

FRANCISZEK SZTABIŃSKI
ORBS IFiS PAN
TERESA ŻMIJEWSKA-JĘDRZEJCZYK
ORBS IFiS PAN/MillwardBrown SMG/KRC

MIXED MODE SURVEY DESIGN: PROBLEM EFEKTU TECHNIKI

Streszczenie

W sytuacji pogarszających się warunków realizacji badań surveyowych poszukuje się sposobów ograniczenia błędu pomiaru (*Total Survey Error*), związanego z nieuczestniczeniem wylosowanych jednostek w wywiadzie (*Non Response Error*). Jedno z rozwiązań stanowi procedura *Mixed Mode Survey Design*, która polega na uzyskiwaniu tych samych informacji od różnych osób, przy wykorzystaniu odmiennych technik badawczych, np. wywiad PAPI (*Paper and Pencil Interviewing*), CATI (*Computer Assisted Telephone Interviewing*), CAWI (*Computer Assisted Web Interviewing*). Jednakże wykorzystanie tej procedury może powodować wystąpienie efektu techniki (tzw. *Mode Effects*). W artykule analizujemy efekt techniki na płaszczyźnie: (i) sformułowania pytania; (ii) sposobu kontaktowania oraz (iii) sposobu komunikowania się z respondentem, poddając ocenie wskaźnik realizacji (tzw. *Response Rate*) oraz porównując odpowiedzi respondentów na te same pytania, uzyskane przy zastosowaniu różnych technik.

W analizach wykorzystano materiały zebrane w projektach realizowanych techniką PAPI (badanie European Social Survey, round 3) oraz CATI (badanie w ramach metodologicznego grantu ESS Infrastructure: i3.). Wyniki przeprowadzonych analiz potwierdzają istnienie poważnych różnic związanych z efektem techniki (w przypadku wywiadu PAPI i CATI). Stawia to pod znakiem zapytania ideę ich łączenia celem ograniczenia błędu braku odpowiedzi.

Słowa kluczowe: