanych kosztów podjęcia określonego działania oraz oczekiwanych nagród związanych z jego podjęciem²². W oparciu o tę tezę, Dillman formułuje wniosek, iż warunkiem efektywności ankiety jest minimalizacja kosztów związanych z odpowiadaniem na pytania, przy jednoczesnej maksymalizacji samej nagrody oraz przekonania o jej uzyskaniu (zarówno pojęcia „nagroda” oraz „koszty” używa się tutaj w szerokim rozumieniu). Jest to zatem swoisty „trade off” między kosztami a nagrodami. Powyższe stwierdzenie ma kilka bardzo doniosłych implikacji. Otóż, aby w ankiecie osiągnąć satysfakcjonujące rezultaty (tzn. zredukować TSE – *Total Survey Error*, w szczególności błąd pomiaru i braku odpowiedzi) badacz musi przede wszystkim²³:

- maksymalizować nagrody poprzez: okazywanie szacunku i pozytywnego wyobrażenia o Respondencie; wyrażane werbalnie uznanie dla Respondenta; postawienie Respondenta w roli eksperta, z którym chcemy przekonsultować pewne problemy; wskazanie użyteczności społecznej badania (dla określonych grup, których Respondent jest członkiem); wzbudzenie identyfikacji grupowej; proponowanie gratyfikacji niematerialnych (np. raport z wynikami badania) i materialnych (pieniężne i niepieniężne); przygotowanie atrakcyjnego i interesującego kwestionariusza; wskazanie ograniczonych możliwości uczestnictwa w badaniu, o ile decyzji nie podejmie się natychmiast;
- minimalizować społeczne koszty ponoszone przez Respondenta poprzez: unikanie sformułowań w języku „nakazowo-rozdzielczym” i eliminowanie wszystkich negatywnych implikacji związanych z poczuciem „konieczności subordynacji”; eliminowanie pytań trudnych i drażliwych, wzbudzających zakłopotanie (np. ze względu na brak kompetencji); eliminowanie wszelkich niedogodności związanych z udziałem w badaniu; wzbudzanie przekonania, iż wypełnienie kwestionariusza zajmie Respondentowi niewiele czasu; ograniczenie liczby pytań dotyczących danych osobowych; upodobnienie pytań do innych, na które Respondenci już udzielali odpowiedzi (chodzi zarówno o formę, jak i ich przedmiot); redukcja fizycznego i psychicznego wysiłku związanego z odpowiadaniem na pytania;
- budowanie wzajemnego zaufania poprzez: zastosowanie bezwarunkowych gratyfikacji (upominek z badania, niezależnie od faktu uczestnictwa w nim, który jest „symbolicznym gestem zaufania”. Może on być bardziej efektywny niż znacznie większa gratyfikacja warunkowa); identyfikację badań z instytucją cieszącą się powszechnym zaufaniem i poważaniem; nadanie odpowiedniej rangi badaniu,

²⁰ D.A. Dillman, *Mail and Internet Surveys: The Tailored Design Method*, New York, John Wiley & Sons. Inc. Edith 2000. De Leeuw twierdzi, iż jest to jedynie różnica środka przekazu za pośrednictwem którego komunikujemy się z Respondentem. Za: E.D. De Leeuw, *Mixed Mode Surveys...*

²¹ D.A. Dillman, *Mail and Telephone Surveys. The Total Design Method*, New York, John Wiley & Sons. Inc., 1978.

²² Chodzi tutaj o koncepcje wymiany społecznej, które rozwinięte zostały przede wszystkim przez Homansa (1961), Blaua (1964) Thibaut i Kelleya (1959). Zob.: G.C. Homans, *Social Behaviour: Its Elementary Forms*, New York, Harcourt, Brace & World, 1961; P.M. Blau, *Exchange and Power in Social Life*, New York, John Wiley & Sons. Inc. 1964; J.W. Thibaut, H.H. Kelley, *The Social Psychology of Groups*, New York, John Wiley & Sons. Inc. 1959. Jednocześnie, aczkolwiek Dillman tego nie pisze, jest to klasyczny przypadek wymiany dokonujący się wg zasady: „do ut des”. Zob. B. Malinowski, *Zwyczaj i zbrodnia w społeczności dzikich*, [w:] B. Malinowski, *Dzieła*, t. 2, Warszawa, PWN, 1984.

²³ D. A. Dillman, *Mail and Internet Surveys...*; 2000, s. 15–21. Zob. także: F. Sztabiński, *Ankieta pocztowa i wywiad...*; 1997, s. 143–146.

poprzez wskazanie jego użyteczności; rozszerzenie zakresu wymiany poprzez nadbudowanie innych relacji.

Połączenie tych trzech elementów wymiany społecznej, które mogą i powinny wzajemnie się uzupełniać i wspomagać, oznacza przede wszystkim podmiotowe traktowanie Respondenta. Mówiąc wprost: traktujemy Respondenta w taki sposób, w jaki chcielibyśmy, aby nas traktowano²⁴. Oznacza to przede wszystkim:

- zerwanie z bezosobowością Respondenta. Osoba, do której zwracamy się z prośbą o udzielenie odpowiedzi na pytania, nie jest dla nas osobą anonimową, zazwyczaj znamy przynajmniej jej płeć. Należy zatem przygotować narzędzie w dwóch równoważnych wersjach: dla kobiet i mężczyzn. Zabieg ten ogranicza również „koszty” ponoszone przez Respondenta. Przykładowo, tekst pytania: „W którym roku się Pani urodziła?” jest bardziej „przyjazny” niż: „W którym roku się Pan/Pani urodził/urodziła?”. Co więcej, zabieg ten jest jednocześnie „nagrodą” dla Respondenta, który czuje się „dowartościowany” faktem, iż uwzględniona została jego płeć;
- zerwanie z anonimowością kontaktu ze strony badacza. To nie instytucja, np. Instytut Filozofii i Socjologii Polskiej Akademii Nauk, a konkretna osoba – Franciszek Sztabiński, pracownik Instytutu Filozofii i Socjologii Polskiej Akademii Nauk zwraca się z prośbą o wypełnienie ankiety;
- symetryczność układu ról: badacz–badany, która wyraża się w dyspozycyjności wobec Respondenta („jestem dostępny pod telefonem: ... oraz adresem e-mail’owym: ...”);
- bezpośrednie dowartościowanie Respondenta poprzez gratyfikacje materialne i niematerialne. Te ostatnie (np. udostępnienie wyników badania) mogą się okazać bardziej efektywne niż gratyfikacje materialne;
- minimalizację wysiłku Respondenta związanego z wypełnieniem ankiety. Chodzi tutaj nie tylko o „przejrzystość graficzną” kwestionariusza, ale także „uproszczenie” pytań;
- naturalizację kontaktu z Respondentem poprzez rozciągnięcie badania w czasie i „zaprzyjaźnienie się” z Respondentem.

O ile spełnienie powyższych „wymogów” na „papierze” (chodzi o ankiety „tradycyjne”) nie nastręcza zbytnich trudności, to w przypadku ankiet wspomaganych komputerowo sprawy się nieco komplikują. Kwestionariusz ankiety CASI winien być bowiem pisany, nie tak jak mówi Dillman, w dwóch językach (języku słów oraz języku symboli graficznych, które prowadzą Respondenta po kwestionariuszu), a w trzech językach: języku słów, języku symboli graficznych oraz języku programowania²⁵. Badacz powinien zatem być równocześnie: sobą (badaczem), grafikiem i programistą. A nie jest to z pewnością łatwe.

Kwestionariusz na papierze, kwestionariusz w komputerze i kwestionariusz w Internecie

Czy zasady konstruowania pytań zadawanych w ankietach internetowych różnią się od zasad konstruowania pytań stosowanych w innych technikach? Zasadniczo nie, ale ... pamiętać musimy, iż w wielu przypadkach widziany przez nas kwestionariusz nie będzie kwestionariuszem widzianym przez Respondenta. Dzieje się tak za sprawą różnic sprzętu i oprogramowania. O ile w przypadku kwestionariusza papierowego mamy do czynienia z sytuacją²⁶:

²⁴ Nie bez powodu zatem, mówiąc o Respondencie, używam zawsze dużej litery.

²⁵ D.A. Dillman, *Mail and Internet Surveys...*; s. 24. Zob. także: S.H. Kinsey, D.M. Jewell, *A Systematic Approach to Instrument Development in CATI*, [w:] M. Couper, R.P. Baker, J. Bethlehem, C.Z. Clark, J. Martin, W.L. Nicholls II, J.M. O'Reilly, *Computer Assisted Survey...*; 1998, s. 105–123.

²⁶ D. A. Dillman, *Mail and Internet Surveys...*; s. 360.

BADACZ ➞**➞ Kwestionariusz ➞****➞ RESPONDENT**

o tyle, w przypadku kwestionariusza przekazywanego przez Internet mamy do czynienia z następującą:

BADACZ ➞**➞ Hardware i Software Badacza ➞****➞ Kwestionariusz widziany przez Badacza ➞****➞ Hardware i Software Respondenta ➞****➞ Kwestionariusz widziany przez Respondenta ➞****➞ RESPONDENT****Kwestionariusz na papierze**

Pytanie jest bardzo wrażliwym instrumentem pomiaru. Sposób jego sformułowania, obok błędów próby, stanowi podstawowe źródło błędu wyniku badania. Niewielkie różnice sformułowania pytania (przy zachowaniu ekwiwalentności logicznej) mogą powodować znaczące różnice w rozkładach odpowiedzi²⁷.

Wspomniany „potrójny” język kwestionariusza nie ma oczywiście zasadniczego znaczenia przy konstruowaniu pytań, niemniej należy zawsze o tym pamiętać. Zasady formułowania pytań są praktycznie uniwersalne, co oznacza, iż należy²⁸:

- używać krótszych, mniej skomplikowanych terminów i sformułowań (np.: zmęczony, zamiast wyczerpany), nie używać skrótów (np.: Powiatowy Urząd Pracy, a nie PUP);
- stosować proste i jasne pytania, aby respondenci wiedzieli, jakiego typu informację chcemy od nich uzyskać. Przykładowo, pytanie kafeteryjne winno mieć następującą konstrukcję: 1. Wprowadzenie; 2. Zadanie (co ma zrobić Respondent); 3. Możliwe odpowiedzi (kafeteria);
- formułować pytanie w formie rozwiniętej – pytanie nie może być sformułowane hasłowo lub w postaci zdania oznajmującego, np. „wiek” lub „liczba osób w gospodarstwie domowym”, „dochody” itp.;
- unikać niejasnych lub trudnych określeń i terminów;
- dążyć do uzyskania informacji bardziej precyzyjnych, zamiast ogólnych i przez to niejasnych. Przykładowo, w pytaniu o częstotliwość udziału w mszach, nabożeństwach lub spotkaniach religijnych zastosowana skala winna być następująca: „1. Codziennie, 2. Częściej niż raz w tygodniu, 3. Raz na tydzień, 4. Co najmniej raz w miesiącu, 5. Tylko z okazji szczególnych świąt, 6. Jeszcze rzadziej, 7. Nigdy”; zamiast: „1. Często, 2. Dość często, 3. Od czasu do czasu, 4. Bardzo rzadko, 5. Nigdy”;
- unikać żądania od respondenta zbytnej precyzji. Przykładowo, jeśli chcemy dowiedzieć się, ile książek przeczytał respondent w ostatnim roku, nie należy zadawać pytania otwartego, lecz zamknięte, wskazując przedziały liczbowe np.: „1, żadnej, 2, 1-2; 3, 3-5 itd.”;
- konstruować symetryczne skale słowne, tzn. liczba pozytywnych i negatywnych punktów skali winna być jednakowa;
- w skalach słownych odróżnić punkt „niezdecydowany” (nie wiem, trudno powiedzieć) od „neutralny” (ani, ani). Ten pierwszy nigdy nie powinien być „środkiem” skali²⁹;

²⁷ Zob. przykładowo: H.-J. Hippler, N. Schwarz, *Response Effects in Surveys*, [w:] H.-J. Hippler, N. Schwarz, S. Sudman (eds) *Social Information Processing and Survey Methodology*, New York, Springer-Verlag, 1988, s. 105–122; J.-P. Gremy, *Wpływ sposobu sformułowania pytań kwestionariuszowych na uzyskiwane odpowiedzi*, ASK „Społeczeństwo, Badania, Metody”, 2; 1995, s. 7–19.

²⁸ D.A. Dillman, (2000), *Mail and Internet Surveys...*, s. 50–78 oraz D.A. Dillman, (1978), *Mail and Telephone Surveys...*; s. 97–117.

²⁹ Zob. także: P.B. Sztabiński, *Punkt „środkowy” w skalach bipolarnych: zamieszczać czy nie zamieszczać?*, ASK „Społeczeństwo, Badania, Metody”, 11, 2002, s. 45–81.

- unikać w pytaniach wartościujących określeń. Przykładowo jako przyczynę przemocy w szkole wskazuje się, obok: „szkoły” i „programów TV”, „nieodpowiedzialnych rodziców”;
- unikać pytań niezbalansowanych, tzn. pokazywać obydwie możliwe postawy, np. „Czy zgadza się, czy też nie zgadza ze stwierdzeniem, że . . .”;
- unikać pytań typu „wskazać wszystkie możliwe odpowiedzi”. Należy pytać o każdy punkt osobno, aby uniknąć efektu pierwszeństwa (*primacy effect*) lub, o ile punktów nie jest ich zbyt wiele, rangować³⁰;
- kategorie odpowiedzi muszą być rozłączne i z tego samego porządku. Przykładowo, w pytaniu o źródła informacji kateria nie może obejmować źródła i miejsca uzyskania informacji: Radio, TV, koledzy z pracy, w domu, w podróży itd.;
- stosować elementy wywiadu kognitywnego (*cognitive design techniques*) ułatwiające przypominanie faktów przeszłych. Polega to na „rozłożeniu” pytania na serię pytań prostych. Przykładowo, zamiast pytania: „Czy podczas ostatniej jazdy samochodem miał P. zapięty pas przez cały czas, prawie przez cały itd.”, należy pytać: „Kiedy ostatnio jechał P. samochodem?”, „Skąd – dokąd jechał P. samochodem?”, „Jak długa była to podróż?”, „Czy był kierowcą czy pasażerem na tylnym/przednim siedzeniu”, „Czy miał P. zapięty pas przez cały czas, prawie przez cały itd.”³¹;
- wskazać właściwe odniesienia czasowe (horyzont czasowy) w pytaniach o częstotliwość określonych zachowań. Raportowanie częstości występowania określonych zachowań musi uwzględniać ich rodzaj. Przykładowo, inną skalę częstości musimy zastosować w przypadku oglądania TV, inną pytając o jadenie w restauracji, a jeszcze inną w przypadku chodzenia do teatru. Należy także unikać odniesień typu „w ciągu ostatnich 3 dni”, ponieważ aktywność respondenta może być różna w różnych dniach tygodnia (np. poniedziałek – środa i weekend). To samo dotyczy określenia: „w tym roku” (czy chodzi o rok kalendarzowy czy ostatnie 12 miesięcy) oraz „ostatnio” (?);
- kontrolować merytoryczną poprawność. Przykładowo, pytanie o częstotliwość chodzenia do kościoła jest nieadekwatne dla Świadków Jehowy;
- formułować pytania w ten sposób, aby możliwe były porównania z innymi badaniami;
- unikać podwójnego przeczenia, tzn. formy przeczącej w pytaniu. Przykładowo, „Czy jest P. za czy przeciw, aby nie zabraniać . . .”;
- unikać pytań zawierających żądanie udzielenia jednej odpowiedzi na dwa różne pytania (tzw. *double barreled*). Przykładowo: „Czy powinno się budować nowe baseny i brodziki, które byłyby otwarte zimą?”. Pytanie to byłoby poprawne, gdybyśmy zamieścili kateria uwzględniającą odpowiedź dotyczącą basenów, odpowiedź dotyczącą brodzików oraz basenów otwartych zimą i brodzików otwartych zimą;
- minimalizować drażliwość pytań, ponieważ Respondent może je pominąć. Przykładowo, zamiast pytać o wysokość dochodów, lepiej jest zapytać o przedział, w którym one się mieszczą; zamiast pytać, czy Respondent „kiedykolwiek coś ukradł w sklepie”, lepiej zapytać, czy „wziął coś ze sklepu nie płacąc”;
- unikać pytań, które zmuszają Respondenta do zbędnych obliczeń, zbyt dużego wysiłku intelektualnego np. pytanie o dochód w gospodarstwie domowym na osobę (pytanie to wymaga od Respondenta zsumowania dochodów wszystkich członków gospodarstwa domowego oraz podzielenia wyniku tej operacji przez liczbę osób w gospodarstwie domowym).

³⁰ Innym rozwiązaniem, w przypadku takich pytań jest rotacja możliwych odpowiedzi. Nie eliminuje ona efektu porządku, ale ogranicza związany z nim błąd.

³¹ Zob. także: T. DeMaio, *Attempting to Validate Cognitive Testing Methodology*, European Survey Research Association Conference, June 29-July 3, Warsaw, 2009. Oznacza to w praktyce stosowanie pytań „prostych” (*unfolded*) w miejsce „złożonych” (*folded*).

Tego typu „wyliczanki” przedstawione przez Dillmana, zawsze budzą moje obiekcje. Po pierwsze dlatego, że można do nich „dodać” jeszcze wiele innych zaleceń. Przykładowo, czy w pytaniach rozstrzygnięcia, przedstawione odpowiedzi nie tworzą tzw. fałszywej dychotomii; czy w skalach zamieszczony jest punkt środkowy, o ile istnieje logiczny „środek”; czy w pytaniach kafeteryjnych kategorie odpowiedzi są wyczerpujące; czy stosowane skale nie są zbyt „czułe”, czy nie zawierają „fałszywego środka”, czy opisy jej krańców stanowią opozycje semantyczne, czy skala nie jest zbyt długa, a struktura pytania zbyt skomplikowana (np. seria stwierdzeń do oceny na skalach horyzontalnych) itd. Musimy także pamiętać, iż opis skali może także „dookreślać” treść pytania i użyte w nim terminy (np. częstotliwość czyszczenia zębów: „codziennie”, – chodzi o „mycie zębów” oraz „kilka razy w roku” – chodzi o „czyszczenie zębów przez stomatologa”).

Po drugie, dlatego, iż klóci się to, przynajmniej w moim odczuciu, z ogólną ideą TDM. Czym bowiem są powyższe reguły? Niczym innym niż zasadami sformułowanymi w języku „nakazowo-rozdzielczym”, przed którym „ostrzegał” sam Dillman. Dlatego też warto może powiedzieć nie o tym, co „należy robić”, ale czego „warto unikać”, a więc o błędach pytań³². Wyróżnić można pięć podstawowych źródeł błędów pytań. Są to:

- 1) niedostosowanie pytania do sytuacji Respondenta,
- 2) trudność pytania dla Respondenta,
- 3) niejasność pytania dla Respondenta,
- 4) drażliwość pytania dla Respondenta oraz
- 5) sugestie zawarte w pytaniu.

Ad 1. O pytaniach niedostosowanych do sytuacji Respondenta mówimy wtedy, gdy pytanie nie ma zastosowania do sytuacji wszystkich Respondentów i odpowiedź ma dwójaki sens. Przykładowo, pytanie: „Czy w trudnych sytuacjach może P. liczyć na pomoc kościoła lub związku wyznaniowego?”. Odpowiedź „nie” może oznaczać: „Nie należę do żadnego kościoła”, ale także „Należę do kościoła, ale nie mogę na niego liczyć”. Inny przykład niedostosowania pytania do sytuacji Respondenta, to pytanie o zamiar zmiany miejsca pracy w przypadku osób niepracujących. Podobna sytuacja występuje w przypadku pytania o preferowany poziom wykształcenia dzieci, zadawanego osobom bezdzietnym lub w gospodarstwach 1-osobowych.

Ad 2. Trudność pytania dla Respondenta może wynikać z kilku względów. Może być ona związana z jego długością (Respondenci odpowiadają wtedy tylko na część pytania) lub żądaniem dwóch lub więcej informacji w jednym pytaniu (np. nazwa uczelni, wydziału i kierunku, który Respondent ukończył). Innym źródłem trudności pytania dla Respondenta może być zbyt wysoki poziom jego ogólności. Często zdarza się, iż pytanie, na które badacz chce odpowiedzieć na podstawie badań, jest wprost zamieszczane w kwestionariuszu. Przykładem może tutaj być następujące pytanie: „Jak pan oceniał swoją reakcję emocjonalną na sytuację stresu przed szkoleniem? Czy ta reakcja może ulec zmianie w wyniku szkolenia?”. Musimy także pamiętać, iż osoby o niższym poziomie wykształcenia myślą konkretnymi i podkładają różne treści pod sformułowania zawarte w pytaniu, co pociąga za sobą ryzyko artefaktów. Kolejne źródło trudności pytań dla Respondentów to założenie, iż Respondenci posiadają wiedzę, której de facto nie posiadają. Jest to częsty błąd w badaniach konsumenckich, w których pyta się Respondenta o ocenę produktów, o których tylko słyszał lub widział ich reklamy. I wreszcie, przyczyną trudności pytań mogą być niezrozumiałe dla Respondenta terminy i określenia (nie można zakładać, że język używany w mediach jest zrozumiały dla wszystkich Respondentów). I wreszcie, nie jest wskazane formułowanie pytania w formie przeczącej, zwłaszcza przy ocenie stwierdzeń na skali słownej (podwójna negacja, przykładowo: „nie zgadzam

³² Rzeczą oczywistą jest, iż poniższe uwagi nie do końca są rozłączne z „zaleceniami” sformułowanymi przez Dillmana.

się, że nie jest to produkt niezdrowy”), podobnie jak używanie terminów obcojęzycznych (np. *full flavor*, *King Size*).

Ad 3. Kolejnym źródłem błędów jest niejasność pytania dla Respondenta. W wielu przypadkach wynika ono z niejasności zapotrzebowania informacyjnego badacza: jeśli sami nie wiemy, o co pytamy, to skąd Respondent ma wiedzieć, o co pytamy? Przykładem może być następujące pytanie: „Ile wynoszą przeciętne miesięczne P. wydatki związane z nauką dzieci? Proszę uwzględnić nie tylko opłaty szkolne, ale także koszt wszystkich zajęć dodatkowych, np. sportowych, kursów językowych, komputerowych itp.”. Czy Respondent odpowiadając na to pytanie powinien uwzględnić koszt podręczników, zeszytów itp., czy też nie? W efekcie istnieje ryzyko podkładania przez Respondentów różnych treści, co prowadzi do niehomogeniczności wyników. Niejako klasycznym przykładem niejasności może być poniższe pytanie: „Ludzie mają w życiu różne cele. Chcielibyśmy dowiedzieć się, jakie cele chciałby P. w najbliższym czasie osiągnąć – jakie ważne sprawy życiowe załatwić?”, w którym mówi się o „celach życiowych”, „celach, które Respondent chciałby w najbliższym czasie osiągnąć” oraz „ważnych spraw życiowych, które chciałby On załatwić?”. O co właściwie chodzi w tym pytaniu? Źródłem niejasności może być także niezręczność sformułowania pytania (tzw. „zamęt” językowy). Czy możliwe jest bowiem odczytanie Respondentowi poniższego pytania: „Jaki dochód miesięczny przy obecnych cenach chciałby P., aby P. rodzina miała na 1 osobę, aby zapewnić utrzymanie na minimalnym poziomie?”³³. Ponieważ nie jest to możliwe, ankieterzy będą „opowiadali” Respondentom to pytanie, każdy w inny sposób i w efekcie nastąpi destandaryzacja bodźca: uzyskamy jakieś odpowiedzi, ale nie wiadomo, na jakie pytania.

Ad 4. Innym źródłem błędu może być drażliwość pytania. Zasadniczo wyróżnić można dwie kategorie tych pytań: pytania skłaniające do udzielania społecznie akceptowanych odpowiedzi oraz pytania wzbudzające niepokój³⁴. Te pierwsze, to pytania o zachowania, posiadanie cech, poglądów, przedmiotów itp., w przypadku których normy społeczne określają, co jest pożądane, a co nie jest (np. choroby, zachowania nielegalne lub społecznie potępiane, wykształcenie, posiadanie dóbr, antysemityzm, kurtuazja wobec kobiet, higiena). W ich przypadku istnieje ryzyko uzyskania odpowiedzi zgodnej z normami. Z kolei pytania wzbudzające niepokój to pytania dotyczące spraw, które zgodnie z normami społecznymi nie nadają się do rozmowy (np. życie seksualne, śmierć bliskich, przemoc w rodzinie, posiadane oszczędności). W przypadku ankiet wypełnianych samodzielnie przez Respondenta, między innymi ze względu na wyższe poczucie anonimowości, problem drażliwości pytań i możliwości irytacji Respondenta schodzi „na dalszy plan”³⁵.

Ad 5. Ostatnie źródło błędów to sugestia zawarta w pytaniu. Musimy mieć świadomość, iż na dobrą sprawę sugerujące jest każde pytanie, choćby poprzez sposób postawienia problemu, w pytaniach kafeterijnych – poprzez sformułowanie i liczbę możliwych odpowiedzi (kafeteria „zamyka mentalnie” Respondentów, którzy ograniczają się do wyboru spośród możliwych odpowiedzi. Dlatego tak ważne jest, aby kafeteria była wyczerpująca). Podstawowym źródłem sugestii jest jednak użycie w pytaniu terminów wartościujących (istnieje wtedy ryzyko, że Respondenci uznają to za wskazówkę, które odpowiedzi są „właściwe” i wyboru dokonają, zgodnie z nią) oraz nierówna liczba punktów kafeterii, zawierających określenia pozytywne i negatywne. Jednym z najbardziej niebezpiecznych błędów związanych z sugestią jest jednak ukryte w pytaniu założenie, które odzwierciedla system wartości i przekonania autora pytań.

³³ Warto tutaj podkreślić, iż nie jest to pytanie wymyślone: zostało one sformułowane przez badacza z wieloletnim stażem.

³⁴ Zob. P. B. Sztabiński, *Ankieterzy i ich respondenci. Od kogo zależą wyniki badań ankietowych?* Warszawa, Wydawnictwo IFIS PAN, 1997, s. 80–82.

³⁵ Tradycyjnie, za pytania drażliwe uznaje się również pytania o wiedzę. W przypadku ankiet, ze względu na ich specyfikę, zadawanie tego typu pytań mija się z celem. Respondent, jeśli nie posiada wiedzy na dany temat, bez problemu może ją uzyskać (sprawdzić w Internecie, książkach, zapytać itd.).

W zdecydowanej większości przypadków Respondenci zaakceptują to założenie i zgodnie z nim udzielają odpowiedzi. Przykładowo, w pytaniu: „W naszym społeczeństwie istnieje wiele grup ludzi bogatych. Są to wybitni naukowcy, gwiazdy sportu, wybitni artyści, biznesmeni. Bogactwo której z tych grup budzi najczęściej niechęć w naszym społeczeństwie?”, zawarte są dwa założenia: pierwsze, iż grupę ludzi bogatych stanowią naukowcy, sportowcy, artyści i biznesmeni; drugie, iż ich bogactwo budzi niechęć (pomijam tutaj kwestię wartościujących sformułowań: wybitni naukowcy, gwiazdy sportu, wybitni artyści vs. biznesmeni). Mówiąc o sugestii zawartej w pytaniu pamiętać również należy o jego zbalansowaniu, co daje równą szansę wyboru możliwych odpowiedzi.

Należy tutaj także wspomnieć o tzw. błędzie multiplikacji, a więc stosowaniu zbitek pojęciowych (np.: „stosunek do religii i Kościoła”), co może powodować, iż respondenci będą oceniać jeden z elementów lub będą je rozdzielać.

Ułożenie pytań nie jest równoznaczne z przygotowaniem kwestionariusza. Nie jest on bowiem zbiorem pytań w przypadkowej kolejności. Należy pamiętać, iż właściwa konstrukcja kwestionariusza, ale także i jego grafika, mogą redukować błąd braku odpowiedzi (chodzi o *unit-nonresponse*) oraz błąd pomiaru. Można to interpretować również w kategoriach minimalizacji kosztów ponoszonych przez Respondenta oraz maksymalizacji nagród. W kwestionariuszu pytania muszą być ułożone według określonego „klucza” i stąd można mówić o strukturze kwestionariusza. Porządek ten uzależniony jest od: tematyki pytań, logicznych związków między pytaniami oraz psychologicznej dynamiki interogacji. Stąd też często mówi się o:

- 1) tematycznej,
- 2) logicznej i
- 3) psychologicznej strukturze kwestionariusza.

Ad 1. Tematyczny układ kwestionariusza określa kolejność poruszanych tematów, tzn. kolejność bloków pytań (modułów). W poszczególnych badaniach bloki te mogą mniej lub bardziej różnić się między sobą. Generalnie jednak, praktycznie w każdym kwestionariuszu wyróżnić można dwie podstawowe jego części: część zasadniczą (merytoryczną) oraz tzw. „metryczkę”.

Część „merytoryczną” kwestionariusza tworzą pytania, które bezpośrednio związane są z problemem badania i jego przedmiotem. Są one zgrupowane w moduły tematyczne, które czasami są ze sobą „luźno” powiązane (w takim przypadku, w kwestionariuszu zamieszczone zostają tzw. „łagodne przejścia” lub „wprowadzenia”, a więc tekst informujący Respondenta, iż zakończyliśmy jeden temat, a obecnie przechodzimy do innego). W ramach każdego modułu znajdują się baterie pytań. Pomiedzy poszczególnymi pytaniami i bateriami pytań powinna być zachowana ciągłość tematyczna. Tzw. „skakanie” po tematach (nawet w ramach baterii pytań) bardzo utrudnia Respondentowi udzielanie odpowiedzi i prowadzi do zmęczenia i znużenia wywiadem (są to dodatkowe „koszty” przez niego ponoszone). Musimy mieć także świadomość, że „układ” kwestionariusza, a więc kolejność zadawanych modułów, czy bloków może mieć wpływ na uzyskane rezultaty³⁶.

Drugą stałą częścią kwestionariusza jest tzw. „metryczka”, a więc zestaw pytań o podstawowe cechy społeczno-demograficzne Respondenta (ewentualnie współmałżonka lub innych członków gospodarstwa domowego) oraz pytań dotyczących gospodarstwa domowego, np.: o dochody, wyposażenie gospodarstwa domowego w sprzęt. Ich zestaw oraz rodzaj będzie oczywiście inny w przypadku badania

³⁶ Przykładowo, w badaniach prasowych, jeśli moduł dotyczący tygodników zamieszczony został po module dotyczącym miesięczników, to ze względu na większą częstość ich ukazywania się, stwierdzono większą skłonność Respondentów do częstszego ich wskazywania. Jeśli natomiast moduł „miesięczników” zamieszczony został po „tygodnikach”, to z racji ich rzadszego ukazywania się, wskazywane były one rzadziej. Zob.: F. Dalbard-Martin, *Wpływ struktury kwestionariusza na uzyskiwane odpowiedzi w badaniach prasowych*, ASK „Społeczeństwo, Badania, Metody”, 2(4), 1996, s. 63–70.

instytucji. W takim przypadku nie będzie nas interesował wiek czy wykształcenie Respondenta, tym bardziej iż czasami jest ich kilku (np. w nieco większych instytucjach, informacji finansowych, ekonomicznych czy kadrowych udzielają zazwyczaj różne osoby). W takim przypadku interesuje nas zazwyczaj jedynie stanowisko, ewentualnie staż pracy informatorów. Mówiąc o metryczce pamiętać musimy o dwóch sprawach: po pierwsze, zestaw informacji winien być wyznaczony przez pytania problemowe (naczelne i szczegółowe) oraz plan analiz zebranego w badaniach materiału. Dlatego też jakiegokolwiek braki w tej części kwestionariusza powodują, że nie jest możliwa analiza odpowiedzi według np. wieku Respondenta, jego wykształcenia itd. lub wielkości instytucji czy jej struktury. Pytania te są zatem niezwykle ważne. Po drugie, biorąc pod uwagę powyższe względy, a dodatkowo to, że dla niektórych Respondentów część pytań „metryczkowych” może być drażliwa, dla innych zaś „bezzasadna”, przechodząc do tego bloku pytań należy wyjaśnić celowość ich zadawania. Stąd też, w kwestionariuszu należy zamieścić tzw. „wprowadzenie do metryczki”. W ten sposób „chronimy” Respondenta przed dociekaniem typu „po co te pytania?”, „co to ma wspólnego z tematyką badania?” itp.

Ad 2. Struktura logiczna kwestionariusza jest przede wszystkim określana przez związki logiczne między pytaniami danego bloku pytań. Generalnie chodzi tutaj o tzw. efekt porządku (*order effect*), a więc wpływ odpowiedzi na dane pytanie, na odpowiedź na inne, w szczególności następne. W zależności od kolejności zadawania dwóch powiązanych pytań możemy uzyskiwać różne odpowiedzi. Wymienić należy tutaj cztery podstawowe rodzaje efektu porządku: efekt normy bezstronności (*norm of evenhandedness*), efekt zakotwiczenia, (*anchoring effect*), efekt dodania (*addition effect*) i cofnięcia (*subtraction effect*) oraz efekt kolejności pytań ogólnych i szczegółowych (*part – whole/general-specific*)³⁷.

Efekt normy bezstronności (czasami nazywany efektem wzajemności) po raz pierwszy zaobserwowali Herbert H. Hyman i Paul B. Sheatsley w badaniach z 1948 r., zadając pytanie o to, czy powinno się zezwolić komunistycznym dziennikarzom relacjonować wizytę w Ameryce, w sposób, jaki ją odbierają. Jeśli to pytanie zadawane było przed analogicznym, ale dotyczącym dziennikarzy amerykańskich i ich pobytu w Związku Radzieckim, to 37% respondentów udzieliło odpowiedzi twierdzącej. Jeśli jednak to pytanie zadawane było jako drugie, to odsetek ten wyniósł 73%. Efekt ten wyjaśnia się tym, że respondenci odwołują się do poczucia „sprawiedliwości” i udzielając odpowiedzi na drugie z pytań, biorą pod uwagę swoje odpowiedzi na pierwsze pytanie³⁸.

Efekt zakotwiczenia polega na tym, że odpowiedź na jedno pytanie służy jako „kotwica” przy udzielaniu odpowiedzi na inne. Efekt ten zaobserwowała Roberta L. Sangster, zadając studentom dwa pytania. Pierwsze dotyczyło opinii na temat tego, czy oszustwa są powszechnym problemem na ich uniwersytecie, drugie natomiast, czy oszustwa są powszechnym problemem na uniwersytetach w Stanach Zjednoczonych. Jeśli pytanie to zadawane było jako pierwsze, odsetek studentów zgadzających się z tym stwierdzeniem wyniósł 65% (dla ich Uniwersytetu: 60%). Odsetek ten znacząco zmniejszył się, jeśli pytanie to było zadawane po pytaniu o ich własny uniwersytet (49% vs. 44%).

Dwa dalsze rodzaje efektu porządku, to efekt dodania (*addition effect*) oraz efekt cofnięcia (*subtraction effect*), które czasami określa się jako efekt asymilacji i kontrastu. Zostały one opisane przez Howarda Schumana i Stanleya Pressera w 1981 r. Stwierdzili oni, że w przypadku pytania szczegółowego o zadowolenie z małżeństwa oraz pytania ogólnego o zadowolenie z życia, odsetek odpowiedzi na to ostatnie pytanie różnił się w zależności od tego, czy było zadawane jako pierwsze, czy jako drugie. O ile 52%

³⁷ Zob. D.A. Dillman, *Mail and Internet Surveys...*, 2000, s. 88–91.

³⁸ Jeśli kolejność była odwrotna, tzn. najpierw zadawano pytanie o dziennikarzy amerykańskich, a potem rosyjskich odsetki odpowiedzi twierdzących wyniosły odpowiednio: 90 i 66%. Za: H. Schuman, *Method and meaning in polls and surveys*, Cambridge, Harvard University Press; 2008, s. 93.

Respondentów wyrażało zadowolenie z życia, jeśli pytanie to było zadawane jako drugie (po pytaniu o zadowolenie z małżeństwa), o tyle tylko 38% udzielało takiej odpowiedzi, jeśli porządek był odwrócony. Wyjaśnia się to tym, że ludzie dalej myślą o małżeństwie (70% stwierdzało, że są bardzo zadowoleni), odpowiadając na pytanie ogólne. Odwrotny efekt (efekt cofnięcia/kontrastu) opisali Robert Mason, John E. Carlton i Roger Tourangeau³⁹. Zaobserwowali oni, że odsetek odpowiedzi wskazujących na poprawę sytuacji ekonomicznej stanu, który zamieszkują Respondenci, jest znacząco wyższy, gdy pytanie to jest zadawane przed pytaniem o ocenę przyszłej lokalnej (nie całego stanu) sytuacji ekonomicznej. Tłumaczą oni to tym, że udzielając odpowiedzi na drugie pytanie, Respondenci nie uwzględniają informacji, na których bazowała odpowiedź na pytanie pierwsze.

W wielu badaniach zaobserwowano również wzrost odsetka ocen pozytywnych w pytaniach ogólnych, gdy były zadawane po serii pytań szczegółowych na ten sam temat (*part – whole/general-specific questions effect*).

Występowanie efektu porządku (w różnych jego „odmianach”) wskazuje, że konstruując kwestionariusz nie możemy ignorować istniejących między pytaniami związków logicznych. Jednocześnie efekt porządku pytań nie jest możliwy do wyeliminowania. Rotacja modułów pytań minimalizuje jedynie prawdopodobieństwo wystąpienia błędu systematycznego, który polega na jednokierunkowym wypaczeniu wyniku badania spowodowanym kolejnością zadawanych pytań. Jedyne możliwe rozwiązanie to kontrola wyników rotowanych pytań zadawanych losowo dobranym Respondentom.

Innym przykładem związków logicznych są relacje między odpowiedziami na zadawane w tzw. „pionowy” i „poziomy” pytania tabelaryczne. Inne bowiem wyniki uzyskamy pytając najpierw o ocenę, przykładowo, jednego produktu ze względu na dane cechy, a następnie drugiego, ze względu na te same cechy i zupełnie inne, gdy będziemy pytać o ocenę, ze względu na daną cechę, najpierw jednego, a potem drugiego produktu itd. Oznacza to, iż zmieniając procedurę pytania wpływamy na wynik badania.

Mówiąc o związkach logicznych między pytaniami pamiętać należy także o tzw. pytaniach zależnych i wzajemnie kontrolujących się (tzw. *Tally Questions*). Chodzi tutaj o pytania dotyczące na przykład liczby osób w gospodarstwie domowym i składu gospodarstwa (liczba osób wymienionych w składzie musi być zgodna z liczbą osób w gospodarstwie), czy pytania o zarobki Respondenta, jego dochód i dochód gospodarstwa domowego (zarobki Respondenta mogą być co najwyżej równe jego dochodowi, a ten może być co najwyżej równy dochodowi gospodarstwa domowego). Respondent udzielając odpowiedzi na takie pytania oczywiście może się pomylić, jednakże w ich przypadku, w skrypcie, przewidzieć należy ich kontrolę logiczną, a więc kontrolę ich spójności. Dotyczy to zresztą nie tylko pytań metryczkowych. Mówiąc o pytaniach wzajemnie się kontrolujących nie można pominąć problemu pytań powtórzonych, a więc zadawanych dwukrotnie w różnych miejscach kwestionariusza, których celem jest „sprawdzenie” wiarygodności Respondenta. Ich zadawanie wydaje się niecelowe, przynajmniej z dwóch względów. Po pierwsze, co zrobić, jeśli się uzyska dwie różne informacje (jest to zupełnie prawdopodobne ze względu na efekt kontekstu: Respondenci mogą podkładać inny sens pod dwa takie same pytania) – zadać pytanie ponownie? Po drugie, Respondent może się zirytować, widząc, że chcemy sprawdzić Jego wiarygodność.

³⁹ R. Mason, J.E. Carlson, R. Tourangeau, *Contrast effects and subtraction in part-whole questions*, „Public Opinion Quarterly”, 58, 1994, s. 569–578. Wcześniej te dwa efekty opisali: N. Schwarz, F. Strack, H.-P. Mai, *Assimilation and contrast effects in part-whole question sequences: a conversational logic analysis*, „Public Opinion Quarterly” 55, 1991, s. 3–23. Autorzy ci używają jednak terminów: efekt asymilacji (jeśli informacja ta jest użyta do formułowania opinii na następne pytanie) oraz efekt kontrastu (jeśli informacja ta jest celowo pomijana przy formułowaniu opinii).

Ad 3. Struktura psychologiczna kwestionariusza określana jest przez wszystkim przez psychologiczną funkcję pytań. Niektóre pytania w kwestionariuszu pełnią określoną funkcję. I tak, wskazać można tzw. pytania „uczące” (lub „rozbiegowe”), których celem jest przygotowanie rozmówcy do roli Respondenta, ale i zachęcenie do udziału w badaniu. Pytania takie zamieszcza się zawsze na początku kwestionariusza i z reguły są one bardzo proste, interesujące i dotyczą wszystkich respondentów i spraw im bliskich. Wyznacza je zasadniczo list zapowiedni i tematyka badania. Wybór pierwszego pytania jest sprawą niezwykle istotną ze względu na odsetek zwrotów.

W dalszej części kwestionariusza można zamieszczać pytania trudne i bardziej skomplikowane formalnie (nie będą one stanowiły utrudnienia, jeśli Respondent nawet wypełnia kwestionariusz „w częściach”). W końcowej części zamieszczamy oczywiście pytania metryczkowe (Respondenta lub instytucji). W przypadku trudnych i długich kwestionariuszy stosuje się czasami tzw. pytania „watę”. Są to pytania z reguły „luźno” związane z tematyką wywiadu, łatwe i przyjemne dla Respondenta, np. „plany na przyszłość”, które mogą być interesujące (podtrzymują „motywację”) zarówno Respondenta indywidualnego, jak i instytucjonalnego.

Niezwykle istotnym elementem struktury psychologicznej kwestionariusza są wspomniane tzw. wprowadzenia („Na początek chcielibyśmy dowiedzieć się, ...”) i „łagodne” przejścia między modułami lub bateriami pytań („Obecnie chcielibyśmy zapytać o „Przejdźmy teraz do . . .”). Mają one na celu ułatwić Respondentowi przedstawienie się na myślenie o poruszanej w danym bloku tematyce i czynią ankietę bardziej „przyjazną”, „naśladując” styl narracyjny.

Przygotowując kwestionariusz „na papierze” musimy pamiętać, że zawarte w nim pytania muszą być, jak wskazałem, przemyślane i być efektem długotrwałej pracy, a ich układ, kolejność i brzmienie nie może być przypadkowe. Przygotowanie kwestionariusza nie jest i nie może sprowadzać się do operacji typu: „Ctrl+C” i „Ctrl+V”. Podobnie, żadne pytanie w kwestionariuszu nie jest mniej lub bardziej „ważne”.

Nie mniej istotna jest pierwsza i ostatnia strona kwestionariusza. Pierwsza strona powinna motywować Respondenta do wypełnienia kwestionariusza. Przede wszystkim musi być rozpoznawalna i działać jak tytuł książki, a więc zawierać logo, nazwę i adres instytutu realizującego badanie, dalej tytuł badania oraz jego logo (element graficzny identyfikujący badanie). Generalnie, musi ona mieć związek z treścią listu wprowadzającego. Z kolei ostatnia strona winna zawierać podziękowanie za wypełnienie kwestionariusza oraz prośbę o dodatkowe uwagi.

Kwestionariusz w komputerze

Powyzsze „zasady” tworzenia kwestionariusza papierowego obowiązują również rzecz jasna przy tworzeniu kwestionariusza w komputerze. Wcześniej wspominałem bowiem, iż zdaniem Dillmana technika ankiety „papierowej” nie różni się zasadniczo od ankiety internetowej, ale . . . Podobnie jest w przypadku kwestionariusza. Kwestionariusz w komputerze nie powinien różnić się od kwestionariusza na papierze, ale . . . pamiętać należy, iż tworzenie kwestionariusza w komputerze to połączenie języka pisanego i języka graficznego w taki sposób, aby wzajemnie się wspomagały i minimalizowały wysiłek Respondenta. Tworząc kwestionariusz w komputerze (w formie e-maila lub załącznika do niego), musimy pamiętać jeszcze o trzech dodatkowych sprawach:⁴⁰

- 1) określeniu jednolitego sposobu odczytywania wszystkich informacji i pytań zawartych w kwestionariuszu (tzw. *navigational path*), tzn. zasad: (i) pisania pytań w ten sposób, aby minimalizować konieczność powtórznego ich odczytywania w celu zrozumienia zadania (np. w pytaniach kafeteryj-

⁴⁰ Por.: D.A Dillman, *Mail and Internet Surveys...*; s. 96–134.

- nych: wprowadzenie, zadanie, możliwe odpowiedzi); (ii) instrukcje szczegółowe należy zamieszczać w miejscu, gdzie są one konieczne, nie zaś w liście wprowadzającym lub instrukcji na początku kwestionariusza; tam znaleźć się mogą jedynie ogólne uwagi dotyczące sposobu wypełniania ankiety (zaznaczania odpowiedzi np. poprzez wstawienie symbolu 'x' w przygotowanych nawiasach - [x]); (iii) pytania (stwierdzenia) z tymi samymi kategoriami odpowiedzi należy zamieszczać w tabeli (itemów nie może być jednak zbyt wiele) i należy pamiętać, iż w takim układzie związek między nimi jest dla Respondenta silniejszy (efekt porządku). Jeśli jest ich więcej, należy je rotować; (iv) nie zadawać dwóch pytań „zależnych” równocześnie (np.: jak ważna jest ta usługa i jeśli ważna, to jak często z niej korzysta się); (v) należy ograniczać liczbę „matryc”, tzn. złożonych pytań tabelarycznych (np. tabela gospodarstwa domowego); (vi) wprowadzić opcję odpowiedzi; „nie wiem/trudno powiedzieć”⁴¹;
- 2) stworzeniu jednolitego systemu znaków i symboli graficznych ułatwiających respondentom poruszanie się po kwestionariuszu i zapewniających poprawne wypełnienie kwestionariusza (np. poprzez: powiększanie rozmiaru pewnych elementów treści pytania, zmiana ich koloru; zwiększenie białych przestrzeni między elementami; markowanie elementów). Tutaj także muszą obowiązywać pewne zasady. Należy między innymi: (i) zaczynać pytanie w lewym górnym rogu (najważniejsza informacja), najmniej ważne – zamieszczać w prawym dolnym rogu); (ii) używać większych znaków lub kolorowych wyróżników dla identyfikacji punktu „START” lub punkt ten (np. numer pytania) oznaczać następująco: **1**); (iii) używać tych samych symboli graficznych dla identyfikacji kolejnego pytania; np.: **1**, **2**, **3** itd.; (iv) stosować „prostą” numerację pytań wg ich kolejności (np. **1**, **2**, **3** itd., zamiast np.: Q.1. lub 1.1., 2.1. itp.); (v) używać stałego formatu „zmuszającego” Respondenta do odczytania całego pytania; (vi) należy ograniczać stosowanie „odwróconego” druku; (vii) białe przestrzenie między pytaniami należy utrzymywać większe niż między jego elementami; (viii) używać innej czcionki (np. **bold**) dla pytań, innej dla odpowiedzi; (ix) instrukcje zamieszczać w treści pytania, nie zaś w ramkach; (x) instrukcje dla Respondenta pisać inną czcionką (np. *italic*); (xi) nie należy zamieszczać instrukcji w osobnym dokumencie; (xii) używać jasnych kolorów jako tła w polu tekstowym (o ile chcemy je „zamarkować” kolorem); (xiii) jeśli tekst pytania jest edytowany na kolorowym tle, pola odpowiedzi winny zostać białe; (xiv) możliwe odpowiedzi winny być edytowane w układzie pionowym, nie zaś poziomym, dwukolumnowym; (xv) miejsce na zaznaczenie odpowiedzi można zostawić po lewej lub prawej stronie⁴²; (xvi) używać pudełek dla zaznaczania odpowiedzi; (xvii) zachować „poziomy” układ pytań, nie zaś dwukolumnowy; (xviii) w skalach graficznych utrzymywać jednakowe interwały między punktami skali; (xix) utrzymywać ten sam kierunek skal we wszystkich pytaniach (nie stosować alternacji skal); (xx) nie zajmować całego obszaru strony tekstem pytania (wiersze winny być krótsze, większy prawy margines) ze względu na ewentualną konieczność śledzenia długiego wiersza (opcja „zawijania” może „rozformatować” plik);
- 3) stworzeniu dodatkowego systemu wizualnych reguł nawigacji, tzn. wprowadzenie strzałek kierunkowych: poziomych, pionowych i „łamanych” (np.: „→”, „↓” i „↘” lub „↙”, „↗” oraz „↖”), zamiast tradycyjnych instrukcji „pisanych” (przejdź do pyt. XX). Może to oznaczać również wprowadzenie innego **kroju czcionki**, w celu wizualnego wyodrębnienia pewnych fragmentów (nie mogą to jednak być zbyt duże fragmenty)⁴³.

⁴¹ Zalecenie to wydaje się dyskusyjne, ponieważ może „prowokować” Respondentów do zaznaczania takiej właśnie odpowiedzi i może istotnie wpłynąć na rozkłady odpowiedzi „znaczących”. Dlatego też innym rozwiązaniem może być zamieszczenie w instrukcji informacji, że jeśli Respondent nie potrafi udzielić odpowiedzi, winien wpisać to w pole tekstowe: „inna odpowiedź”.

⁴² Lepszym rozwiązaniem wydaje się zaznaczanie po „lewej” stronie, ponieważ poszczególne odpowiedzi mogą mieć „różną długość” i tym samym byłaby różna odległość od końca odpowiedzi do „pudełka”, w którym należy zaznaczyć odpowiedź.

⁴³ Stosując powyższe zalecenia należy jednak pamiętać o jednej podstawowej sprawie: ponieważ sugestii przygotowania pytań „na papierze” i „w komputerze” jest bardzo wiele, nie można ich mechanicznie „łączyć”. Może się bowiem zdarzyć, iż zamiast ułatwić Respondentowi wypełnienie kwestionariusza, utrudnią tę czynność.

Podobnie jak w kwestionariuszu na papierze, bardzo ważna jest także pierwsza i ostatnia strona kwestionariusza oraz to, co „obowiązuje” w przypadku kwestionariusza papierowego – zamieszczenie informacji, dotyczącej sposobu zwrotu wypełnionej ankiety.

Kwestionariusz w Internecie

Opisane powyższe „zasady” tworzenia kwestionariusza w komputerze (oraz na papierze) obowiązują także przy tworzeniu kwestionariusza w Internecie. Kwestionariusz w Internecie nie powinien różnić się od kwestionariusza w komputerze, ale ... pamiętać należy, iż tworzenie kwestionariusza w Internecie to połączenie języka pisanego, języka graficznego oraz języka programowania (z uwzględnieniem różnic), w taki sposób, aby wzajemnie się wspomagały i minimalizowały wysiłek Respondenta. Tworząc kwestionariusz w Internecie, musimy pamiętać przede wszystkim o tym, że z jednej strony, w porównaniu z badaniami e-mailowymi, są one wbrew pozorom prostsze w przygotowaniu, ponieważ nie mają tylu ograniczeń. Internet stwarza możliwości dodania koloru, umożliwia wirtualną interakcję między nadawcą i respondentem, skomplikowane reguły przejścia, wyświetlanie instrukcji przy poszczególnych pytaniach w formie „pop-up box”, rozwijane listy (*drop-down boxes*), zamiast pytań otwartych (np. w pytaniach o miejsce zamieszkania: ⇒ województwo ⇒ powiat ⇒ gmina ⇒ miejscowość), obrazy (do obejrzenia) można skanować, można załączać animacje, wideoklipy lub odwołania do stron WWW. Jednakże z drugiej strony, należy pamiętać o różnicach sprzętowych (różnice w Hardware i Software: przede wszystkim różne systemy operacyjne, różne przeglądarki internetowe oraz zróżnicowana rozdzielczość monitorów), w efekcie czego, na ekranie może być widoczny jedynie fragment wyświetlanej strony. Chodzi o wspomniany wcześniej problem kto? i co? widzi na monitorze komputera. Tego niestety nie jesteśmy w stanie przewidzieć. Różnice te mogą sprowadzać się do różnic w kolorach, dystansie między punktami w skalach graficznych, rozmieszczeniem tekstu na ekranie itp.

Bez wątpienia, część ogólnych zasad przygotowania kwestionariusza „papierowego” oraz „mailowego” ma też zastosowanie w przypadku kwestionariusza Webowego. Chodzi tutaj przede wszystkim o to, aby format pytań był zbliżony do formatu w kwestionariuszu papierowym (np. odpowiedni sposób numerowania pytań, zamieszczanie odpowiedzi z prawej strony itd.). Pamiętać także należy o graficznych (wizualnych) wyróżnikach pytań i odpowiedzi, także ich numeracji. Aczkolwiek „początek” ankiety Webowej winien być zupełnie inny. Respondent dostaje (w liście zapowiednim) adres strony internetowej, nie zaś kwestionariusz ankiety. Na stronie tej musi znajdować się ekran wprowadzający (tzw. *Welcome screen*), na którym znajduje się list wprowadzający i instrukcja, jak przejść do następnej strony (w związku z tym list musi być krótki i mieścić się na jednym ekranie). Na pierwszej stronie nie zamieszcza się pytań. Można natomiast zamieścić pytanie „identyfikujące” (potwierdzające) płeć Respondenta, aby kolejne strony wyświetlały się w odpowiedniej wersji językowej. Ponieważ musimy mieć kontrolę nad tym, kto wypełnił ankietę, a kto nie, Respondentom dostarcza się numer identyfikacyjny (kod PIN lub hasło) lub prosi się o ustalenie hasła dostępu (chodzi o ograniczenie dostępu do strony i „dublowanie” wypełnionych ankiet)⁴⁴. Aby przejść dalej, należy wpisać kod PIN lub hasło. Zasady przygotowania kolejnych stron są zbliżone do tych wspomnianych przy e-mailingu i są one następujące:⁴⁵ (i) pierwsze pytanie powinno być łatwe (nie może zawierać np. rozwijanej listy), interesujące, dotyczyć wszystkich Respondentów i nawiązywać bezpośrednio do problematyki badania. Musi także zawierać instrukcję, jak odpowiadać (np.: „Proszę zaznaczyć wybraną odpowiedź („kliknąć”). Jeśli się P. pomyli, proszę zaznaczyć właściwą – poprzednia

⁴⁴ W przypadku e-mailingu jest to zbędne: Respondenta możemy identyfikować na podstawie adresu pocztowego.

⁴⁵ Por.: Dillman D.A., *Mail and Internet Surveys...*; s. 377–399.

zniknie"). Pierwsze pytanie wraz z instrukcją musi się w całości mieścić się na jednym ekranie; (ii) należy ograniczać użycie kolorów, aby nie utrudniały czytania (kolory na monitorze Respondenta mogą być zupełnie inne niż na naszym). Dotyczy to także różnego rodzaju kształtów: kwestionariusz nie jest choinką, powinien być przejrzysty. Generalnie, kwestionariusz na stronie powinien być edytowany czarnymi literami na białym tle. Szare pola można przewidzieć dla zapisu wypowiedzi tekstowych lub „rozwijanej listy”. Kolory mogą być natomiast używane na pierwszej i ostatniej stronie kwestionariusza; (iii) należy unikać pytań, które ze względu na konfigurację systemu, różną rozdzielczość ekranu mogą różnić się wyglądem lub wymagają przewijania w poziomie lub pionie (stąd kwestionariusz należy testować na różnych komputerach i różnych monitorach); (iv) instrukcje mówiące, jak należy odpowiadać na pytania, zmieniać omyłkową odpowiedź, jak rozwinąć okno itp., mogą mieć postać zamkniętą („Kliknij tutaj, aby ...” i pojawia się okienko) lub otwartą (okienko samo pojawia się obok pytania). Ten ostatni rodzaj jest szczególnie przydatny w przypadku instrukcji merytorycznych. Tak jak w przypadku e-mailingu, niewskazane jest natomiast zamieszczanie osobnej instrukcji, którą respondent musi przeczytać, aby zacząć wypełniać kwestionariusz⁴⁶; (v) nie używać zbyt często rozwijanych list (*drop-down boxes*) i w każdym z nich (w *boxie*) zamieścić instrukcję „kliknij tutaj”, a nie kraniec skali, np. „zdecydowanie się zgadzam” (może być to źródłem sugestii lub błędu Respondenta, który widząc odpowiedź, może pomyśleć, iż na dane pytanie udzielił już odpowiedzi). Rozwijane listy są wskazane przy ocenie dużej liczby stwierdzeń, które muszą być zamieszczone na jednej stronie, ale zawsze należy rozważyć przyjęcie formatu ankiety papierowej (pytanie tabelaryczne, ze skalą w „główce”). Bezwzględnie nie należy ich używać przy odpowiedziach „tak-nie”, ponieważ zmuszają tylko do wykonania dodatkowych operacji; (vi) zachęcać, nie zaś zmuszać Respondenta do odpowiedzi na każde pytanie (chodzi o blokowanie okien lub ekranów, o ile nie jest zaznaczona odpowiedź na pytanie). Często mówi się, że jest to jedna z podstawowych zalet technik wspomaganych komputerowo. Jednocześnie, ze względów etycznych nie można zmuszać Respondenta do udzielenia odpowiedzi; ze względów praktycznych zaś, „lepsz” jest *item nonresponse*, niż *unit nonresponse* (jeśli Respondent nie będzie mógł „przejsć dalej”, wyloguje się i zakończy wypełnianie kwestionariusza). Dlatego, jak wspominałem, należy zamieszczać opcje „Nie wiem”, „Odmawiam odpowiedzi” lub pole tekstowe na jej wpisanie. Nie oznacza to jednak, iż nie należy informować Respondenta o jego pomyłkach (tzw. *summary screen*). Tego typu „interaktywne” pytania są wskazane, przede wszystkim wymagana jest wewnętrzna spójność odpowiedzi (tzw. *Tally Questions*); (vii) reguły przejścia czasami konstruuje się w ten sposób, że po wybraniu odpowiedzi przechodzi się automatycznie do właściwego pytania, co uniemożliwia przeczytanie i zaznaczenie odpowiedzi na pytania pominięte. Inny rodzaj formatowania, to tworzenie hiperlinków (respondent musi kliknąć na „regułę przejścia”). W tym wypadku instrukcja powinna brzmieć „Po zaznaczeniu odpowiedzi, KLKNIJ TUTAJ, aby przejść do pyt. XX.” (tego typu rozwinięta forma reguły przejścia jest niezbędna, ze względu na to, że Respondenci niekiedy zapominają zaznaczyć odpowiedzi i od razu „klikają” na link); (viii) jeśli liczba możliwych odpowiedzi nie mieści się na jednym ekranie, należy je zamieścić w dwóch kolumnach i zamieścić we wspólnej ramce (w ten sposób wskazujemy Respondentowi, iż jest to jedno pytanie); (ix) wskazane jest zamieszczenie graficznej informacji (w postaci „paska postępu” – *Progress Bar*) wskazującej, na jakim etapie wypełniania kwestionariusza jest Respondent. Nie jest to konieczne przy kwestionariuszu przewijanym (pokazuje to „pasek przewijania”); (x) należy ograniczać liczbę pytań typu „zaznaczyć wszystkie odpowiednie” (chodzi o minimalizację efektu pierwszeństwa – *primacy effect*). Dlatego wskazane jest „zatomizowanie” takiego pytania na serię pytań dychotomicznych: „tak – nie”. Podobnie, ograniczać należy liczbę pytań otwartych, ze względu na brak możliwości sondowania i dopytywania. Jeśli jednak decydujemy się z nich skorzystać, nie należy ograni-

⁴⁶ Innym rozwiązaniem jest zamieszczenie instrukcji w zakładce „FAQ” lub „Help”.

czać pola zapisu: „ramka” powinna automatycznie dopasowywać się do długości tekstu; (xi) w „górnym”, bądź „bocznym menu” należy oprócz standardowych przycisków (FAQ, Help, „do przodu – do tyłu” itp.) przewidzieć przycisk „ZAPISZ”. Jest to bardzo ważne, ponieważ nie zmusza się w ten sposób Respondenta do wypełnienia od razu całej ankiety; (xii) ostatnia strona zawierać podziękowanie za wypełnienie kwestionariusza, prośbę o ewentualne uwagi, ale przede wszystkim o kliknięcie „WYŚLIJ”.

Ostatnia, ale nie najmniej ważna sprawa związana z przygotowaniem kwestionariusza, to decyzja, czy ma być on skonstruowany na zasadzie „ekran po ekranie” (każde pytanie wyświetla się na osobnym ekranie), czy też ma być „przewijany” i dalej, czy mamy umożliwić Respondentowi zapoznanie się z nim w całości, przed udzieleniem odpowiedzi, czy też nie? Biorąc pod uwagę założenia *Tailored Design Method*, Respondent winien mieć możliwość przejrzenia całego kwestionariusza, pytanie po pytaniu. Skoro tak, to może być on „przewijany” od początku do końca i z powrotem. Inna możliwość to zastosowanie przycisków funkcyjnych (po ich kliknięciu przechodzi się do następnego pytania lub poprzedniego). Pierwsze rozwiązanie wydaje się lepsze, ponieważ bardziej przypomina sposób poruszania się w Internecie, aczkolwiek trudno jest jednoznacznie stwierdzić, jaki wpływ na odpowiedzi mają te dwie różne konstrukcje kwestionariusza.

Uwagi końcowe

Przedstawione powyżej uwagi nie wyczerpują oczywiście problemu konstrukcji narzędzia w badaniach e-mailingowych i Web’owych. Pomijam bowiem całkowicie nie tylko problem przygotowania i sposobu administrowania innych elementów narzędzia badawczego, takich jak np. list zapowiedni, monity, podziękowania itd., ale także kluczowe dla badań internetowych zagadnienia deontologiczne: bezpieczeństwo i poufność danych. Nie było zresztą moim celem choćby ich zasygnalizowanie⁴⁷. Było nim natomiast zwrócenie uwagi na dwa problemy: czy techniki internetowe są zupełnie nowymi technikami oraz, czy są technikami „idealnymi”? Wskazując na wspólne cechy ankiet papierowych i internetowych udzieliłem odpowiedzi na pierwsze z pytań. Odpowiedź na drugie pytanie jest bardziej skomplikowana, aczkolwiek omawiając zalety i ograniczenia tych technik, pośrednio również udzieliłem na nie odpowiedzi. Z pewnością nie są to techniki idealne. W chwili obecnej powszechne ich wykorzystanie zarówno na próbach instytucji, a tym bardziej jednostek, nie jest możliwe i to nie tylko w Polsce. Dzieje się tak ze względu na relatywnie niski poziom penetracji i związany z tym błąd pokrycia. Czy to jednak oznacza, iż są one, przynajmniej na dzisiaj, mało użyteczne? Z pewnością nie, jeśli będą wykorzystywane w projektach, w których zakładamy możliwość zbierania informacji różnymi technikami. Mam tutaj na myśli tzw. *Mixed-Mode Survey* lub *Mix-Mode Data Collecting*, w których dopuszcza się łączenie nie tylko różnych technik wywiadu (np. PAPI i CATI), ale także technik wywiadu z technikami ankiety (np. CATI i CAWI)⁴⁸. Z jednej strony, idea „mieszania” technik jest tak „stara”, jak technika ankiety wypełnianej pod kontrolą ankietera w trakcie, lub po wywiadzie. Z drugiej, jest ona zupełnie „nowa”, ponieważ nie tylko dopuszcza, ale wręcz zakłada uzyskiwanie danych – w ramach tego samego badania – różnymi technikami⁴⁹.

⁴⁷ Nie byłoby to nawet możliwe na kilkunastu stronach. Ostatnie, 3 wydanie książki Dillmana liczy 499 stron. Zob.: D.A. Dillman, J. Smyth, L.M. Christian, *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method*, New York, John Wiley Co., 2009.

⁴⁸ Zob.: E.D. De Leeuw, D.A. Dillman, J.J. Hox, *Mixed Mode Surveys: When and Why*, [w:] E.D. De Leeuw, J.J. Hox, D.A. Dillman, (eds), *International Handbook of Survey Methodology*, Lawrence Erlbaum Associates, New York, London, 2008, s. 299–316.

⁴⁹ Przykładowo, w badaniach instytucji (tzw. badania B2B), w wielu przypadkach, od początku przyjmuje się takie założenie. Jako jednostkę badania traktuje się tutaj bowiem daną instytucję, a informacje na jej temat mogą być dostarczane, jak wcześniej wspominałem, przez różne osoby. W takich badaniach efekt techniki (mode effect) nie powinien jednak być duży, ponieważ badane problemy zasadniczo dotyczą faktów, zachowań i cech, nie zaś opinii. Por. D.A. Dillman, *Mail and Internet Surveys...*; s. 342–343.

„Mieszanie” technik w standardowych badaniach (tzn. badaniach jednostek, nie instytucji) rodzi jednak wiele problemów związanych przede wszystkim z efektem techniki i tego problemu zasadniczo się nie uniknie. Jak wcześniej wspomniałem, w zależności od „środka komunikacji”, ludzie na to samo pytanie mogą udzielać różnych odpowiedzi. Czy zatem można je kwantyfikować, a następnie uogólniać na całą populację? Jednocześnie musimy mieć świadomość, iż ich „mieszanie” może kompensować słabości poszczególnych technik.

Dillman mówi o pięciu sytuacjach, w których stosuje się „mieszanie” technik⁵⁰. Są one następujące:

1. Zbieranie tych samych informacji, różnymi technikami od różnych osób. Jest to najpowszechniejszy typ badań MxM (skrót pochodzi od autora). Zazwyczaj stosuje się tę procedurę, aby zmniejszyć błąd non-responсу, przy danych kosztach. Przykładowo można stosować technikę ankiety pocztowej, a w przypadku braku zwrotu, wywiad osobisty.
2. Zbieranie danych „panelowych” różnymi technikami od tych samych osób w kolejnych falach. Przykładowo pierwsza fala panelu, ze względu na rodzaj dostępnych informacji w operacie losowania, realizowana jest techniką CAPI, kolejne zaś fale, techniką CATI lub e-mailingu.
3. Zbieranie różnych informacji różnymi technikami od tych samych osób wylosowanych w trakcie „jednorazowego” badania. Chodzi tutaj o wspomnianą sytuację, w której stosuje się technikę ankiety wypełnianej pod kontrolą ankietera w trakcie lub po wywiadzie, bądź pozostawia do wypełnienia „dzienniczek” (jest to dodatkowa technika zbierania materiałów).
4. Zbieranie tych samych informacji różnymi technikami w różnych populacjach. Chodzi tutaj przede wszystkim o projekty międzykrajowe, które w poszczególnych krajach, w całości, realizowane są różnymi technikami. Przykładem może być projekt Europejski Sondaż Społeczny, który w niektórych krajach realizowany jest techniką PAPI, w innych zaś CAPI⁵¹.
5. Wykorzystanie różnych technik „wspomagających” realizację badania prowadzonego inną techniką. Przykładowo, mogą to być monity telefoniczne w przypadku badań realizowanych techniką ankiety pocztowej⁵².

Powyższe, przedstawione przez Dillmana rozumienie *Mix-Mode Data Collecting* należy uznać za bardzo szerokie. W wąskim rozumieniu, MxM to zasadniczo pierwsza z wyróżnionych przez niego sytuacji, a więc zastosowanie różnych technik zbierania informacji od różnych jednostek stanowiących elementy danej próby⁵³.

Jak wspomniałem, „mieszanie” technik w standardowych badaniach jednostek (nie instytucji) rodzi jednak problemy związane przede wszystkim z efektem techniki: te same osoby, w zależności od kanału – sposobu komunikowania (komunikacja ustna i pisemna) oraz sposobu dotarcia – formy kontaktu (bezpośrednia – „naoczna” lub osobista oraz pośrednia, w tym: przez telefon, Internet, za pośrednictwem poczty itp.) mogą w różny sposób reagować na ten sam bodziec, a więc udzielać różnych odpowiedzi na to samo pytanie⁵⁴. Oznacza to, iż nie należy „mieszać” technik, ale... na badanie musimy spojrzeć całościowo, z punktu widzenia „kosztów”, „strat” i „korzyści”. Na ile wprowadzenie MxM zmniejszy błąd braku odpowiedzi (*nonresponse*), a na ile zwiększy błąd pomiaru (*measurement error*) przy danych kosztach. Jak

⁵⁰ Za: D.A. Dillman, *Mail and Internet Surveys...*; s. 219–222. Zob. także: E.D. De Leeuw, D.A. Dillman, J.J. Hox, *Mixed Mode Surveys...*, s. 299–316.

⁵¹ Zob.: <http://www.europeansocialsurvey.org>

⁵² Określenie tego rodzaju badań mianem MxM może budzić zastrzeżenia, ponieważ większość realizowanych badań należałoby określić w ten sposób. Przykładowo, jeśli przed badaniem realizowanym techniką PAPI wysyłamy do respondentów list zapowiadający, to należałoby to uznać za MxM. W tym przypadku lepiej jest mówić chyba o Multiple Mode lub MxM w szerokim rozumieniu.

⁵³ Za: E. D. De Leeuw, *Mixed Mode Surveys...*

⁵⁴ Wiąże się to przede wszystkim z dystansem „sytuacyjnym” i „czasowym”, charakterystycznym dla komunikacji pisemnej. Zob. F. Sztabiński, *Ankieta pocztowa...*, s. 21–47.

wspomniałem, jest to swoisty „trade off”: czy lepiej nie „mieszać” technik i uzyskać odpowiedzi od 50% wylosowanych osób (mówię tutaj o badaniach realizowanych na próbach „bez zastępowania”), czy może lepiej jest „mieszać” i uzyskać realizację na poziomie powyżej 70%?⁵⁵. Obecnie wszystko wskazuje na to, iż większym „zagrożeniem” dla wartości wnioskowania jest błąd braku odpowiedzi, niż efekt techniki (oczywiście w pewnych granicach).

Jeśli zatem decydujemy się w badaniu na *Mixed Mode Data Collection*, rodzi się pytanie, jakie techniki można lub należy łączyć, a jakich nie? Chodzi tutaj o problem „co mieszać?”. Istnieją bowiem cztery grupy technik: (i) techniki oparte na komunikacji ustnej, wspomagane komputerowo (CAPI, CATI); (ii) techniki oparte na komunikacji ustnej, niewspomagane komputerowo (PAPI); (iii) techniki oparte na komunikacji pisemnej, wspomagane komputerowo (e-mailing, web-survey) i (iv) techniki oparte na komunikacji pisemnej, niewspomagane komputerowo (ankieta pocztowa). Odpowiedź na pytanie czy można łączyć „każdą technikę z każdą” nie jest prosta. Wydaje się, iż dążyć należy do tego, aby łączyć techniki maksymalnie do siebie zbliżone, przede wszystkim ze względu na sposób komunikowania (sposób dotarcia do Respondenta wydaje się mieć mniejsze znaczenie) i unikać łączenia technik całkowicie odmiennych. Oznacza to, iż łączyć należy raczej techniki oparte na komunikacji ustnej wspomagane komputerowo z technikami niewspomagany komputerowo (np. PAPI, CAPI i CATI) oraz techniki oparte na komunikacji pisemnej wspomagane komputerowo z technikami niewspomagany komputerowo (np. ankieta pocztowa, ankieta e-mailowa, ankieta Web-owa). Wydaje się, iż należy unikać ich „krzyżowania”, aczkolwiek, to wszystko zależy ...⁵⁶.

Problem ten musimy jednak rozstrzygnąć sami i to na najwcześniejszym etapie, tzn. etapie konceptualizacji badania, uwzględniając oczywiście problematykę i badaną zbiorowość. Wybierając sposób „mieszania technik” musimy przede wszystkim odwoływać się do własnej wiedzy, doświadczenia badawczego i wreszcie, intuicji badawczej. Wydaje się jednak słuszne, aby przy wyborze technik, które mają być łączone, kierować się jeszcze jedną, dodatkową zasadą. Otóż, jeśli któraś z nich stawia przed nami większe „wymagania” w stosunku do kwestionariusza pod względem sformułowania pytań, grafiki bądź języka programowania, niż inne, to należy je uwzględnić w przypadku pozostałych technik. Innymi słowy, jeśli w badaniach internetowych stosujemy technikę e-mailingu (która z założenia jest mniej skomplikowana) oraz ankietę Web-ową (która jest bardziej „zaawansowana”), to ta pierwsza, na poziomie formalnym, nie tylko nie powinna być bardziej skomplikowana niż webowa, ale co więcej, winny być przygotowane w podobnych językach. I wreszcie, musimy pamiętać o możliwościach i tradycjach kulturowych związanych z poszczególnymi sposobami kontaktowania się.

W przypadku MxM, jeśli rozstrzygniemy pytanie „co mieszać?”, kolejnym pytaniem jest „jak mieszać?”. Istnieją dwa sposoby implementacji MxM⁵⁷. Jeden z nich to stosowanie sekwencyjne (*sequential*) kolejno różnych technik, w sposób następujący po sobie. Polega on na tym, że określamy jedną podstawową (preferowaną) technikę oraz dodatkowe (niejako „uzupełniające”), które będą stosowane w drugiej kolejności. Przykładowo, jeśli podstawową techniką jest CATI, to do grupy respondentów, z którymi nie udało się zrealizować wywiadu telefonicznie, możemy wysłać ankietę, aby zrealizowali z nimi wywiad F2F (techniką PAPI lub CAPI). I wreszcie, do tych, z którymi nie udało zrealizować wywiadu F2F, możemy wysłać ankietę pocztową. Celem MxM sekwencyjnego jest przede wszystkim redukcja błędu braku odpowiedzi (*unit nonrespon*s). Przy czym to oczywiście badacz decyduje o tym, w jakiej kolejności po-

⁵⁵ Odsetek taki, w badaniach europejskich, przyjmuje się za minimum, biorąc pod uwagę błąd braku odpowiedzi. Zob.: <http://www.europeansocialsurvey.org>

⁵⁶ Istnieją badania, w których obok wywiadu PAPI i CATI wykorzystywano również ankietę CAWI. Za: E.D. De Leeuw, *Mixed Mode Surveys*....

⁵⁷ Zob. E.D. De Leeuw, D.A. Dillman, J.J. Hox, *Mixed Mode Surveys*..., s. 306–307, 310–312.

szczególne techniki są stosowane. Drugi rodzaj MxM polega na równoczesnym (*concurrent*) stosowaniu różnych technik zbierania informacji. Nie określa się tutaj techniki podstawowej, lecz traktuje wszystkie jako ekwiwalentne i wykorzystuje się je jednocześnie, w tym samym czasie. Przykładowo możemy wysłać do wylosowanych osób list z informacją o badaniu i propozycją wypełnienia ankiety „papierowej” (ankieta pocztowa), ankiety e-mailowej lub ankiety Web-owej. W zależności od preferencji Respondenta, wysyłamy mu właściwą ankietę. Celem równoczesnego MxM jest przede wszystkim redukcja błędu pokrycia (osoby niemające dostępu do Internetu, mogą wypełnić ankietę „papierową”). Odpowiedzi na powyższe dwa pytania („co i jak?” łączyć) są ze sobą oczywiście powiązane i zależą od tego, jakiego rodzaju błąd chcemy minimalizować, przy danych kosztach.

Kolejny problem związany z wykorzystaniem MxM, to pytanie o sposób przygotowania narzędzia badawczego, w szczególności kwestionariusza. Generalnie zależy to rodzaju MxM, a więc sposobu jego zastosowania (sekwencyjnie vs. równocześnie). W pierwszym przypadku – sekwencyjny MxM, należy określić, jak wspomniano, metodę podstawową, przygotować narzędzie adekwatne dla tej techniki, a następnie stworzyć jego ekwiwalentne wersje w innych technikach. Przykładowo, tworzymy kwestionariusz CATI, a następnie „przerabiamy” – tworzymy jego odpowiednik dla techniki PAPI i CAPI. W przypadku równoczesnego stosowania MxM, brak jest, jak pamiętamy, jednej podstawowej metody. Wszystkie techniki są ekwiwalentne. Czym zatem kierować się i jaki kwestionariusz przygotować. Otóż, w tym przypadku możliwe są 3 podejścia⁵⁸:

- 1) pierwsze zakłada optymalizację każdego z narzędzi (tzw. *Mode Specific Design*). Oznacza to, że dla każdej techniki przygotowujemy „osobne” narzędzie, przy czym, co jest oczywiste, muszą być one porównywalne, a właściwie równoważne;
- 2) drugie podejście zakłada maksymalne upodobnienie każdego narzędzia na poziomie bodźca – sformułowania pytania (tzw. *Uni-Mode Design*)⁵⁹. Przy tym podejściu, aby minimalizować efekt techniki, ważne jest: zunifikowanie formatu pytań, formatu odpowiedzi oraz instrukcji. Unifikacja ta polega między innymi na: zachowaniu jednolitej struktury pytań dla poszczególnych technik (zmiana struktury może zmienić bodziec); redukowaniu liczby możliwych odpowiedzi (winny być one takie same i włączone w treść pytania, także odpowiedź „Trudno powiedzieć”); ustaleniu ekwiwalentnych instrukcji i reguł przejścia (jeśli byłyby one bardzo skomplikowane, należy rozważyć restrukturyzację kwestionariusza) itd.;
- 3) trzecie podejście zakłada ekwiwalencję pytań na poziomie kognitywnym (tzw. *Generalized Mode Design*)⁶⁰. Odchodzi się tutaj od literalnej unifikacji bodźca (pytania) na poziomie formalnym, ponieważ identyczne sformułowanie nie zawsze zapewnia ekwiwalencję. Zakłada się, natomiast, iż różne bodźce (pytania) mogą przynosić ekwiwalentne odpowiedzi na poziomie zapotrzebowania badawczego. Przykładowo, jeśli chcemy się dowiedzieć, czy skontaktowaliśmy się z właściwą osobą, to w technice CATI zadamy pytanie: „Czy rozmawiam z: . . . ? (imię i nazwisko)”; z kolei stosując ankietę Web-ową, pytanie to będzie brzmiało: „Czy nazywasz się: . . . ? (imię i nazwisko)”⁶¹.

Właściwe stosowanie *Mix-Mode Data Collecting* z pewnością redukuje błąd pokrycia (*coverage error*) oraz błąd braku odpowiedzi (*nonresponse error*). Z drugiej jednak strony, z pewnością komplikuje schemat i sposób doboru próby, co może owocować błędem próby (*sampling error*). Podobnie, komplikuje od strony logistycznej realizację badania. Jeśli chodzi o błąd pomiaru (*measurement error*), to zależy:

⁵⁸ Zob.: E.D. De Leeuw, D.A. Dillman, J.J. Hox, *Mixed Mode Surveys* . . . , s. 312–313.

⁵⁹ Zob.: D.A. Dillman, *Mail and Internet Surveys* . . . , s. 232–240.

⁶⁰ Zob.: E.D. De Leeuw, D.A. Dillman, J.J. Hox, *Mixed Mode Surveys* . . . , s. 312–313.

⁶¹ Przykład za: E.D. De Leeuw, *Mixed Mode Surveys* . . .

w niektórych przypadkach błąd ten może być większy, w innych mniejszy (z pewnością jednak zawsze występuje efekt techniki). I wreszcie koszty: znowu, to zależy. Zależy od tego, „co mieszmamy” i „z czym” porównujemy. Jeżeli z wywiadem PAPI, to zastosowanie MxM w jakiegokolwiek konfiguracji obniża nam koszty; jeśli natomiast z ankietą pocztową, czy ankietami internetowymi, to zazwyczaj je podnosi. Z pewnością jest jednak tak, iż MxM właściwie stosowany w sumie redukuje TSE (*Total Survey Error*). Stosowany jednak w sposób uproszczony może zwiększyć ten błąd.

Franciszek Sztabiński – doktor socjologii, adiunkt w Instytucie Filozofii i Socjologii Polskiej Akademii Nauk. Zajmuje się problemami oceny wartości danych w badaniach typu ilościowego, w szczególności jej związku z fazą terenowej realizacji badania. Autor lub współautor ponad 60 publikacji z zakresu metodologii badań, w tym książki „Ankieta pocztowa i wywiad kwestionariuszowy” (Warszawa 1997). Współredaktor i współautor „Fieldwork jest sztuką” (Warszawa 2005), „Nowe metody, nowe podejścia badawcze w naukach społecznych” (Warszawa 2004) oraz „Podręcznika Ankietera” (Warszawa 2000). Jeden z założycieli i członek redakcji czasopisma „ASK. Społeczeństwo, Badania, Metody”. Wykładowca w Szkole Nauk Społecznych w IFIS PAN oraz Collegium Civitas. Konsultant metodologiczny licznych projektów badawczych krajowych i międzynarodowych. Koordynator Narodowy w projekcie Europejski Sondaż Społeczny. W swojej karierze zawodowej przeszedł przez wszystkie szczeble działalności badawczej, pracując jako ankieter, koordynator regionalny, instruktor metodyczny sieci ankierskich, konsultant badań rynku, a także pełnił funkcję kierownika samodzielnych zespołów metodologii badań w CBOS oraz w Instytucie SMG/KRC Poland Media S.A. Jest członkiem European Survey Research Association (ESRA) od momentu powołania organizacji.

Bibliografia

- Blau P. M., (1964), *Exchange and Power in Social Life*, New York, John Wiley& Sons. Inc.
- Cichomski Bogdan (kierownik programu), Tomasz Jerzyński, Marcin Zieliński. *Polskie Generalne Sondaże Społeczne: struktura skumulowanych wyników badań 1992–2008*. Instytut Studiów Społecznych, Uniwersytet Warszawski, Warszawa, styczeń 2009.
- Couper M., Nicholls II W.L., (1998), *The History and Development of Computer Assisted Survey Information Collection Methods*, [w:] Couper M., Baker R.P., Bethlehem J., Clark C.Z, Martin J., Nicholls II W.L., O'Reilly J.M., *Computer Assisted Survey Information Collection*, New York, John Wiley Co.
- Dalbard-Martin F., (1996) *Wpływ struktury kwestionariusza na uzyskiwane odpowiedzi w badaniach prasowych*, ASK „Społeczeństwo, Badania, Metody”, 2(4).
- Dillman D.A., (1978), *Mail and Telephone Surveys. The Total Design Method*, New York, John Wiley& Sons. Inc.
- Dillman D.A., (2000), *Mail and Internet Surveys: The Tailored Design Method*, New York, John Wiley& Sons. Inc.
- Europejski Sondaż Społeczny, <http://www.europeansocialsurvey.org>
- Europejski Sondaż Społeczny Runda 4. Zob.: <http://ess.nsd.uib.no/ess/round4/>
- Gremy J-P. (1995), *Wpływ sposobu sformułowania pytań kwestionariuszowych na uzyskiwane odpowiedzi*, ASK „Społeczeństwo, Badania, Metody”, 2.
- Groves R. M., (1989), *Survey Errors and Survey Costs*, New York, John Wiley & Sons.

- Hippler H-J., Schwarz N., (1988), *Response Effects in Surveys*, [w:] Hippler H-J., Schwartz N. Sudman S. (eds) *Social Information Processing and Survey Methodology*, New York, Springer-Verlag.
- Homans G. C., (1961), *Social Behaviour: Its Elementary Forms*, New York, Harcourt, Brace & World.
- <http://biuroprasowe.tp.pl/pr/107315/instytut-laczności-potrzebna-strategia-inwestycji-telekomunikacyjnych?changeLocale=PL>
- <http://www.internetworldstats.com.stats.htm>
- Internet Advertising Bureau. Zob.: <http://www.iabeurope.eu/media/6238/directory%20of%20definition%20ed1ii.pdf>.
- Kinsey S.H., Jewell D.M. (1998), *A Systematic Approach to Instrument Development in CATI*, [w:] Couper M., Baker R.P., Bethlehem J., Clark C.Z., Martin J., Nicholls W. L., O'Reilly J.M., *Computer Assisted Survey Information Collection*, New York, John Wiley Co.
- de Leeuw E.D. (2009), *Mixed Mode Surveys – 6th ESS Training Course*, 30. 04–01. 05., Ljubljana, Slovenia.
- de Leeuw E.D., Dillman D.A, Hox J.J, (2008) *Mixed Mode Surveys: When and Why*, [w:] de Leeuw E.D., Hox J.J, Dillman D.A, (red.). *International Handbook of Survey Methodology*, New York, London, Lawrence Erlbaum Associates.
- deMaio T. (2009), *Attempting to Validate Cognitive Testing Methodology*, European Survey Research Association Conference, June 29-July 3, Warsaw.
- Malinowski B., (1984) *Zwyczaj i zbrodnia w społeczności dzikich*, [w:] Malinowski B., *Dzieła*, t. 2, Warszawa, PWN.
- Mangione T. W., (1995), *Mail Surveys. Improving the Quality*, Thousand Oaks, Sage.
- Mason R., Carlson J.E., Tourangeau R., (1994) *Contrast effects and subtraction in part-whole questions*, „Public Opinion Quarterly”, 58.
- Polski Internet 2008/2009, http://pliki.gemius.pl/Raporty/2009/02_2009_Polski_internet_2008_2009.pdf
- Saris W., E., (1991), *Computer-Assisted Interviewing*, Newbury Park. Sage.
- Schuman H., (2008), *Method and meaning in polls and surveys*, Cambridge, Harvard University Press.
- Schwarz N., Strack F., Mai H-P., (1991) *Assimilation and contrast effects in part-whole question sequences: a conversational logic analysis*, „Public Opinion Quarterly”, 55.
- Stoop I., (2005), *The Hunt for the Last Respondent: Nonresponse in Sample Surveys*. The Hague: Social and Cultural Planning Office of the Netherlands.
- Sztabiński F., (1997), *Ankieta pocztowa i wywiad kwestionariuszowy*, Warszawa, Wydawnictwo IFiS PAN.
- Sztabiński F., (2003), *Logika badacza i logika respondenta. Problem adekwatności narzędzia pomiaru*, ASK „Społeczeństwo, Badania, Metody”, 12.
- Sztabiński P. B., (1999), *Wywiad telefoniczny ze wspomaganie komputerowym (CATI) a wywiad kwestionariuszowy. Problem porównywalności wyników*, [w:] Domański H., Lutyńska K., Rostocki A. W., (red.) *Spojrzenie na metodę. Studia z metodologii badań socjologicznych*, Warszawa, Wydawnictwo IFiS PAN.
- Sztabiński P. B., (2001), *Wywiad osobisty ze wspomaganie komputerowym (CAPI): czy rzeczywiście idealna technika*, ASK „Społeczeństwo, Badania, Metody”, 10.
- Thibaut J.W., Kelley H. H., (1959), *The Social Psychology of Groups*, New York, John Wiley& Sons. Inc.
- Żmijewska-Jędrzejczyk T., (2004), *Badania internetowe*, [w:] Sztabiński P. B., Sztabiński F., Sawiński Z. (red.), *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Warszawa, Wydawnictwo IFiS PAN.

Metody kontroli w badaniach ewaluacyjnych

Wstęp

Zgodnie z polskim prawodawstwem publicznych o udzielenie zamówienia publicznego na realizację badań ma prawo ubiegać się każdy, kto ma wpis w działalności gospodarczej „Badania Opinii” oraz wykonał jedno badanie o podobnym charakterze. W praktyce oznacza to, że w zależności od warunków brzegowych postępowania (warunków SIWZ) do przetargu może przystąpić każdy podmiot niezależnie czy robi badania dobrej czy złej jakości.

Ponieważ na podstawie badań ewaluacyjnych dokonywana jest ocena oraz w oparciu o wyniki badań podejmowane są decyzje zarządcze na Zamawiającym, ciąży obowiązek zadbania o wysoką jakość i wiarygodność informacji płynących z procesu badawczego. Zadanie to jest o tyle utrudnione, że z jednej strony mamy do czynienia z niedużą liczbą podmiotów pracujących w oparciu o wysokie standardy realizacji i kontroli badań (legitymizujące się certyfikatami bądź własnymi standardami zapewnienia jakości procesu badawczego) z drugiej strony w znacznej liczbie przetargów jednym z kluczowych kryteriów oceny jest najniższa cena oferty.

Nikogo nie trzeba przekonywać, że profesjonalizm firmy badawczej oraz jakość produktu końcowego, czyli przede wszystkim jego skuteczność w dostarczaniu użytecznych informacji, kosztuje. Nikogo też nie trzeba zapewne przekonywać, że decyzje oparte o błędne informacje kosztują znacznie więcej. Biorąc pod uwagę powyższe kwestie, bezdyskusyjne staje się stwierdzenie, iż kroki, które musi podjąć zleceniodawca, muszą zapewnić, by podmiot realizujący projekt wykorzystując zarówno swoje procedury, jak i strukturę organizacyjną, wykona projekt zgodnie z kanonami sztuki badawczej. Dlatego też po stronie zleceniodawcy (ogłaszającego postępowanie na zamówienie publiczne) leży skonstruowanie specyfikacji istotnych warunków zamówienia/umowy tak, by wyegzekwować od Wykonawcy stworzenie właściwego otoczenia i środowiska dla zagwarantowania jakości całego procesu organizacji i realizacji badania ewaluacyjnego.

Należy jednak podkreślić, iż zdaniem autorów, rola zleceniodawcy nie jest bierna. Nie ogranicza się jedynie do zamówienia badania (konstrukcji SIWZ), odebrania wyników (raportu) i zapłacenia rachunku. Przedstawiciel zleceniodawcy (na ogół wyspecjalizowana komórka ewaluacyjna) po pierwsze jest bardzo dobrze zorientowany zarówno w celu, jak i w naturze badanego problemu, zna specyfikę respondentów, jak i możliwości, ograniczenia zastosowania w przedmiocie badania konkretnych technik badawczych. Zamawiający posiada ponadto informacje/dane zgromadzone we wcześniejszych projektach o podobnym charakterze oraz poprzez identyfikację problemu badawczego, konstrukcję zamówienia, unikalną wiedzę kontekstową chcąc nie chcąc, mniej lub bardziej uczestniczy w procesie planowania i realizacji procesu badawczego. I choć modelowa relacja firmy badawczej – zlecającego badanie, bez wątpienia, powinna mieć cechy daleko posuniętego partnerstwa, współpracy i zaufania, to

jednak wymaga określenia mechanizmów weryfikacji jakości gromadzonych danych zarówno po jednej jak i drugiej stronie. Standardy i mechanizmy kontroli powinny być czymś zupełnie oczywistym i podobnie jak kwestie związane z innymi etapami procesu ewaluacji powinny być uwzględnione zarówno w projekcie badania, jak i kosztorysie. Na pytanie „czy tak jest” czytelnik powinien odpowiedzieć sobie na podstawie własnego doświadczenia lub też na podstawie lektury SIWZ poszczególnych ewaluacji.

Budowanie jakości w środowisku projektu

W celu uzyskania wyników wysokiej jakości potrzebne jest działanie wielopoziomowe, zgodne z wypracowanymi standardami, obejmujące cały proces badawczy. Za trzon tego procesu należy uznać trzy główne etapy: 1) konceptualizację i operacjonalizację badania (w tym jakość narzędzi badawczych), 2) terenowe prace gromadzenia danych (pracę ankietatorów), 3) przetwarzanie danych i analizy. Przy czym o ile samo dobre przygotowanie narzędzia badawczego może zminimalizować błędy terenowego gromadzenia danych to nawet najlepsze analizy merytoryczne nie naprawią błędów popełnionych podczas realizacji badania.

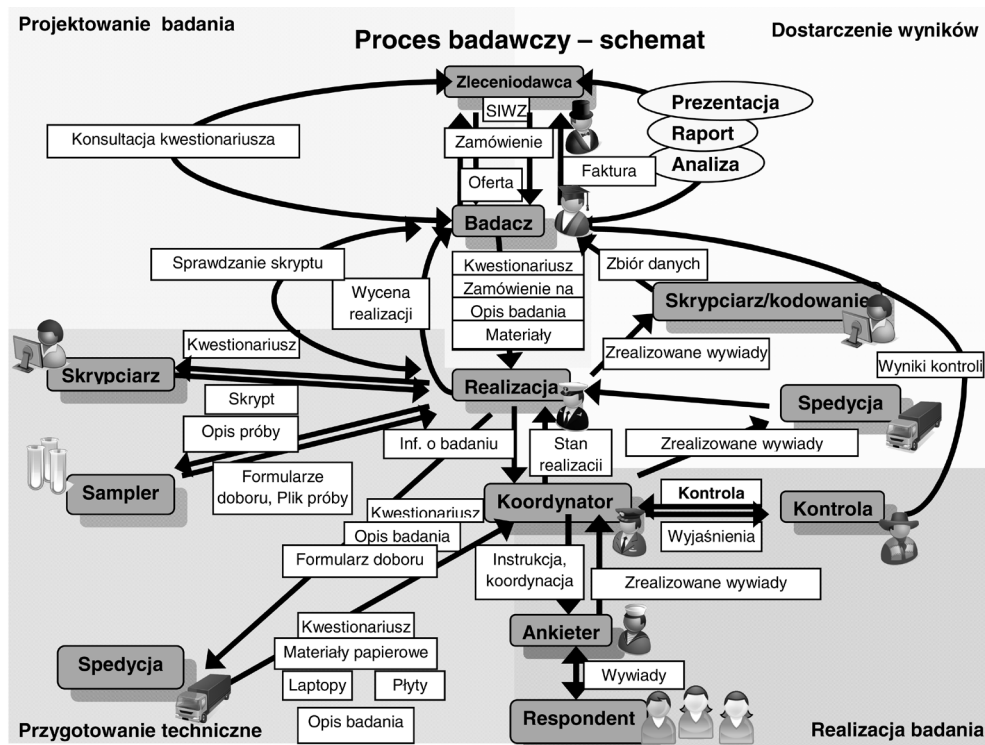
Wychodząc z założenia, że **podstawą rzetelnych wyników są wysokiej jakości prace terenowe** w niniejszym artykule zajmiemy się obszarem kontroli jakości terenowych prac gromadzenia danych w podziale na kwestię: 1) przygotowania badania do realizacji terenowej (czyli m.in. doboru i przygotowaniem ankietatorów i koordynatorów), 2) realizacji badania zgodną z założeniami teoretycznymi oraz 3) weryfikacji zebranego materiału i eliminacją materiału niespełniającego kryteriów.

Przygotowanie badania do realizacji terenowej

Zanim przejdziemy do kluczowego zagadnienia niniejszego artykułu – szczegółowych zagadnień kontroli warto zwrócić uwagę na kilka zagadnień. Pierwszym z nich jest **struktura organizacyjna zleceńiobiorcy**. Z punktu widzenia realizacji konkretnego projektu badawczego warto przyrzeć się takim problemom jak:

- a) to czy wykonawca ma podwykonawców procesu, jeżeli tak, to jakie czynności należą do podwykonawcy oraz jak to podwykonawstwo jest organizowane i czy oraz jak prace podwykonawcy są kontrolowane?
- b) jakie osoby są odpowiedzialne za bezpośrednią obsługę zleceniodawcy (zwykle badacze)?
- c) kto odpowiada za dobór lub losowanie próby?
- d) kto odpowiada za koordynację i prowadzenie badania w terenie, szkolenia ankietatorów i monitorowanie badania?
- e) kto odpowiada za przetwarzanie danych, kontrolę, a także logistykę wszystkich tych procesów?
- f) jak zorganizowany jest nadzór w terenie i jak załatwia się problemy rodzące się w trakcie realizacji prac terenowych?

Już sama analiza liczby czynności składających się na realizację standardowego projektu badawczego pokazuje, ile uczestników po stronie wykonawcy musi być w niego zaangażowanych oraz jak ich przygotowanie do wykonywania swojej roli (kompetencje i zrozumienie procesów) nie może być przypadkowe.



Rys. 1. Przykładowy schemat procesu badawczego z wyszczególnieniem jego uczestników

Kolejnym zagadnieniem, jakie należy rozpatrzyć przed rozpoczęciem realizacji badania, jest **plan realizacji procesu**. Przygotowanie i realizacja badania wymaga spełnienia odpowiednich standardów zgodnych z dokumentami polityki jakości, które opisują dokładną chronologię i aktorów procesów organizacji projektu (zob. poniższą tabelę). Warto, niejako powtórnie, zadać sobie pytanie:

- Czy istnieją procedury/dokumentacja „cyklu produkcyjnego”? (przykład tabela 1)
- Czy przykładą się należyta wagę do fazy przeprowadzania wywiadów w porównaniu z innymi fazami projektu?
- Czy proponowane metody doboru próby są realne do wykonania w terenie?
- Czy i jakie procedury kontrolne są przewidziane w celu kontroli realizacji zaplanowanych procedur doboru próby?
- Jakie mogą pojawić się obszary ryzyka.

Trzecią ważną kwestią są procedury badania, narzędzia badawcze i ich pilotaż oraz materiały pomocnicze wykorzystywane w badaniu. Podstawą tych rozważań powinna być ocena wykonalności zadania w kontekście budżetu, czasu i metodologii. Nawet najlepiej dobrana metoda pomiaru może zawieść, jeżeli jest niewykonalna w zbyt krótkim czasie bądź jeśli wycena projektu nie uwzględniała faktycznej pracochłonności i wysiłku potrzebnego do wykonania zadania.

Po analizie powyższych aspektów można przyglądać się zaplanowanym ostatecznym krokom badawczym, jak i samej treści oraz formie materiałów prezentowanych respondentom (kwestionariusz, karty, nagrania itp.).

Tabela 1. Szczegółowy schemat procesu realizacji w siedzibie instytutu badania.

1. Informacja o badaniu od Zespołu Badawczego
2. Zaproponowanie terminarza badania:
 - 2.1. Ustalenie dat dostarczenia kwestionariusza oraz dostarczenia materiałów kolorowych do Działu Realizacji
 - 2.2. Ustalenie dat wysyłki w teren
 - 2.3. Ustalenie dat realizacji i spływu (ew. międzyspływów)
 - 2.4. Ustalenie dat przekazania danych do Działu Przetwarzania Danych
3. Zamówienie na próbę od Zespołu Badawczego.
4. Przekazanie zamówienia na badanie do Specjalisty Doboru Prób.
5. Skonsultowanie z Zespołem Badawczym niejasności w zamówieniu i doprecyzowanie zamówienia.
6. Sprawdzenie kwestionariusza pod względem logiczności i spójności.
7. Sprawdzanie kwestionariusza, w tym zwłaszcza:
 - sprawdzenie poprawności długości numeru respondenta,
 - sprawdzenie filtrów (np. sprawdzenie granic wieku),
 - sprawdzenie kompletności kafeterii,
 - sprawdzenie czy wszędzie pojawiają się komentarze dla ankietera, (np. „Pokazać respondentowi kartę D2”),
 - sprawdzenie czy wyświetlają się kody wyjściowe typu: „odmowa odpowiedzi”, „nie wiem/ trudno powiedzieć.”
8. Zatwierdzić do realizacji (po uprzednim zatwierdzeniu preliminarza przez badacza).
9. Przygotowanie wyceny.
10. Akceptacja wzoru i wydruk formularzy doboru.
11. Przygotowanie kart do pytań.
12. Odebranie od badacza opisu badania dla ankietera.
13. Sporządzenie spisu materiałów.
14. Powiadomienie Działu Przetwarzania Danych o terminach realizacji i spływów badania.
15. Informacja do Działu Spedycji o wysyłce.
16. Poinformowanie mailem o realizowanym badaniu, jego terminach oraz terminie ostatecznego spływu do Działu Przetwarzania Danych.
17. Wysyłka.
18. Monitorowanie postępów w badaniu:
 - 18.1. Kontakt z Koordynatorami Regionalnymi,
 - 18.2. Kontakt z Zespołem Badawczym,
 - 18.3. Kontakt z Działem Przetwarzania Danych.
19. Przekazywanie zainteresowanym stronom informacji o ewentualnych problemach w realizacji, a następnie podjętych środkach zaradczych i decyzjach.
20. Spływ i oddanie danych.
 - 20.1. Przeliczenie i sprawdzenie kompletności kwestionariuszy z poszczególnych ośrodków.
 - 20.2. Rozdzielenie formularzy doboru od kwestionariuszy.
 - 20.3. Ponumerowanie kwestionariuszy.
 - 20.4. Oddanie kwestionariuszy do Działu Przetwarzania Danych.
 - 20.5. Przekazanie formularzy do Działu Kontroli.
21. Informacja do Zespołu Badawczego o zakończeniu realizacji terenowej i przekazaniu danych do Działu Przetwarzania Danych.

z „Poradnika Badacza”:

„Lepiej przygotuj narzędzie do końca i dopiero wtedy je wypuść, niż uzupełniaj w czasie realizacji nawet o drobne szczegóły. Każdy taki szczegół potrafi zwiększyć koszty badania”.

„Staraj się przygotować narzędzie jak najstaranniej. Ankieter nie da w badaniu z siebie więcej niż Ty przy projektowaniu narzędzia. Jeżeli się nie postarasz, on czuje, że też za bardzo nie musi”.

„Myśląc o ankiecie i jego respondencie myśl jak o sobie”.

„Ani ankieter, ani respondent nie zrobią bezinteresownie czegoś, czego Ty byś nie zrobił”.

„Ani ankieter, ani respondent nie zrobią czegoś, na co Ty sam byś się nie zgodził”.

„Pamiętaj, że swoją wyceną i procedurą badawczą możesz generować konkretne zachowania uczestników procesu badania, które będą wpływały na wyniki badania...”

„...my nazwiemy to pójściem na łatwiznę, inni usprawnieniem pracy”.

„Staraj się, aby procedura badania nie pozwalała na niepożądane zachowania.

Kontrola badania niczego nie uratuje, dowiesz się co najwyżej gorzkiej prawdy”.

Czwartym zagadnieniem, na które warto zwrócić uwagę to dobór ankierów do badania oraz przygotowanie sieci (koordynatorów i ankierów) do procesu gromadzenia danych. Merytoryczne przygotowanie „terenu” powinno objąć takie kwestie, jak:

- a) Jak ankierzy byli rekrutowani do projektu, ze względu na jakie cechy (np. doświadczenie, wiek, płeć, staż pracy w firmie itp.), o ile jest to istotne dla realizacji badania?
- b) Jak jest zbudowana sieć ankierska?
- c) Na jakich warunkach/w jakim wymiarze czasu są zatrudnieni (istotne np. przy sesjach CATI, pojedyncza sesja pracy ankiera nie powinna trwać dłużej niż 4–5 godzin)?

Odrębną kwestią są szkolenia koordynatorów i ankierów zarówno te organizowane dla zapoznania ankierów z tematem i narzędziem badawczym, jak i te pozaprojektowe w zakresie technik polepszania warsztatu pracy¹. Ponieważ zleceniodawca musi mieć prawo zagwarantowania swojego udziału w szkoleniu/szkoleniach, zleceniobiorca winien poinformować, gdzie i kiedy będą się odbywać wszystkie szkolenia. Jeżeli z powodów logistycznych zamawiający nie może być obecny na wszystkich szkoleniach, wykonawca badania winien przekazać zleceniodawcy nagranie audio lub wideo z przeprowadzonych szkoleń, materiały szkoleniowe oraz listę osób, które wzięły w nim udział. Szkoleniem muszą być objęte wszystkie osoby odpowiedzialne za pozyskiwanie danych. Nie powinno się dopuszczać możliwości, aby badanie realizowała jakkolwiek nieprzygotowana osoba. Aktywny udział Zamawiającego badanie w szkoleniach jest dobrą praktyką. Zamawiający może występować w nich jako bierny obserwator lub towarzyszyć badaczowi (twórcy narzędzia) jako osoba prezentująca cel, kontekst badania, aparat pojęciowy towarzyszący przedmiotowi badania itp.

¹ Istotne z punktu widzenia potrzeby posiadania przekonania, że badanie jest realizowane przez grupę przygotowanych zawodowo, nieprzypadkowych osób. Aranżacja kontaktu z respondentem, sposób prowadzenia wywiadu, techniki zapisywania odpowiedzi, czy w końcu prawne wymagania o poufności zbierania informacji, czy też o ustawie danych osobowych należą do tych rzeczy, które profesjonalny ankier powinien znać zanim przystąpi do realizacji jakiegokolwiek badania.

Tabela 2. Zakres standardowego szkolenia dla ankieterów

Standardowe szkolenie do badania powinno przebiegać według następującego scenariusza:

1. Sprawdzenie listy obecności.
2. Omówienie celu badania.
3. Rozdanie materiałów badawczych uczestnikom szkolenia.
4. Przedstawienie ogólnych informacji o badaniu.
 - 4.1. Typy próby
 - 4.2. Wielkość próby (ile w całej Polsce, ile poszczególne ośrodki, ile dany ankieter).
 - 4.3. Rodzaj próby.
 - 4.4. Terminy realizacji badania.
5. Szczegółowe omówienie poszczególnych etapów badania.
 - 5.1. Aranżacja wstępna.
 - 5.1.1. Przedstawienie się i przekazanie ogólnych informacji o wykonawcy projektu.
 - 5.1.2. Ogólne przedstawienie celu i procedury badania.
 - 5.1.3. Wyjaśnienie zasad doboru.
 - 5.1.4. Wyjaśnienie roli respondenta w badaniu.
 - 5.1.5. Poinformowanie respondenta o zgodności naszych działań z Ustawą o ochronie danych osobowych.
 - 5.1.6. Poinformowanie o poufności zbieranych w badaniu informacji.
 - 5.1.7. Wręczenie respondentowi ulotki o wykonawcy i o celu badania.
 - 5.2. Rekrutacja.
 - 5.2.1. Rodzaj próby.
 - 5.2.2. Przedstawienie kryteriów doboru respondentów.
 - 5.2.3. Omówienie poszczególnych pytań w kwestionariuszu rekrutacyjnym.
 - 5.2.4. Przeanalizowanie formularzy doboru respondenta.
 - 5.3. Analiza poszczególnych pytań kwestionariusza zasadniczego ze zwróceniem uwagi na wykorzystywane materiały.
 - 5.3.1. Ogólne przedstawienie budowy kwestionariusza (wersje, bloki pytań itp.)
 - 5.3.2. Sposób zadawania pytań i notowania odpowiedzi.
 - 5.3.2.1. Szczegółowe omówienie pytań trudnych i niestandardowych.
 - 5.3.2.2. Poinstruowanie ankieterów o dopytywaniu w pytaniach otwartych.
 - 5.3.2.3. Zwrócenie uwagi na kierunki skal.
 - 5.3.2.4. Zwrócenie uwagi na konieczność instruowania respondentów o korzystaniu z odpowiedzi zawartych w kartach do pytań.
 - 5.3.2.5. Podkreślenie konieczności dopytywania się w przypadku odpowiedzi typu: „proszę zaznaczyć czwartą odpowiedź”, „tak jak poprzednio”.
 - 5.3.3. Przeprowadzenie wywiadu treningowego przez dwóch ankieterów, w którym jeden wciela się w rolę respondenta (pozostali uczestnicy szkolenia uważnie przysłuchują się wywiadowi, starając się wychwycić błędy ankietera) – w trakcie treningu prowadzona jest dyskusja.
 - 5.4. Aranżacja końcowa.
 - 5.4.1. Pozyskanie danych osobowych respondenta.
 - 5.4.2. Poinformowanie respondenta o możliwości ewentualnej kontroli.
 - 5.4.3. Podziękowanie respondentowi, ewentualnie wręczenie listu, upominku itp.
6. Omówienie kwestii organizacyjnych i technicznych.

- 6.1. Szkolenie do obsługi sprzętu komputerowego wykorzystywanego podczas badania (w przypadku badania CAPI).
- 6.2. Sprawdzenie kompletności materiałów niezbędnych do realizacji badania.
- 6.3. Przedstawienie stawek za realizację.
- 6.4. Wyznaczenie terminów raportowania o stanie realizacji.
7. Pytania od ankierów.
8. Krótkie podsumowanie (przypomnienie najważniejszych rzeczy).

Podobnie jak badacz, Zamawiający powinien upewnić się czy narzędzie badawcze, pojęcia w nim zawarte oraz instrukcje dla ankierów są zrozumiałe, a także czy i jak informuje się ankierów o istocie procedury doboru respondentów oraz o możliwościach odejścia od przyjętych procedur na wypadek zaistnienia różnych trudności. Należy precyzyjnie określić sposób dotarcia do respondenta – czy ma się odbywać w sposób losowy, udziałowy czy też celowy. Jeżeli dobór jest losowy należy ustalić *response rate*, zdefiniować wszystkie możliwe powody niezrealizowania wywiadu (*non response*), wielkość wiązki, czyli ile dziennie wywiadów ma realizować ankier, liczbę powrotów do wylosowanych respondentów wypadku niezastania lub innych niedefinitywnych powodów niezrealizowania wywiadu. Ustalenie dokumentacji doboru, w tym instrukcji doboru respondenta², leży w równej mierze do badacza, jak i zamawiającego. Dla komórki realizującej badanie/ankierów pomocny w tych działaniach jest **formularz doboru respondenta**. Jego zakres powinien obejmować przedmiot badania, rodzaj próby, procedurę dotarcia do respondenta oraz, co równie istotne, informację o rezultatach kontaktu z respondentem. (zob. poniższe formularze doboru udziałowego i imiennego respondentów do badania).

MillwardBrown
SMG/KSC
Centrala: (0-22) 54-52-000

Nazwa ankiety: **Miś 8** Symbol badania: **09N12886** Miasto: **Warszawa** Okręg: **31 / WA3**

Numer respondenta: **31-5010-043** Województwo: **mazowieckie** Klasa wielkości miejscowości: **miasto od 200 tys. mieszk.**

Transza: **001** Wiązka: **342**

Branża: **Produkcja** Wielkość firmy (zatrudnionych na stałe): **101-250 pracowników**

Respondent spełnia następujące kryterium:
* jest osobą odpowiedzialną w firmie za sprawy związane z telekomunikacją

Klucz: **G335** Numer respondenta z 2 transzy (wisać jeśli dotyczy):

Pieczętka firmy:

Dane firmy (wisać po dobraniu):
nazwa firmy: _____
miejscowość: _____
ulica: _____
nr domu: _____ nr lokalu: _____
imię i nazwisko respondenta: _____
ID z bazy: _____
telefon stac.: **0** | | | - | | | - | | |
lub telefon kom.: **0** | | | - | | | - | | |

Typ firmy (zaznaczyć X):
☐ podstawowa ☐ rezerwowa (udziałowa)

Oświadczenie:
Pozwieniam własnoręcznym podpisem, że w obecności koordynatora zapoznałem (-am) się z instrukcjami i materiałami do badania i następnie respondenta dobrałem (-am) i wywiad zrealizowałem (-am) osobiście, zgodnie z instrukcjami.

Imię: _____ Nazwisko: _____ Nr Legitymacji: _____ Podpis: _____

Rys. 2. Przykładowy formularz doboru udziałowego B2B

² Należy ustalić, jaka powinna być wielkość i rodzaj próby mając na względzie jak liczna jest populacja, która ma reprezentować badanie. Po stronie wykonawcy leży sformułowanie precyzyjnej definicji, kto jest respondentem, w tym także ustalenie czy w wypadku badanego problemu nie następuje konieczność dopuszczenia respondenta zbiorowego (szczególnie w badaniach Bussines to bussines).

		Nazwa badania CAPIBUS	Symbol badania 09Omni-010-011	Województwo BIALOSTOCKIE	Okręg 02 / BIA
Centrala: (0-22) 64-62-000 Dane wylosowanej osoby, dobór rezerwy wg kroków (+7, +1, +1) Jan Kowalski ul. Hamasiów 6 Białystok Telefon (wpisać)		Numer respondenta 02-5091-001	Dane osoby (wpisać po dobru)		
Cechy dobranej osoby mężczyzna, 35-44 lat(a)		imię i nazwisko: miejscowość: ulica: nr domu: nr lokalu: telefon stac.: [] [] [] - [] [] [] - [] [] [] lub telefon kom.: [] [] [] [] - [] [] [] []			
Wielkość miejscowości miasto 200 - 499 tys.		Nr Wiązki 021	Uwaga Adres respondenta jest adresem startowym w badaniu Pikus		

Charakterystyka budynku i użytkowanie mieszkania			Respondent zasadniczy	Respondent rezerwow
Liczba pięter budynek parterowy - 0, budynek wielopiętrowy (wpisać liczbę pięter, np. dwupiętrowy - 2), (jeśli budynek składa się z standardu: parter, 1. piętro, 2. piętro... zaznaczyć X z prawej strony)				
Piętro respondenta na parterze - 0, na piętrze (wpisać liczbę piętra, np. na drugim piętrze - 2)				
Terminy i przebieg wizyt u wylosowanej osoby			Badanie zrealizowano z respondentem	
Data wizyty	Godz. wizyty	Rezultat	1. Wylosowanym 2. Rezerwowym Czy do ankiety dołączono kartę zapowiednią?	
		Symbole rezultatu wizyty 1. Wywiad zrealizowano ze wskazanym respondentem 2. Wskazano niewłaściwy adres, podany adres nie istnieje 3. Człowiek wylosowany nie mieszka pod wskazanym adresem 4. Pod wskazanym adresem nie można nikogo znaleźć; respondent nie mieszka 5. Respondent odmówił udziału w badaniu 6. Inne przyczyny niezrealizowania wywiadu	1. Tak 2. Nie -> Dlaczego?	
Oświadczenie				
Potwierdzam własnoręcznym podpisem, że w obecności koordynatora zapoznałem (-am) się z instrukcjami i materiałami do badania i następnie respondenta dobrałem (-am) i wywiad zrealizowałem (-am) osobiście, zgodnie z instrukcjami.				
Imię		Nazwisko	Nr Legitymacji	Podpis

Rys. 3. Przykładowy formularz doboru imiennego

Piątym obszarem, istotnym dla jakości oraz procesu wykonania badania, są kwestie zaplecza technicznego – niezbędnego do zrealizowania zlecenia zgodnie z jego specyfikacją. Jest oczywiste, że aby zrealizować badanie CAPI (*Computer Assisted Personal Interview*) trzeba dysponować nie tylko odpowiednią liczbą mobilnych ankierów, ale także odpowiednią liczbą laptopów. Od zaplecza technicznego, jakim dysponujemy, zależy również:

- szybkość spływu i obróbki gromadzonych danych; techniki CAPI i CATI m.in.: eliminują konieczność ponownego wprowadzania danych do komputera; umożliwiają szybkie przesyłanie danych przez Internet;
- ułatwienie pracy ankierom i eliminacja niektórych możliwych błędów popełnianych przez ankierów, np. oprogramowanie do badań CATI i CAPI umożliwia eliminację błędów przejścia w pytaniach złożonych (ankier odczytuje jedynie pytanie pojawiające się na ekranie komputera);
- monitorowanie / kontrola pracy ankierów (kontrola realizacji poprzez → GPS w telefonach komórkowych; rejestrację wywiadów za pomocą dyktafonów, laptopów, rejestrację obiektów / zdarzeń → aparaty fotograficzne, kamery wideo itp.).

Powszechnie wiadomo, że badania są tylko tak dobre, jak dobre są prace terenowe i najbardziej wyrafinowane metody statystyczne analizy danych nie pomogą, gdy zebrane w terenie dane są kiepskiej jakości. Choć w procesie zamówienia publicznego zamawiający nie ma możliwości wymagania konkretnego certyfikatu to jednak fakt posiadania przez firmę certyfikatu jakości realizacji badań³ wydawanego przez niezależne organizacje jest niezwykle istotny. Jeżeli firma posiada taki certyfikat, standardy współpracy na linii zamawiający – wykonawca stają się o wiele prostsze oraz bardziej przejrzyste. Co jednak, jeżeli firma realizująca badanie nie posiada takiego certyfikatu? Jak klient może kontrolować badanie? Na to pytanie postaramy się odpowiedzieć w kolejnych częściach artykułu.

³ Np.: PKJPA - Program Kontroli Jakości Pracy Ankierów administrowany przez Organizację Badania Opinii i Rynku, norma ISO 20252, CASRO Code of Standards and Ethics for Survey Research, The Interviewer Quality Control Scheme (IQCS).

Realizacja badania zgodna z założeniami teoretycznymi czyli jak klient może kontrolować badanie?

By minimalizować ewentualne obszary błędów, jakie mogą pojawić się w procesie gromadzenia danych, musimy rozważyć następujące kwestie: (1) **cel i funkcja kontroli**, (2) założenia **przedmiotu kontroli**, (3) założenia **zakresu kontroli**, (4) założenia **metody kontroli**.

(1) Cel i funkcja kontroli. Nikogo nie trzeba przekonywać, że kontrola jakiegokolwiek procesu powinna być standardem zarządzania. W przypadku realizacji badań celem kontroli jest podnoszenie jakości, skuteczności oraz wiarygodności wyników badań. Kryteria kontroli są zróżnicowane i zależne od wielu czynników, w tym od funkcji, jakie kontrola ma pełnić. Zasadniczo możemy wyróżnić dwie główne funkcje: **kontrolną** – funkcja rozliczeniowa i egzekwowania odpowiedzialności, czyli innymi słowy selekcja i sprawozdawczość oraz **instrumentalną** – funkcja stymulowania usprawnień procesowych oraz wzmacniania poczucia współodpowiedzialności za jakość produktu końcowego. Jej zadaniem jest zagwarantowanie wydajnej, racjonalnej oraz skutecznej alokacji zasobów poprzez weryfikację logiki, struktury, procedur oraz procesów realizacji badania.

Poszczególne funkcje kontroli w praktyce nie są rozłączne, jednak to, na jakie akcenty kontroli położymy nacisk, zależy w dużej mierze od skali badania, czasu realizacji, zastosowanego podejścia badania oraz konkretnych technik badawczych, ale również od zasobów finansowych przeznaczonych na takie działania, wymagań formalnych zamawiającego, posiadanych certyfikatów oraz procedur po stronie firmy badawczej.

Zanim przejdziemy do omówienia poszczególnych obszarów kontroli warto naszkicować wynikające z nich perspektywy. I tak pierwsza z wymienionych funkcji służy głównie zamawiającemu do oceny końcowych efektów – jakości, wiarygodności i tym samym użyteczności produktu końcowego badania ewaluacyjnego, druga wzmacnia przede wszystkim wykonawcę badania w działaniach operacyjnych nakierowanych na usprawnienia procesów, rozwiązywanie aktualnych problemów, wątpliwości oraz bieżącej szybkiej diagnozy i ewentualnych sugestii korekt.

(2) Założenia badania kontrolnego. W relacji zleceniodawca–zleceniobiorca w kwestii diagnozowania jakości danych od początku powinny obowiązywać jasne zasady. Konieczne jest precyzyjne określenie, co, dlaczego, kiedy oraz w jaki sposób ma być kontrolowane. Kontrola nie powinna nikogo zaskakiwać, a więc jej szeroki zakres oraz czas powinny być oczywiste. Ponieważ jakość danych powinna być przedmiotem wspólnej troski wykonawcy i zamawiającego, to głęboka i transparentna współpraca w zakresie kontroli staje się koniecznością. Biorąc pod uwagę powyższe uwarunkowania, warto, by kontrola badania była realizowana według ściśle określonych zasad i procedur.

Wyznacznikiem dla nich jest **przedmiot kontroli**. Warunkowany jest on m.in.: (1) charakterem badania zasadniczego – np. badanie jednorazowe, badanie panelowe; (2) metodologią oraz techniką/mi zastosowanymi w badaniu; (3) rodzajami próby (np. imienna, adresowa, udziałowa) oraz wielkością próby; a także (4) poziomem kontroli – oceną rzetelności pomiaru lub oceną pracy ankietatorów.

Ocena **rzetelności pracy ankietatorów ma obecnie najbardziej rozbudowane formy**. Może być przeprowadzona na podstawie – **kontroli pośredniej** (analizy pracy ankietatorów) oraz na podstawie **kontroli bezpośredniej** (powtórzonego zadania respondentom części pytań z kwestionariusza zasadniczego). Najczęściej weryfikowane aspekty realizacji wywiadu obejmują (1) sam **fakt realizacji** wywiadu (z właściwym respondentem przez konkretnego ankietera) oraz miejsce realizacji wywiadu; (2) **popraw-**

ność realizacji procedury: poprawność dotarcia do właściwego respondenta; czas trwania wywiadu; fakt zastosowania i poprawność wykorzystania materiałów pomocniczych (jeżeli takie były przewidziane) itp. (3) **zachowanie standardów realizacji** przewidzianej **dla danej techniki**, w tym ocena pracy ankietera dokonywana przez respondenta (np. aranżacji wywiadu, sposobu zadawania pytań, sposobu zachowania itp.); (4) **poprawność/rzetelność wykonania wywiadu ankietowego** – kontrola: (a) zapisu zgodności odpowiedzi respondenta; (b) zakresu (czy wszystkie) i (c) poprawności zadanych pytań (czy poprawnie) oraz (d) ewentualnego wystąpienia wpływu ankietera na odpowiedzi respondenta.

Zakres kontroli, a więc to ile i które wywiady chcemy skontrolować, zależy przede wszystkim od wielkości **próby zasadniczej** oraz **od procedury wyboru przypadków** do kontroli. Nie ma z góry określonej wielkości procentowej zrealizowanych wywiadów, jaką należy skontrolować w celu weryfikacji poprawności. Standardowo przewiduje się ok. 5% (PKJP) całości próby przewidzianej do realizacji. Niemniej jednak ostateczna wielkość zależy z jednej strony od indywidualnych standardów przyjętych w wyniku zapisów umowy, procedur wynikających z danej certyfikacji lub praktyki danego wykonawcy, z drugiej strony od przewidzianych środków finansowych na ten obszar realizacji badania. Ponadto zasady doboru przypadków obejmują takie kwestie jak: czy chcemy kontrolować jedynie realizację wywiadów kwestionariuszowych, czy również pracę ankieterów. W przypadku pierwszego rozstrzygnięcia podejmujemy decyzję co do wielkości próby oraz sposobu wyboru przypadków – zrealizowanych ankiet (losowo, celowo). W przypadku kontroli pracy ankieterów należy rozważyć czy kontrolujemy wszystkich ankieterów, czy tylko część, jak będą wybierani (losowo, celowo); czy w przypadku celowego wyboru kontroli pracy ankieterów będziemy badać całość ich pracy, czy jedynie losowo wybrane przypadki oraz w przypadku kontroli pracy wszystkich ankieterów jaki % lub liczbę realizowanych ankiet obejmujemy kontrolą itd.

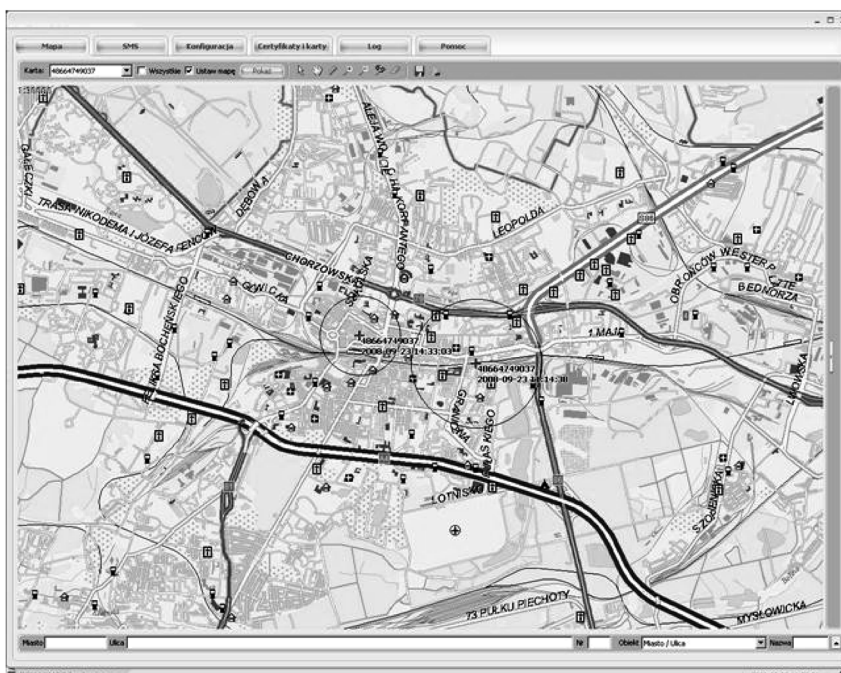
Ankieter	Okręg	Suma kontroli	Liczba bł. podst.	Procent skontrolowanych	Liczba błędów istotnych	Procent skontrolowanych	Liczba przewinień	Procent skontrolowanych	Podgląd
RAZEM		16486	209	1,27%	1508	9,15%	1140	6,91%	Pokaż
XXX YYY	Poznań	17	1	5,88%	1	5,88%	0	0,00%	Pokaż
XXX YYY	Warszawa	23	0	0,00%	2	8,70%	1	4,35%	Pokaż
XXX YYY	Kraków	20	0	0,00%	3	15,00%	0	0,00%	Pokaż
XXX YYY	Wrocław	2	0	0,00%	0	0,00%	0	0,00%	Pokaż
XXX YYY	Łódź	17	0	0,00%	2	11,76%	0	0,00%	Pokaż
XXX YYY	Kalisz	79	2	2,53%	5	6,33%	7	8,86%	Pokaż
XXX YYY	Gdańsk	4	0	0,00%	1	25,00%	1	25,00%	Pokaż
XXX YYY	Białystok	38	0	0,00%	3	7,89%	0	0,00%	Pokaż
XXX YYY	Lublin	38	0	0,00%	0	0,00%	1	2,63%	Pokaż
XXX YYY	Dębica	1	0	0,00%	0	0,00%	0	0,00%	Pokaż
XXX YYY	Katowice	16	0	0,00%	1	6,25%	4	25,00%	Pokaż
XXX YYY	Koszalin	13	0	0,00%	0	0,00%	0	0,00%	Pokaż
XXX YYY	Elbląg	54	3	5,56%	12	22,22%	5	9,26%	Pokaż
XXX YYY	Kielce	1	0	0,00%	0	0,00%	0	0,00%	Pokaż
XXX YYY	Gdynia	1	0	0,00%	0	0,00%	0	0,00%	Pokaż
XXX YYY	Szczecin	1	0	0,00%	0	0,00%	0	0,00%	Pokaż
XXX YYY	Siedlce	1	0	0,00%	0	0,00%	0	0,00%	Pokaż
XXX YYY	Opole	15	0	0,00%	0	0,00%	0	0,00%	Pokaż
XXX YYY	Bolków	15	0	0,00%	1	6,67%	1	6,67%	Pokaż
XXX YYY	Tarnów	22	0	0,00%	1	4,55%	2	9,09%	Pokaż
XXX YYY	Częstochowa	65	1	1,54%	9	13,85%	4	6,15%	Pokaż
XXX YYY	Gorzów Wlk	2	0	0,00%	0	0,00%	0	0,00%	Pokaż

Rys. 4. Tabela prezentująca wyniki kontroli pracy ankieterów

Od przedmiotu i zakresu kontroli zależy **metoda kontroli**. Wskazać można dwie podstawowe grupy technik kontroli: terenowe i nieterenowe. **Kontrola terenowa** kładzie nacisk na powtórny kontakt bezpośrednio z respondentem (kontrolerzy docierając do respondenta bezpośrednio *face to face*, za pośrednictwem poczty lub telefonicznie, pytają respondentów o fakt odbycia wywiadu, cechy respondenta, sprawdzając czy właściwie dostał się on do próby, weryfikując procedurę przeprowadzenia badania).

Kontrola nieterenowa skoncentrowana jest na spójności danych (np. ojciec starszy od syna) lub na koherencji danych (np. analogiczne zachowania takich samych respondentów u różnych ankierów).

Do metod KONTROLI TERENOWEJ zaliczyć można działania w tzw. trakcie trwania wywiadu oraz po jego zakończeniu. Do pierwszej grupy zaliczyć możemy działania takie jak: (1) **nagrywanie ukryte** w sytuacji, gdy realizacja wywiadu odbywa się w „warunkach kontrolowanych”, np. pomieszczenia specjalnie do tego przystosowanego obiektu posiadającego system rejestracji audio wideo oraz (2) **nagrywanie jawne** do którego ankier zostaje zobowiązany (np. przez dyktafon; kamerę, mikrofon w laptopie; kamerę wideo), kontrola ta polega na analizie zarejestrowanego materiału przekazywanego wraz z zrealizowanym kwestionariuszem; (3) **podstawienie respondenta** (przy czym należy pamiętać, by przed analizami wyeliminować dany wywiad ze zbioru danych); (4) **obserwacja ukryta** (np. **podśluch w trakcie realizacji** - w badaniach realizowanych techniką CATI gdzie supervisor bez wiedzy ankiera w każdej chwili ma możliwość podsłuchu rozmowy ankiera i respondenta); (5) **obserwacja jawna** w trakcie której kontroler uczestniczy w wywiadzie jako obserwator oraz (6) **obserwacja miejsca i czasu realizacji wywiadu za pomocą GPS** np. w telefonie komórkowym, który ankier otrzymuje wraz ze zleceniem badania, a także (7) tzw. rytmiczność spływów w przypadku badań CAPI oraz realizacja w ramach sesji przy badaniach CATI.



Rys. 5. Miejsce realizacji wywiadów ankierskich wraz z czasami realizacji naniesione na mapę – program do namierzania ankierów

Do drugiej grupy metod kontroli terenowej należą działania realizowane po zakończeniu sytuacji wywiadu ankierskiego, tj. **powtórny wywiad (1) osobisty lub (2) telefoniczny oraz (3) ankieta pocztowa**. Należy przy tym pamiętać, by kontrola terenowa była przeprowadzana najpóźniej kilka dni po badaniu (ze względu na zapamiętywanie respondenta) i obejmowała te elementy badania, które są łatwo zapamiętywane, a więc nie treść 43 pytań kwestionariusza, a raczej fakty łatwe do skojarzenia – laptop/ jego brak, płeć ankietera (ale nie jego wiek), ogólna tematyka wywiadu (ale raczej nie czas jego trwania gdyż jest to subiektywne), stosowanie materiałów pomocniczych (zdjęcia, karty do pytań). Kontroli powinna podlegać także generalna ocena zachowania ankietera. Niemniej jednak taka ocena zachowania ankietera służy bardziej do poprawy działania sieci, a w odniesieniu do konkretnego badania nic nie wnosi.

Ponadto kontrola terenowa powinna generalnie mieć charakter statystyczny (losowy dobór respondentów / wstępnych adresów do kontroli). Przy celowym doborze do kontroli nie ma możliwości oceny statystycznej jakości zebranego materiału, dlatego też warto to ustalić z wykonawcą przed badaniem. Kontrola terenowa – celowa stosowana jest w przypadku wykrycia uchybień w pracy jakiegoś ankietera. Mając na względzie, że ogólna charakterystyka może nie być zbyt rzetelna, warto „pociągnąć wątek”, by usunąć dane nieprawidłowe, jeżeli zostały zrealizowane niezgodnie z instrukcją i zastąpić je nowymi (o ile pozwoli na to czas) lub jedynie usunąć. Jeśli dobór respondenta był losowy, to przed kontrolą terenową potrzebna jest analiza kontaktów ankietera z respondentem metodą nieterenową (odpowiednia karta doboru i jej zakodowanie). Kontrola terenowa powinna objąć zarówno wywiady zrealizowane jak i niezrealizowane oraz powody ich niezrealizowania. Pozwala to wskazać ankieterów do kontroli terenowej o nietypowych statystykach charakterystyki poruszania się w terenie oraz efektywności ich pracy. W przypadku próby losowej warto, aby na formularzu doboru znalazł się opis budynku, w którym mieszka respondent (liczba pięter i piętro, na którym mieszka). Daje to ewentualną możliwość obserwacji kolejnych aspektów pracy ankietera.

Niezwykle istotnym jest również, by w trakcie kontroli terenowej zwrócić uwagę na sposób sformułowania pytań kontrolnych. Pytania nie mogą sugerować respondentom odpowiedzi. Tak jak stwierdziliśmy wcześniej, respondentów należy pytać o kwestie faktualne, a nie zmienne i bezsporne, ale także niewymuszające po prostu potakiwania. Należy pamiętać, że w wyobrażeniu około 30% respondentów, ankieter to biedny student, który chce sobie „dorozić”. Toteż nawet kiedy dany respondent nie brał udziału w badaniu, będzie mimo wszystko potwierdzać swój udział, gdyż nie chce skrzywdzić wyimaginowanego młodego człowieka.

KONTROLA NIETERENOWA oparta jest o analizę wypełnionych kwestionariuszy. Jej zadaniem jest dostarczenie informacji w zakresie poprawności informacji gromadzonych przez ankietera. Zakres tego typu kontroli uwarunkowany jest tematyką badania i zależy od logicznych relacji między poszczególnymi odpowiedziami respondentów na poszczególne pytania. Do metod nieterenowych należą przede wszystkim:

- (1) **analiza kwestionariusza**, jego kompletności i poprawności oraz ewentualnych wewnętrznych sprzeczności odpowiedzi. Popularną metodą jest wstawianie do kwestionariusza pytań bądź odpowiedzi nieprawdziwych lub stwierdzeń wykluczających się. Prawdopodobieństwo, że na przykład lekarz lub inny fachowiec w jakiejś dziedzinie, wskaże nazwę leku, który nie istnieje, jest stosunkowo małe. Jeżeli jednak tak się stanie, może to być sygnał, że niektóre pytania nie były zadawane respondentowi, a ankieter zaznaczył tę odpowiedź sam;
- (2) **analiza efektu ankierskiego** np. w obrębie różnic odpowiedzi uzyskiwanych przez różnych ankieterów;

- (3) **opis i weryfikacja sytuacji wywiadu na podstawie formularza dla ankietera**, np. można prosić ankieterów o fotografie charakterystycznego obiektu lub budynku dla danej miejscowości, w której jest realizowany wywiad lub jak w wypadku badań „business to business” prosić pieczętkę instytucji, (zebranie takich pieczętek, fotografii uwiarygodnia dotarcie do respondenta). Można również podczas analizy zebranego materiału analizować liczbę danych kontaktowych, gdyż, jak utrzymują działy kontroli, uczciwi ankieterzy mają znikomy odsetek braków i odmów podawania numerów telefonów, co w prosty sposób przekłada się na łatwość ich skontrolowania. Standardowy odsetek kontaktów telefonicznych to jakieś 85–90% – ankieterzy zbierający mniej danych do potwierdzenia wywiadu albo mają coś do ukrycia, albo mają zły warsztat ankieterski (np. zła aranżacja końcowa – nie wzbudzają zaufania i nie potrafią wytłumaczyć respondentowi celu kontroli i zbierania danych);
- (4) **analiza czasów realizacji badania**. Zdigitalizowanie technik badawczych zwiększyło paletę sposobów kontroli ankieterów i bardzo uwiarygodniło ich pracę. Obecnie istnieje możliwość porównania czasu realizacji badania z namierzaniem poruszania się ankieterów – ankieterzy są namierzani poprzez GPS lub telefon komórkowy – dane geolokacyjne położenia ankieterów łączone są z czasami realizacji wywiadów z ich laptopów lub ze średnim czasem trwania realizacji wywiadu w przypadku badania realizowanego techniką PAPI. (por. rys. 4).

Weryfikacja zebranego materiału i eliminacja materiału niespełniającego kryteriów

Wynikiem kontroli powinien być raport pokontrolny. Raport pokontrolny, który trafia do rąk klienta, powinien rozpoczynać się od podstawowych informacji, takich jak:

- przedmiot, czas realizacji kontroli, metody kontroli i kwestionariusz kontrolny;
- liczbę skontrolowanych przypadków w relacji do wielkości efektywnie zrealizowanej próby;
- liczbę skontrolowanych ankieterów w poszczególnych regionach z uwzględnieniem liczby przypadków dla każdego kontrolowanego ankietera;
- podstawowe statystyki prezentujące zbiorcze wyniki kontroli z uwzględnieniem a) liczby przypadków zastrzeżeń podstawowych, b) liczby wywiadów niezrealizowanych c) liczbę wywiadów zrealizowanych z inną osobą niż wskazana jako respondent,
- wskazania zastrzeżeń proceduralnych, tj.: a) złego doboru respondenta lub adresu, b) niezgodności danych demograficznych respondenta, c) niezgodności czasu trwania wywiadu, d) niewykorzystania kart respondenta oraz innych materiałów, e) pominięcia pytań, pytań filtrujących, bloków pytań, f) liczby rozbieżności w pytaniach powtórzonych (ocena rzetelności pomiaru).

Jeżeli kontrola badania była przeprowadzona przez wykonawcę, raport powinien także zawierać informację, jakie kroki poczyniono na skutek stwierdzenia zaistniałych uchybień i jakie podjęto działania w celu wyeliminowania ich wpływu na wyniki badania. Oczywiście, choć nie zawsze możliwe, wydaje się powtórzenie nieprawidłowo zrealizowanych przypadków. Uchybienia ankieterskie wykrywane podczas kontroli terenowej można zasadniczo podzielić na 3 rodzaje:

- **Błędy podstawowe** – najczęściej zaliczają się do nich potocznie zwane „sufty” oraz „podmiany”. Suft to sytuacja, w której kontrola stwierdziła fakt niezrealizowania wywiadu z osobą wskazaną przez ankietera na formularzu doboru lub ewentualnie, kiedy dane kontaktowe na formularzu doboru są nieprawdziwe – na przykład nieistniejący adres. Podmiana to sytuacja, kiedy w wyniku kontroli okazuje się, że zamiast respondenta spełniającego parametry doboru (na formularzu doboru znaj-

Nazwa badania: XXXX
 Typ badania: CLT
 Termin kontroli: 22 kwietnia br.
 Rodzaj kontroli: Terenowa CATI
 Zasięg terytorialny: Warszawa, Katowice, Białystok, Lublin, Gdańsk, Poznań, Radom, Wrocław
 (wszystkie ośrodki, które realizowały badanie)
 Próba badania zasadniczego: 130
 Próba kontrolna: 23 przypadki

Kontrolę zrealizował Dział Kontroli Jakości Instytutu. Kontrolę realizowało dwóch kontrolerów. Byli to profesjonalni kontrolerzy, niezależni od Sieci Ankieterskiej realizującej badanie zasadnicze. Oznacza to, że osoby te nie uczestniczą w badaniach realizowanych przez Instytut. Kontrola ta miała charakter powtórnego kontaktu CATI, opartego o wywiad kwestionariuszowy, co pozwoliło na uzyskanie jednakowego zestawu informacji od wszystkich respondentów. Wszystkie wywiady kontrolne zostały nagrane. Zostało zrealizowanych 23 wywiadów kontrolnych, co stanowi 17,7% próby.

Badanie było realizowane przez 48 ankieterów i 25 rekruterów, jeden ankieter zrealizował maksymalnie 7 wywiadów, a jeden rekruter zaprosił maksymalnie 16 respondentów.

Ankieterzy mają obowiązek notowania danych teleadresowych respondentów, które następnie są analizowane pod względem powtarzających się numerów telefonów, adresów, występowania adresów i telefonów do instytucji budek telefonicznych itp. W wyniku tej analizy nie stwierdzono uchybień.

	Błędy w realizacji
Sfalszowanych wywiadów (wywiad nie odbył się)	0 (0%)
Nie testowano produktu	0 (0%)
Nie pokazywano kolorowych materiałów	0 (0%)
Brak legitymacji ankieterskiej	0 (0%)
Niezachowana 6-miesięczna karencja	1 (4,3%)
Respondent używał wymaganych produktów	0 (0%)
Respondent zbyt rzadko używa wymaganych produktów	0 (0%)
Zły wiek respondenta	1 (4,3%)
Inne zastrzeżenia	0 (0%)

W wyniku kontroli stwierdzono 1 przypadek, w którym nie została zachowana 6-miesięczna karencja. Również w jednym przypadku kontrola wykazała niezgodność wieku respondenta (o 4 lata). W tym przypadku respondent powinien zostać zakwalifikowany do innej kohorty. Odsetek zgodnych odpowiedzi jest bardzo wysoki, a wyniki kontroli wskazują na dużą rzetelność w doborze respondenta.

Rys. 6. Przykładowa strona z raportu kontrolnego

dują się dane właściwego respondenta), w jego imieniu wypowiadała się inna osoba – najczęściej inny członek tego samego gospodarstwa domowego (żona za męża, matka za córkę etc). Zarówno w wypadku podmian, jak i sufitów mamy do czynienia z sytuacją, kiedy osoba, która powinna być respondentem, w ogóle nie brała udziału w badaniu.

- **Błędy dyskwalifikujące** – zrealizowany wywiad z powodu fundamentalnych kwestii, takich jak to, że respondent nie spełniał parametrów doboru badania (miał być w wieku 25–45 lat, a miał 22 lata, czy też częściej miał czegoś używać, a nie używał, miał być beneficjentem jakiegoś programu, a nie był itp.) lub został dobrany niezgodnie z procedurą badania, np. nie został dobrany metodą losową, a miał być.

- **Błędy proceduralne** – mogące mieć istotny wpływ na uzyskane odpowiedzi, tj. np. błędna aranżacja wywiadu, błędny dobór kwestionariusza, zmiana kolejności bloków pytań itp.
- **Zastrzeżenia** – nie rzutujące na sam fakt przeprowadzenia wywiadu, jednak wskazujące na nieprawidłowości, np. ankieter nie posiadał legitymacji lub według respondenta w pytaniu 23 nie pokazywał karty do pytań.

Trudno pośród wymienionych powyżej metod kontroli wskazać najlepszą ze względu na istotne różnice zarówno techniczne, jak i ze względu na zakres dostarczanych informacji oraz możliwości pragmatycznego zastosowania. Możemy jedynie wskazać te, które są najczęściej stosowane. Wśród metod kontroli terenowej najczęściej przeprowadzaną jest kontrola telefoniczna. Z tego powodu zleceniodawca powinien zagwarantować sobie w niej udział – powinien móc typować wywiady do kontroli, a także móc odsłuchiwać wywiady kontrolne. Analogicznie do kontroli badania CATI, gdzie klient ma prawo odsłuchiwać wszystkie wywiady wraz z supervisorem studia telefonicznego CATI. Zapewne też, wraz z rozwojem techniki, coraz więcej firm będzie sięgało po rozwiązania typu obserwacja miejsca i czasu realizacji wywiadu za pomocą GPS.

Dla zamawiającego drugorzędny jest fakt sięgnięcia po jedną czy drugą metodę, jego perspektywa to „wartość oraz jakość zebranych danych, ich wiarygodność i zgodność ze stanem faktycznym”. Ocena rzetelności pomiaru to proces, w którym nie musi uczestniczyć bezpośrednio, ale który musi uwzględnić w procesie badania zarówno w wymiarze finansowym, ludzkim, jak i czasowym. Wiele elementów kontroli w sposób bezkosztowy można przerzucić od razu na wykonawcę (przykładowo wymagać zestawień statystycznych z realizacji zaznaczając w umowie, że mają one być elementem raportu technicznego). Ponadto zamawiający nie musi zdawać się jedynie na wykonawcę bądź własne doświadczenie i możliwości. Obecnie istnieją niezależne podmioty realizujące kontrolę zewnętrzną. Już sam proces przekazywania danych, a w szczególności procedury ochrony danych respondenta wykażą jakość działań wykonawcy. Wykonawca, który nie ma nic do ukrycia, zrobi to bez żadnych zastrzeżeń, a poszanowanie dla danych osobowych respondenta i sposobu przekazania do kontroli pokaże poziom kompetencji i respektowania prawa.

Podsumowanie

Mamy nadzieję, że ten krótki artykuł nie pozostawia wątpliwości, iż dobra jakość wyników badań nigdy nie jest ani kwestią przypadku, ani rezultatem myślenia życzeniowego. Bez wątpienia wymaga przemyślanych działań długo wcześniej, niż badanie trafi do realizacji terenowej. W przypadku zamówień publicznych możemy nawet mówić o rozpoczęciu tego procesu już po stronie zamawiającego na etapie konstrukcji zapisów SIWZ oraz zapisów umowy.

Na pewno artykuł ten nie wyczerpuje wszystkich kwestii i może być jedynie wstępem do budowania rozważań na temat indywidualnych strategii realizacji oraz jakości badań ewaluacyjnych. Jak wskazuje Edwards Deming „jakość to sposób myślenia, który powoduje, że stosuje się i bez przerwy poszukuje najlepszych rozwiązań”. Możemy mieć jedynie nadzieję, że powyższe przybliżenie zagadnień kontroli jakości danych terenowych ułatwi czytelnikowi implementację najlepszych rozwiązań we własnych projektach.

Agnieszka Haber – absolwentka Instytutu Socjologii UW, Wydziału Stosowanych Nauk Społecznych UW oraz Katedry Zarządzania Projektami SGH w Warszawie. Od 2004 roku kieruje Sekcją Ewaluacji w Polskiej Agencji Rozwoju Przedsiębiorczości. Jest odpowiedzialna za realizację ewaluacji programów przedakcesyjnych oraz strukturalnych wdrażanych przez PARP w obszarach rozwoju sektora MSP, Kapitału Ludzkiego, Infrastruktury. Problematyką badań społeczno-gospodarczych zajmuje się od 10 lat, zaś obszarem ewaluacji i zarządzania interwencjami publicznymi od 7 lat. Od 2005 roku organizator cyklicznej Konferencji Ewaluacyjnej odbywającej się pod auspicjami PARP i MRR (www.konferencja-ewaluacja.pl). Od 5 lat prowadzi szkolenia oraz wykłady w ramach studiów podyplomowych z ewaluacji programów o charakterze społeczno-gospodarczym (UW, SGH oraz WSPiZ im. L. Koźmińskiego). Jest autorką szeregu publikacji i opracowań z dziedziny ewaluacji programów sektora publicznego.

Mirosław Kalinowski – praktyk badań marketingowych, Dyrektor Realizacji Badań w MilwardBrown SMG/KRC. Z branżą badawczą związany od 19 lat, a więc od początków jej funkcjonowania w warunkach gospodarki rynkowej. Posiada doświadczenia w realizacji prawie wszystkich rodzajów badań na każdym poziomie odpowiedzialności – ankietera ilościowego, jakościowego, ankietera CATI, rekrutera, audytora, supervisor, koordynatora regionalnego, kontrolera, szefa sieci ankieterskiej.

Efekt przyczynowy w ewaluacji

Ewaluacja programów publicznych w Polsce wobec nowych wyzwań

W ostatnich latach, wraz z wkroczeniem na arenę polityk publicznych programów finansowanych ze środków Unii Europejskiej, na agendzie pojawiło się zagadnienie ewaluacji programów publicznych. Zaowocowało to szeregiem publikacji, które rozważały pojęcie ewaluacji i jej funkcje (np. Olejniczak 2008a, Olejniczak 2008b, Ferry, Olejniczak 2008, Korporowicz 2007, Górniak 2007). Jest to polski odprysk międzynarodowej debaty nad ewaluacją, czy szerzej: analizą programów i polityk publicznych.

Rozumienie pojęcia ewaluacji, które wyłania się z tych rozważań, jest zróżnicowane, co wynika z odmiennych perspektyw teoretycznych i metodologicznych, jak również z rozłożenia akcentów pomiędzy funkcje pełnione przez ewaluację. Tutaj przyjmiemy dość szeroko i pragmatycznie stosowane rozumienie ewaluacji jako przedsięwzięcia zmierzającego do określenia w oparciu o właściwe zgromadzone i przetworzone informacje, w jakim stopniu dane rozwiązanie (np. interwencja publiczna: polityka, program lub projekt) spełnia ustalone kryteria, w tym, w szczególności, w jakim stopniu osiągnęło zakładane cele oraz jakie są relacje pomiędzy nakładami, działaniami i wynikami tego rozwiązania (por. Górniak 2007). Ewaluacja stanowi integralny element cyklu polityk publicznych. Klasycznie, stanowiła element zamykający ten cykl, dając możliwość analizy, podsumowania i oceny wdrożonego programu. Obecnie pojęcie ewaluacji stosuje się również do oceny programu przed podjęciem decyzji o jego wdrożeniu (*ex ante*) i w trakcie realizacji programu (*on-going, mid-term*), gdy można jeszcze dokonywać modyfikacji zarówno w zakresie samej orientacji strategicznej, jak i sposobu zarządzania wdrożeniem.

Warto również wspomnieć w tym miejscu o rozwoju ewaluacji pod sztandarami audytu wydatków publicznych. Wraz z reformami systemu planowania budżetu państwa i pojawienia się budżetowania zadaniowego w wielu krajach OECD powstała potrzeba zorganizowania systemu pomiaru i analizy efektów wydatków publicznych, który został określony mianem audytu efektywnościowego lub audytu efektów – *performance audit*. (por. OECD 2007). **Audyt efektywnościowy** rozumiany jest jako systematyczny, celowy i zorganizowany proces rzetelnej, obiektywnej i niezależnej oceny planowanych, realizowanych i zakończonych polityk, programów i projektów publicznych oraz oceny bieżącego funkcjonowania jednostek sektora finansów publicznych w oparciu o kryteria gospodarności (*economy*)¹, efektywności (*efficiency*)² i skuteczności (*effectiveness*)³ ich działań. Audyt efektywnościowy służy doskonaleniu zarządzania wewnętrznego działaniami podmiotów publicznych, a jednocześnie ma wpływać na poprawę jakości polityk i programów publicznych, w tym usług publicznych. Audyt efektywnościowy zmierza do uzyska-

¹ Gospodarność pojmowana jest jako relacja pomiędzy wykorzystanymi zasobami a jakością otrzymanego dzięki temu dobra lub usługi.

² Efektywność pojmowana jest jako relacja między nakładem a wynikiem. Poprawa wyniku bez wzrostu nakładów lub otrzymanie tego samego wyniku przy zmniejszeniu nakładów oznacza wzrost efektywności.

³ Skuteczność odnosi się do poziomu osiągnięcia założonych celów. To jeden z najtrudniejszych wymiarów oceny.

Wprowadzanie budżetu zadaniowego w Polsce prowadzi do radykalnego rozszerzenia zakresu ewaluacji. W tym kontekście ewaluacja programów współfinansowanych z funduszy strukturalnych jest prawdziwym poligonem doświadczalnym, z którego bardzo szybko trzeba będzie przenosić doświadczenia na całe „pole walki” w postaci wszystkich jednostek sektora finansów publicznych. Niniejsze rozważania wpisują się zatem w szeroki nurt przygotowań do zmiany modelu planowania, wdrażania i rozliczania wydatków publicznych w sposób zorientowany na osiąganie efektów społecznych, a nie tylko wykonywanie przepisów.

Noblista, prof. J. Heckmann, zwraca uwagę na to, że ewaluacja polityk publicznych staje wobec trzech kluczowych problemów:

1. Ewaluacji wpływu (przyczynowego) przeszłych interwencji na obserwowane efekty, w tym w ich wpływ na dobrobyt osób, na które interwencja jest ukierunkowana.
2. Przewidywanie wpływu interwencji już wdrożonych w jednym środowisku, w innych środowiskach, z uwzględnieniem dobrobytu.
3. Przewidywanie efektów interwencji nigdy niewprowadzanych w nowych warunkach środowiskowych, z uwzględnieniem skutków dla dobrobytu (por. Heckmann 2005).

Każde z tych kluczowych zagadnień ewaluacji jest związane z odpowiedzią na pytania o mechanizm związków przyczynowych uwikłanych we wprowadzaną interwencję. Do tej pory w dość zadowalający sposób jesteśmy sobie w stanie poradzić z pierwszym z tych zagadnień, co w pewnym stopniu można wykorzystać do rozwiązania problemu drugiego. Trzeci stawia nas przed szczególnie trudnymi wyzwaniami i wymaga dysponowania koncepcjami teoretycznymi ujętymi w możliwie pełne modele strukturalne zależności przyczynowych w programie⁴.

Ten artykuł zwraca uwagę na szczególny aspekt ewaluacji, na który w mniejszym stopniu zwraca się uwagę w sytuacji, gdy ciągle powszechne są kłopoty z pomiarem efektów, a nawet prawidłowym definiowaniem celów w programach publicznych (por. Górniak, Keler 2007 i 2008): chodzi w nim o badanie związków przyczynowych pomiędzy podejmowanymi działaniami, które znajdują swój wyraz w produktach (*outputs*) interwencji a efektami w postaci rezultatów i oddziaływań programu⁵. W artykule przedstawiam argumenty na rzecz znaczenia tego aspektu ewaluacji oraz przedstawiam niektóre problemy związane z badaniem związków przyczynowych w programach publicznych. Wskazuję też na znaczenie badania (przyczynowo rozumianych) konsekwencji programów w kontekście przygotowania sektora publicznego do wdrożenia budżetu zadaniowego. W tym aspekcie wykraczam poza ujęcia zagadnienia związków przyczynowych w politykach publicznych rozwijane na gruncie statystyki (Donald B. Rubin) i ekonometrii (James J. Heckmann)⁶.

Dlaczego należy badać związki przyczynowe pomiędzy interwencjami publicznymi a efektami?

Efekty programów (polityk) publicznych – sukcesy w ich wdrażaniu – związane są z wieloma czynnikami (por. klasyczny w tym zakresie tekst: Mazmanian, Sabatier 1989). Wśród nich na pewno są: jasność

⁴ Zdaniem J. Heckmanna dwa poprzednie, dla uzyskania właściwej odpowiedzi, też wymagają konstrukcji opartych na teorii, pełnych modeli strukturalnych. O to, czy jest to konieczne, toczy się spór (por. Heckmann 2005b; Sobel 2005).

⁵ Użycie terminów: produkt, rezultat i oddziaływanie przyjmuję w tym miejscu zgodnie z nieco zmodyfikowaną konwencją obecną w dokumentach metodologicznych Komisji Europejskiej; modyfikacja dotyczy rozumienia terminu rezultat, którego zakres rozszerzony jest nie tylko na natychmiastowe, lecz także odłożone w czasie (trwałe) zmiany na poziomie bezpośrednich beneficjentów, które KE określa mianem oddziaływań lokalnych (por. Górniak, Keler 2008).

⁶ Wymienione tu nazwiska wybitnych profesorów D. Rubina i J. Heckmanna zostały podane jako sztandarowe dla głównych obecnie nurtów debaty nad badaniem efektów przyczynowych, określanych w skrócie jako statystyczny i ekonometryczny.

i precyzja celów, zapewnienie odpowiednich zasobów oraz prawidłowe rozpoznanie związków przyczynowo-skutkowych istotnych dla programu. To ostatnie zagadnienie jest szczególnym przedmiotem niniejszego artykułu.

Rozpoznanie zależności przyczynowych jest ważnym zadaniem diagnozy problemów, często zaniebywanym ze złymi skutkami dla programowania rozwiązań. Prosty przykład problemu diagnostycznego – bez podania rozstrzygnięcia: czy fakt chodzenia do przedszkola sprzyja odnoszeniu przez dzieci sukcesu szkolnego. Jako sukces szkolny możemy operacyjnie przyjąć np. wyniki testu szóstkłasiisty lub – po powszechnym wprowadzeniu – trzeciokłasiisty⁷. Ten argument jest bowiem często przywoływany w debacie nad upowszechnieniem edukacji przedszkolnej. Badanie diagnostyczne powinno zostać zaprojektowane tak, by istniała możliwość analitycznego oddzielenia efektu samego przejścia przez edukację przedszkolną od oddziaływania wielu rozmaitych czynników, które wpływają zarówno na sukces szkolny, jak i na wybór (także możliwość wyboru) drogi przez przedszkole lub z jego pominięciem. Podobnie, przy badaniu wpływu liczebności klas szkolnych na wyniki nauczania nie można po prostu korelować ze sobą wyników powszechnych, porównywalnych testów kompetencji i liczebności klas, lecz trzeba tak skonstruować badanie, by kontrolować wpływ możliwie wszystkich istotnych czynników decydujących o wynikach nauczania i liczebności klas. W analizach diagnostycznych wnioski o zależnościach przyczynowych zbyt często opiera się na prostych zależnościach dwóch zmiennych, co w konsekwencji prowadzi do oparcia programów na nadmiernie uproszczonych, czy wręcz błędnych założeniach.

Każda interwencja publiczna, niezależnie od jej genezy, jest swoistą „teorią”: opiera się na hipotezie, że w danych warunkach podjęcie działań wynikających z projektu interwencji doprowadzi do osiągnięcia założonych skutków. Zakładaną w projekcie strukturę zależności między przyjętym celem a zastosowanymi środkami – instrumentami programu określa się mianem „teorii programu”. Jej dopełnieniem jest zbiór założeń dotyczących najlepszego sposobu zarządzania wdrożeniem. Rozpoznanie „teorii programu” i „teorii wdrożenia” dominującej w gronie osób projektujących interwencję i decydentów pozwala na lepsze zrozumienie polityki czy programu i jej pogłębioną ewaluację⁸. Ewaluacja, w szczególności końcowa, sprawdza, czy w świetle uzyskanych efektów teoria programu daje się utrzymać, czy też należy ją odrzucić. Konceptcje działania, które pomyślnie przejdą ewaluację, tworzą korpus wiedzy podnoszącej skuteczność polityk publicznych. Faktycznie, spontaniczny proces weryfikacji występuje zawsze tak czy owak w postaci społecznej oceny skutków prowadzonych polityk. Ewaluacja jest metodologicznie kontrolowanym i usystematyzowanym sposobem takiej weryfikacji prowadzonym przez ekspertów.

Jeśli tak rozumiemy zadanie ewaluacji – zwłaszcza końcowej – to musi ona opierać się na badaniu zależności przyczynowych w sensie pierwszego zadania z listy Heckmannowskich postulatów wobec ewaluacji. W przypadku ewaluacji *ex ante* stajemy przed spełnieniem co najmniej postulatu drugiego, czyli ekstrapolacji dotychczasowych zależności na nowe warunki. W wielu sytuacjach, wobec braku bazy wiedzy opartej na dobrze przeprowadzonych ewaluacjach, musimy się opierać na eksperymencie intelektualnym wspomaganym ustaleniami o cząstkowych zależnościach. Taki eksperyment myślowy opiera się na modelu teoretycznym, który powinien być zoperacjonalizowany w postaci systemu równań strukturalnych – tu jesteśmy w objęciach trzeciego postulatu Heckmanna, najtrudniejszego do realizacji. Jest to wizja prowadzenia polityki, którą możemy określić mianem rządzenia opartego na przewidywaniach (*predictive governance*). Termin ten jest analogią do nazwy rozwijanej w sektorze przedsiębiorstw orientacji strategicznej określanej jako *predictive enterprise*. Czy *predictive governance* jest możliwe? Sądzę, że

⁷ Można też zawęzić problem do zagadnienia wyrównywania szans lub jeszcze inaczej go sprecyzować; w tym miejscu nie chodzi o faktyczną diagnozę tylko ilustrację metodologiczną.

⁸ Por. J.D. Brickmayer, C.H. Weiss, *Theory-Based Evaluation in Practice. What Do We Learn?* Evaluation Review, Vol. 24 No. 4, August 2000, s. 407–431.

za sporem między nurtami analizy przyczynowej w programach publicznych kryje się też różnica co do wiary w możliwość „zapanowania nad przyszłością”. Jednak nawet przy dużej pokorze wobec takiej perspektywy, badanie efektów przyczynowych jest możliwe i potrzebne.

Ewaluacja na pewno ma potencjał falsyfikacyjny – odrzucania interwencji nieskutecznych, a często wręcz przeciwnie, nawet jeśli mamy wątpliwość co do tego, czy jest w stanie skutecznie potwierdzić, że osiągnięte efekty są wynikiem interwencji, a nie czynników od niej niezależnych, a zwłaszcza co do możliwości przewidywania efektów nowych interwencji. To powoduje pewną rezerwę decydentów wobec ewaluacji. Trzeba jednak podkreślić, że nawet jeśli ewaluacja potrafi tylko wskazać, że program nie przynosi efektu przyczynowego, to pozwoli to zaoszczędzić środki publiczne, które mogą być z pożytkiem wydane na inne cele (co też powinno, oczywiście, podlegać ewaluacji). Ewaluacja powinna jednak zmierzać też do ustalenia, dlaczego program nie działa, gdyż sam brak skuteczności interwencji nie oznacza unieważnienia problemu, dla którego rozwiązania program został podjęty. Brak efektów może wynikać z błędnej „teorii wdrożenia”, a nie wadliwej „teorii programu”. Można też utrzymać cele programu zmieniając koncepcję skutecznej interwencji.

Bardzo popularne jest oczekiwanie wobec ewaluacji zawarte w tytule znanej książki Pattona: „Ewaluacja skoncentrowana na wykorzystanie”⁹. Autor podkreśla, że „głównym wyzwaniem profesjonalnej praktyki pozostaje – wykonywanie ewaluacji, które są użyteczne i rzeczywiście wykorzystywane”. Takie podejście zapewne sięgnie po metody jakościowe i po partycypację, ale nie zrezygnuje z obiektywnego i porównawczego pomiaru oraz dobrze wypracowanych standardów metodologii badań, w duchu „multiplizmu”, czyli wzajemnego uzupełniania i kontrolowania wyników stosowania różnych metod badań, zarówno ilościowych, jak i jakościowych. Kluczowe dla dobrej ewaluacji jest przy tym właściwe rozpoznanie teorii programu i ustalenie zależności przyczynowych pomiędzy interwencją a efektami.

Pragmatyczne podejście do ewaluacji prezentują też inni popularni autorzy: Pawson i Tilley, którzy w swojej wizji „ewaluacji realistycznej” kładą nacisk na rozpoznanie związków przyczynowych funkcjonujących w obrębie programu w celu uzyskania odpowiedzi dlaczego program działa (lub nie), komu przynosi efekty i w jakich warunkach. Punktem wyjścia musi być, oczywiście, zdolność ustalenia samych efektów, a także przyczynowego powiązania ich z podejmowanymi działaniami. W tym celu należy wykorzystywać wszystkie dostępne podejścia badawcze (pluralizm metodologiczny) do rozpoznania struktury, ludzi i ich przekonań, w oparciu o które program jest podejmowany i realizowany (w tym przekonań na temat „co działa”); kontekstu społecznego, ekonomicznego i geograficzno-przyrodniczego; regularności, które występują w realizacji programu i wreszcie – ustalenia zmiany wywołanej przez program i odseparowanie jej od wpływu innych czynników społecznych¹⁰. O ile podejście Heckmanna koncentruje się na samym zagadnieniu operacjonalizacji koncepcji przyczynowych w postaci modelu strukturalnego i jego estymacji, to literatura z nurtu ewaluacji realistycznej mocniej stawia zagadnienie samego dochodzenia do adekwatnej koncepcji teorii programu.

Komisja Europejska stoi wyraźnie na gruncie takiego realistycznego, pragmatycznego podejścia. Przewodniki metodologiczne wyraźnie akcentują potrzebę ustalenia w ewaluacji efektu interwencji. Mocno akcentowana jest kwestia właściwego pomiaru wyników programu, w tym problem adekwatnego wykorzystania wskaźników. Nacisk kładziony jest zatem na metody ilościowe, obiektywne i dające możliwość porównania wyników ewaluacji także pomiędzy krajami członkowskimi. Podkreśla się w szczególności potrzebę ustalenia efektu netto – zmiany wywołanej bezpośrednio interwencją, której nie można przypisać samoistnemu oddziaływaniu innych czynników. Jednym z kluczowych zadań ewaluacji jest w ujęciu

⁹ M.Q. Patton, *Utilization-Focused Evaluation*, 3rd edn, Sage, London 1997.

¹⁰ Por. R.N. Pawson, Tilley *Realistic Evaluation*, Sage, London, 1997.

Komisji Europejskiej ustalenie związku przyczynowego między podjętą interwencją a obserwowanymi wynikami. Ustalenie przyczynowego wpływu interwencji wymaga:

- Zmierzenia efektu całkowitego interwencji jako ogólnej zmiany w zakresie zdefiniowanym przez cel (zmiany na poziomie przyjętych wskaźników).
- Oddzielenia zmiany niezależnej od podjętej interwencji od zmiany, którą można przypisać wpływowi interwencji (efekt netto).

Ustalenie efektu netto wymaga zatem oszacowania, co by się stało, gdyby nie została podjęta interwencja. Tak formułuje to podręcznik wydany przez Komisję Europejską:

„Gdy mówimy, że dane efekty zostały wyprodukowane lub spowodowane przez program, oznacza to, że gdyby program nie miał miejsca, lub miałyby miejsce w innej formie lub stopniu, te efekty nie wystąpiłyby lub nie wystąpiłyby w tym samym stopniu. Oznacza to, że ważne jest, by mieć precyzyjny obraz tego, co stałoby się bez programu. Nazywa się to **sytuacją kontrfaktyczną**”¹¹. Skutki, które wystąpiłyby nawet w sytuacji, gdyby program wydatków publicznych nie miał miejsca (sytuacja kontrfaktyczna), nazywamy efektem *deadweight*. Ponadto na obserwację skutków interwencji ma wpływ także efekt substytucji lub wyparcia (przesunięcia) – sytuacja, w której efekty programu w odniesieniu do pewnych osób, grup lub obszarów zostały uzyskane kosztem innych osób, grup lub obszarów. Ten rodzaj efektów jest bardzo trudny do uchwycenia. Dodajmy, że jego wystąpienie jest bardziej prawdopodobne w sytuacji, gdy interwencja wdziera się na obszar normalnego funkcjonowania konkurencji rynkowej, a nie pozostaje w obszarze, na którym rynek zawodzi jako regulator. Przykładem interwencji naruszających konkurencję rynkową i dlatego na ogół nieskutecznych w sensie efektu netto są bezpośrednie dotacje dla przedsiębiorstw, które odwracają uwagę firm od konkutowania o względy klientów, zastępując to zabieganiem o względy jednostek przyznających środki publiczne. Efektem takiej orientacji interwencji może być w szczególności obniżenie jakości nawet całej oferty branży (tak się może stać wkrótce z rynkiem usług szkoleniowych, na którym skutecznie będą oferowane głównie szkolenia, na które zostanie zdobyta dotacja, a nie te, które w warunkach konkurencji wybraliby klienci).

Nacisk na ustalenie efektu netto wyraźnie wskazuje na oparcie ewaluacji na fundamentach przyczynowości. Wymaga obiektywizmu i ilościowego pomiaru. Pomiar efektu wymaga jednak uzupełnienia o dobre zrozumienie mechanizmu projektu, tego, co sprzyjało, a co hamowało jego realizację. Dobrze jest włączyć do tego wyjaśnienia wiedzę jak najszerszej grupy interesariuszy, także wiedzę dostępną przede wszystkim przy pomocy metod jakościowych.

Efekt netto jako miara przyczynowego efektu interwencji

Jak wynika z przeprowadzonych wyżej rozważań, ustalenie przyczynowego wpływu interwencji wymaga:

- zmierzenia efektu brutto interwencji jako ogólnej zmiany w zakresie zdefiniowanym przez cel (wskaźniki),
- oddzielenia zmiany niezależnej od podjętej interwencji od zmiany, którą można przypisać wpływowi interwencji (efekt netto).

Efekt netto jest to zatem różnica pomiędzy efektem całkowitym programu w grupie poddanej oddziaływaniu a efektem w przypadku niepoddania tej grupy oddziaływaniu. Ustalenie efektu netto (po-

¹¹ Komisja Europejska, *Ewaluacja programów wydatków Unii Europejskiej*. Przewodnik, 1997, s. 39.

miar oddziaływania przyczynowego) wymaga zmierzenia wielkości efektu w sytuacji, która nie może mieć miejsca w grupie docelowej, czyli stanu kontrfaktycznego¹².

Efekt netto jest miarą rzeczywistego skutku interwencji na poziomie beneficjentów bezpośrednich – rezultatu interwencji. Pomiar efektu netto umożliwia ustalenie rzeczywistej efektywności poniesionych nakładów.

Rozpoznanie związków przyczynowo-skutkowych oraz wielkości efektów i ich rzeczywistej efektywności pozwala na właściwe planowanie przyszłych interwencji.

Stan kontrfaktyczny opisuje, co by się stało z beneficjentami, gdyby nie uzyskali wsparcia. Zarówno u osób poddanych, jak i u nie poddanych oddziaływaniu potencjalnie mogą wystąpić oba rodzaje efektów: faktycznie zaobserwowane i te, których nie możemy zaobserwować, ze względu na to, że nie możemy poddać obserwacji tych osób w sytuacji kontrfaktycznej, czyli przeciwnej do tej, w której się znajdują. Dla grupy poddanej oddziaływaniu mamy obserwowany średni wynik interwencji: $E(Y1|W=1)$ i nieobserwowany (ukryty) średni wynik niepoddania interwencji: $E(Y0|W=1)$. Dla grupy niepoddanej oddziaływaniu mamy obserwowany średni wynik braku interwencji: $E(Y0|W=0)$ i nieobserwowany (ukryty) średni wynik poddania interwencji: $E(Y1|W=0)$.

W ramach analizy stanów kontrfaktycznych ewaluacja zmierza do ustalenia, jaka jest różnica pomiędzy efektem interwencji i efektem braku interwencji. Porównanie pomiędzy faktyczną sytuacją wśród beneficjentów interwencji po jej wprowadzeniu a sytuacją faktyczną osób niepoddanych interwencji (efekt całkowity):

$$E(Y1|W=1) - E(Y0|W=0)$$

jest tylko przybliżeniem uzyskanego efektu programu. Obrazuje zmianę wśród objętych interwencją ignorując możliwy wpływ na wielkość zmierzonego efektu końcowego wszystkich tych czynników, które wpłynęły na znalezienie się w grupie beneficjentów. Te czynniki mogą wpływać zarówno na selekcję ze strony dostarczających interwencji, jak i ze strony samych beneficjentów lub ich reprezentantów, np. rodziców osób niepełnoletnich (autoselekcja).

Badanie efektu netto należy traktować jako próbę oszacowania różnicy pomiędzy faktycznym efektem oddziaływania w grupie poddanej interwencji ($W=1$) a kontrfaktycznym efektem braku interwencji ($Y0$) w tej grupie $E(Y0|W=1)$, oszacowanym przy pomocy obserwowalnego efektu $E(Y0|W=0)$:

$$E(Y1|W=1) - E(Y0|W=1)$$

Kluczowym problemem ewaluacji nie jest oszacowanie $E(Y0|W=0)$ – efektu braku interwencji w grupie niepoddanej oddziaływaniu (efekt faktyczny), lecz $E(Y0|W=1)$, czyli: co by było, gdyby nie objęto interwencją tych, których objęto, a w konsekwencji ustalenie efektu netto dla beneficjentów jako różnicy pomiędzy faktycznym efektem interwencji a kontrfaktycznym stanem jej braku dla tej grupy. Im lepsza jest ocena stanu kontrfaktycznego (pełniejsza specyfikacja modelu), tym wyższa trafność wewnętrzna badania, czyli zdolność do uchwycenia prawdziwej wartości efektu przyczynowego w grupie poddanej oddziaływaniu.

Narzucającym się sposobem podejścia do problemu szacowania efektu netto jest analiza regresji, w której wynik na wskaźniku mierzącym efekt programu jest zmienną zależną, a zmienną niezależną fakt uczestnictwa bądź nie w programie oraz szereg zmiennych kontrolnych opisujących cechy badanych jednostek. Model taki byłby oparty o dane obserwacyjne, które nie pochodzą z badań opartych na zran-

¹² Bardzo dobry przegląd koncepcji wnioskowania przyczynowego w analizach społecznych zawiera artykuł: Winship, Sobel 2004.

domizowanych eksperymentach, lecz z badań opisowych, takich jak: sondaże, analiza danych administracyjnych czy ze statystyk publicznych.

$$Y_i = \alpha + \tau W_i + X_i' \beta + \varepsilon_i$$

Zwykła analiza regresji, która normalnie pozwala ustalić efekt netto interwencji, wprowadzanej do modelu jako zmienna 0/1, jest problematyczna, ze względu na występujące skorelowanie składnika resztowego ze zmienną reprezentującą oddziaływanie.

Zmienna niezależna W jest na ogół skorelowana ze składnikiem błędu ε_i , gdyż te same czynniki oddziałują na (samo)selekcję do wsparcia i na wynik (tzw. obciążenie selekcyjne – *selection bias*); ponadto osoby poddane interwencji mogą mieć cechy bardziej sprzyjające sukcesowi niż osoby z próby kontrolnej. Na skutek tego uzyskujemy oszacowanie efektu oddziaływania programu (τ), które jest obciążone i nie jest zgodne (zwiększanie próby nie zmniejsza obciążenia). W tym wypadku efekt deadweight nie jest dobrze odseparowany i w konsekwencji zakłócona jest ocena efektu netto.

W ostatnich dwudziestu latach, z naciskiem na lata ostatnie, notujemy silny rozwój prac nad technikami szacowania efektu kontrfaktycznego dla grup objętych interwencją. Z jednej strony jest to kontynuacja idei sformułowanych jeszcze w latach dwudziestych XX wieku przez Jerzego Neymana, a rozwiniętych zwłaszcza przez Donalda Rubina, które zmierzają do odtworzenia sytuacji kontrfaktycznej w oparciu o pomiar w grupie kontrolnej, nieobjętej interwencją¹³. To podejście, zwane statystycznym, zajmuje się konstrukcją metod doboru prób kontrolnych i dostosowanych do nich metod obliczania efektu interwencji.

Klasycznym podejściem zmierzającym do ustalenia stanu kontrfaktycznego grupy objętej interwencją polega na oparciu badania na próbach zrandomizowanych (podejście eksperymentalne). Eksperyment zrandomizowany traktowany jest tradycyjnie jako idealny układ dla pomiaru przyczynowego wpływu interwencji. Randomizacja polega na losowym podziale grupy docelowej na grupę poddaną interwencji i na grupę kontrolną, niepoddaną interwencji; pozwala na usunięcie systematycznego wpływu innych cech poza interwencją. W przypadku programów publicznych istnieje jednak szereg ograniczeń organizacyjnych, prawnych i etycznych, które utrudniają wprowadzenie randomizacji. Badanie eksperymentalne są kosztowne i trudne. Ponadto w przypadku randomizacji mogą wystąpić zjawiska, które zakłócają ocenę efektu netto. Problem oczywiście sprowadza się do odpowiedzi na pytanie, czy w przypadku uzyskanej przez randomizację próby kontrolnej $E(Y_0|W=0)$ rzeczywiście reprezentuje $E(Y_0|W=1)$? Musimy wziąć pod uwagę fakt, że w przypadku programów nie mamy do czynienia z eksperymentem laboratoryjnym, lecz w najlepszym wypadku z tzw. eksperymentem terenowym, rozciągniętym w czasie i nieodizolowanym od wpływu innych czynników otoczenia zarówno na grupę testową, jak i kontrolną. Może w związku z tym wystąpić tzw. „obciążenie randomizacyjne” oszacowań (Heckman i Smith, 1995): mierzony jest raczej efekt zamiaru uczestnictwa niż efekt uczestnictwa w programie (uczestnicy mogą odpaść, a nie-uczestnicy mogą znaleźć substytuty).

Badania nad czynnikami wpływającymi na efekt netto doprowadziły do wniosku, że kluczowe znaczenie ma dobór prób w taki sposób, by usunąć oddziaływanie efektu selekcji na wynik. Problem ten dobrze uwiadacznia omówiona wyżej próba zastosowania statystycznej metody kontroli wpływu czynników przy szacowaniu wpływu interwencji przy pomocy analizy regresji. W związku z ograniczeniami prób zrandomizowanych rozwinięto popularne w dziedzinie badań epidemiologicznych podejście

¹³ Na gruncie literatury ewaluacyjnej od lat sześćdziesiątych XX wieku istnieje silny nurt kładący nacisk na wykorzystanie podejścia eksperymentalnego i quasi-eksperymentalnego, którego fundatorem był D. Campbell.

oparte na dopasowaniu prób kontrolnych według cech badanych jednostek. Metoda prób dopasowanych (MATCHED SAMPLES) polega na doborze próby kontrolnej spośród osób nieobjętych interwencją w taki sposób, by każda z nich była jak najbardziej podobna (identyczna) pod względem cech wpływających na uczestnictwo w interwencji do odpowiedniej osoby z próby poddanej interwencji, z którą stanowi parę.

W przypadku prób dopasowania ze względu na wszystkie znaczące cechy natrafia się szybko na techniczny problem – brak u wystarczającej liczby odpowiednich przypadków kontrolnych. W odpowiedzi na ten problem opracowana została technika doboru prób kontrolnych w oparciu o tzw. ocenę skłonności – *propensity score matching* (PSM)¹⁴. PSM jest techniką umożliwiającą dobór prób dopasowanych z uniknięciem konieczności dopasowania przypadków pod względem wartości każdej ze zmiennych użytych do konstrukcji próby – dopasowanie odbywa się na podstawie wyniku uzyskanego z modelu regresji logistycznej udziału w interwencji na cechy determinujące ten udział.

Metoda kontrolnych prób dopasowanych, włączając w to PSM, ma też swoje słabości. Opiera się na założeniach, które nie zawsze są spełniane; założenia te są następujące:

- próba kontrolna reprezentuje stan kontrfaktyczny próby poddanej interwencji,
- zostały wzięte pod uwagę wszystkie zmienne wpływające istotnie na udział w interwencji,
- zmienne te wpływają niezależnie na każdą jednostkę,
- udział w interwencji jest decyzją podejmowaną przez każdą jednostkę niezależnie od decyzji innych jednostek.

Metoda ta ignoruje oddziaływania interwencji na równowagę ogólną, tzn. to, jaki wpływ ma program na zachowania i wyniki uzyskiwane przez podmioty niebędące przedmiotem oddziaływania, w tym efektów przesunięcia i substytucji, np. czy przeprowadzone szkolenia prowadzące do zatrudnienia beneficjentów nie spowodowały, że nie trafiły one do nie-beneficjentów, do których by w przeciwnym wypadku trafiły. Jest to nieistotne przy małej skali projektów w stosunku do rozmiarów danego rynku.

Na temat technik PSM i związanych z nimi sposobów liczenia efektu netto ukazały się już dwa teksty w publikacjach wydanych przez PARP (Konarski, Kotnarowski 2007; Strawirski 2008), dlatego nie podejmuję tutaj ich szczegółowego omówienia. Natomiast warto podkreślić, że dotychczas zostały przeprowadzone na zlecenie PARP cztery badania ewaluacyjne z ustaleniem efektu netto. W jednym z nich ustalono brak efektu w przypadku programu dotacji skierowanych do przedsiębiorstw, zaś w drugim – stosunkowo słaby, ale jednak występujący efekt szkoleń skierowanych do bezrobotnych. W przypadku wszystkich badań zanotowano wiele problemów realizacyjnych, na które metoda ta napotyka. Wielu z nich można by uniknąć, gdyby lepiej były zaplanowane systemy sprawozdawczości i monitoringu stosowane w administracji publicznej i systemie wdrażania programów.

Co robić, gdy cel jest niejasny?

Warunkiem tego, by analizować efekt przyczynowy, jest dobre ustalenie, jakiego efektu się oczekuje. Znajduje to swój wyraz w dobrze zdefiniowanych celach interwencji. W tym zakresie jest jeszcze wiele do zrobienia (pisaaliśmy o tym przy okazji omawiania funkcji wskaźników w ewaluacji w Górniak, Keler 2007 i 2008). Sytuacja ta dotyczy programów finansowanych ze środków europejskich, choć w tej dziedzinie nastąpiła poprawa w nowym okresie programowania. Ogromne wyzwanie stanowi jednak cały system

¹⁴ Chodzi o skłonność rozumianą jako prawdopodobieństwo znalezienia się w grupie objętej interwencją ze względu na posiadane cechy.

wydatków jednostek sektora finansów publicznych. Obecnie trwają prace nad wdrożeniem budżetu zadaniowego, co oznacza przejście od planowania i sprawozdawczości w kategoriach wydatków ujętych w pozycje statystyki budżetowej do obowiązku definiowania celów, ujmowania ich w planach, wiązania z wydatkami i sprawozdania z ich wykonania. Dotychczasowe, próbne warianty budżetu zadaniowego nie wyglądają dobrze pod tym względem¹⁵.

Czy fakt nieformułowania celów w klasycznym układzie budżetu i słabości pod tym względem w przypadku dotychczasowych opracowań budżetu zadaniowego oznacza, że środki publiczne wydawane są bez celów (= bez sensu z punktu widzenia podatnika)? Otóż nie. Są wydawane celowo, choć czy efektywnie, to inny problem. Nie ma też ocen skuteczności działań prowadzonych za te środki publiczne w sensie omówionej wyżej analizy przyczynowej (w innym, mniej ambitnym znaczeniu również ich brak). Po to, by analizy takie wprowadzić, trzeba jednak zacząć od zdefiniowania celów.

Pierwszym źródłem celów powinna być strategia rządu uzyskująca poparcie większości parlamentarnej (takiej jasnej strategii jest brak). Jednak ważnym sposobem uchwycenia faktycznych celów, które wynikają z potrzeb funkcjonowania państwa może być stosowna analiza.

W związku ze wskazanymi wyżej problemami występującymi w zakresie zdefiniowania celów programów wieloletnich i mierników ich efektów pierwsze ewaluacje obecnie działających programów należy wykonywać w formule otwartej – tzw. ewaluacji wolnej od celów (*goal-free evaluation*). Jest to określenie zaproponowane przez M. Scrivena, jednego z czołowych teoretyków ewaluacji. Nie oznacza ono uwolnienia programów od konieczności ustanowienia celów, lecz ukierunkowuje ewaluację na wszystkie efekty, które wystąpiły w efekcie realizacji programu. W metodyce ewaluacji przyjętej przez Komisję Europejską oznaczałoby to przypisanie szczególnej roli kryterium użyteczności w jego rozszerzającej definicji. Ewidencja wszystkich konsekwencji zarówno pożądanых, jak i niepożądanych, zawartych w przyjętych celach, jak i tych, które nie były w sposób udokumentowany oczekiwane, pozwala ustalić korzyści z programu, a odniesienie poszczególnych zewidencjonowanych efektów do ich kosztów daje nam miarę efektywności. Skuteczność (a także efektywność) w osiągnięciu założonych celów (o ile zostały zdefiniowane w sposób nadający się do ewaluacji) staje się szczególnym przypadkiem takiego szerokiego ujęcia. To podejście ma wartość polegającą na silniejszym otwarciu na zewidencjonowanie i pomiar wszystkich efektów oraz na tym, że dostarcza wiedzy o efektach, których można realistycznie oczekiwać i najlepszym ujęciu ich w kategoriach wskaźników w procesie planowania. Problemem do rozwiązania w tym podejściu jest wysoki poziom wymagań profesjonalnych stawianych ewaluatorom¹⁶. Taki rodzaj ewaluacji zakłada także większe zaangażowanie (partycypację) po stronie jednostek uczestniczących w różnych rolach w planowaniu i wdrożeniu programu, co jednak często traktowane jest jako zaleta w teorii ewaluacji.

Mając opracowaną mapę celów można przystąpić do ustalenia wskaźników, wykonania pomiaru efektów, a następnie zbadania efektów przyczynowych w przedyskutowany wyżej sposób.

¹⁵ Szczegółowa analiza budżetu zadaniowego pod względem sposobu formułowania celów i podatności na ewaluację została przeprowadzona pod kierunkiem autora w ekspertyzie przygotowanej na zlecenie Ministerstwa Finansów (por. Górniak, Jelonek, Mazur, Strzebońska, Szczucka 2008).

¹⁶ Trudno będzie m.in. zaspokoić tę potrzebę w warunkach obecnej sytuacji na rynku usług ewaluacyjnych w oparciu o stosowane obecnie tryby wylaniania wykonawców ewaluacji w ramach prawa zamówień publicznych, z naciskiem na kryterium ceny: jest to ogólnie biorąc tryb niedostosowany do wylaniania wykonawców przedstawiających rozwiązania problemów otwartych, takich jak prace naukowe, eksperckie, innowacyjne; o ile badania naukowe – przy kontrowersjach co do zaliczania w ich poczet rozmaitych przedsięwzięć badawczych – są zwolnione z postępowania w oparciu o PZZ, to inne, analogiczne przedsięwzięcia już nie, co powoduje negatywne skutki i nieefektywności w wydatkowaniu środków publicznych, których omówienie nie jest jednak przedmiotem niniejszego opracowania.

Zakończenie

Niniejszy artykuł jest zachętą do ambitnego projektu budowania bazy wiedzy w oparciu o ewaluację przyczynowego wpływu interwencji publicznych na zmiany zdefiniowane w celach tych interwencji. Czy to jest postulat realistyczny? Tak, choć w perspektywie ewolucyjnej. W ostatniej części tekstu została wskazana droga do odtworzenia celów z wykonywanych przez administrację publiczną działań. Jest to krok pierwszy, umożliwiający skupienie uwagi na efektach, które mają być osiągnięte. Projekty pilotażowe powinny pomóc wypracować założenia do konstrukcji tak funkcjonującego systemu informacji publicznej, który pozwalałby na tańsze pozyskiwanie informacji niezbędnych do przeprowadzenia ewaluacji. Wcześniej należy prowadzić jak najwięcej tego rodzaju badań w oparciu o środki przeznaczone na ewaluację programów unijnych. Należy też zadbać o stworzenie w kraju sieci dobrych, konkurencyjnych jednostek, które będą w stanie takie badania prowadzić, najlepiej w oparciu o czołowe wyższe uczelnie. W uczelniach będzie to miało efekt mnożnikowy przekładając się na kształcenie kadry i studentów. Do tej pory zbyt wiele środków wydaje się na mało ambitne ewaluacje, realizowane często pośpiesznie, prostymi środkami, które nie przyczyniają się znacząco do rozwoju jakości polityk publicznych w Polsce.

Jarosław Górniak, dr hab., prof. Uniwersytetu Jagiellońskiego i Wyższej Szkoły Europejskiej im. ks. J. Tichnera, twórca i kierownik Centrum Ewaluacji i Analiz Polityk Publicznych UJ, kierownik Zakładu Socjologii Gospodarki, Edukacji i Metod Badań Społecznych Instytutu Socjologii UJ, prodziekan Wydziału Filozoficznego UJ. Socjolog i ekonomista, specjalista w zakresie analizy polityk publicznych i ewaluacji, socjologii gospodarki i organizacji i metod badań społecznych. Ekspert w zakresie statystycznej analizy danych. Doświadczony badacz rynku i opinii, przed dziesięć laty przesunął ciężar swojej aktywności naukowo-badawczej na zagadnienia analizy polityk publicznych i ewaluacji. Był m.in. konsultantem międzynarodowym UNDP, ekspertem Zespołu koordynacyjnego ds. ewaluacji ex-ante dokumentów unijnych na lata 2007–2013 w Ministerstwie Rozwoju Regionalnego. Kierował projektami badawczymi dotyczącymi ewaluacji i wskaźnikowego pomiaru efektów programów publicznych (m.in. w zakresie rynku pracy, wykluczenia społecznego, rozwoju kapitału ludzkiego i rozwoju regionalnego). Konsultant w zakresie programowania strategicznego i ewaluacji, a także w zakresie metodologii badań i analizy danych w zarządzaniu i administracji publicznej. Członek Rady CBOS. Członek komitetu sterującego projektu METRIS (Monitoring European Trends in Social Sciences and Humanities) przy DG Research Komisji Europejskiej.

Bibliografia

- Bridgman P., Davis G. (2004), *The Australian policy handbook*, 3rd edn, Allen & Unwin, Crows Nest, N.S.W.
- Ferry M. & Olejniczak, K. (2008), *Wykorzystanie ewaluacji w zarządzaniu programami unijnymi w Polsce*, Warszawa: Ernst & Young, Program „Sprawne Państwo”: www.sprawnepanstwo.pl.
- Górniak J., Keler K. (2007), *Wskaźniki w ewaluacji ex-post programów publicznych*, [w:] *Ewaluacja ex-post w teorii i praktyce*, PARP, Warszawa; Górniak J., Keler K. (2008), *Ewaluacja i wskaźniki w cyklu planowania i realizacji polityk publicznych*, „Zarządzanie Publiczne” Nr. 1(3)/2008.
- Górniak J., Keler K. (2008), *Rola systemu wskaźników w ewaluacji*, [w:] K. Olejniczak, M. Kozak, B. Lendzion (red.), *Teoria i praktyka ewaluacji interwencji publicznych*, Wydawnictwa Akademickie i Profesjonalne, Warszawa.

- Heckman J.J. (2005), *The Scientific Model of Causality*, Sociological Methodology, Vol. 35, No. 1. (August 2005).
- Heckman, J.J., Smith, J.A. (1995), *Assessing the Case for Social Experiments*, Journal of Economic Perspectives, Vol. 9, Issue 2, Spring 2005.
- Komisja Europejska, *Ewaluacja programów wydatków Unii Europejskiej*. Przewodnik b.d..
- Konarski R., Kotnarowski M. (2007), *Zastosowanie metody propensity score matching w ewaluacji ex-post*, [w:] Haber A. (Red.), *Ewaluacja ex-post. Teoria i praktyka badawcza*, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa.
- Korporowicz L. (2008), *Cele ewaluacji ex ante. Myślenie o rozwoju*, [w:] *Ewaluacja ex ante – podsumowanie doświadczeń administracji polskiej*, Ministerstwo Rozwoju Regionalnego, Warszawa.
- OECD, *Performance Budgeting in OECD Countries* (2007), OECD.
- Olejniczak K. (2008), *Wprowadzenie do zagadnień ewaluacji*, [w:] K. Olejniczak, M. Kozak, B. Ledzion (red. nauk.), *Teoria i praktyka ewaluacji interwencji publicznych*. Podręcznik akademicki, Wydawnictwa Pedagogiczne i Naukowe, Warszawa.
- Olejniczak K. (2008), *Ewaluacja jako proces badawczy*, [w:] K. Olejniczak, M. Kozak, B. Ledzion (red. nauk.), *Teoria i praktyka ewaluacji interwencji publicznych*. Podręcznik akademicki, Wydawnictwa Pedagogiczne i Naukowe, Warszawa.
- Patton M.Q. (2007), *Utilization-Focused Evaluation*, 3rd edn, Sage, London.
- Pawson R., Tilley N. (1997), *Realistic Evaluation*, Sage, London, 1997.
- Sabatier P.A., Mazmanian D.A. (1981), *The Implementation of Public Policy: A Framework of Analysis*, [w:] D.A. Mazmanian, P.A. Sabatier (eds.), *Effective Policy Implementation*, Toronto: Heath, Lexington.
- Strawiński (2008), *Quasi-eksperymentalne metody ewaluacji*, [w:] *Środowisko i warsztat ewaluacji*, Haber A., Szałaj M. (red.), Polska Agencja Rozwoju Przedsiębiorczości, Warszawa.
- Winship Ch., Sobel M. (2004), *Causal Inference in Sociological Studies*, [w:] Hardy, M., Bryman, A., *Handbook of data analysis*, Sage, London-Thousand Oaks-New Dehli.

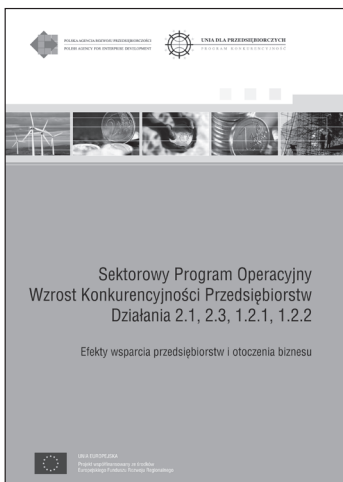
W ramach serii EWALUACJA ukazały się następujące publikacje:



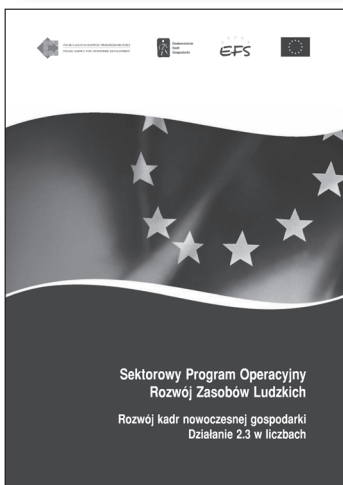
www.parp.gov.pl/index/more/2046



www.parp.gov.pl/index/more/9658



www.parp.gov.pl/index/more/5475



www.parp.gov.pl/index/more/5474



www.parp.gov.pl/index/more/9850

W najbliższym czasie ukażą się:

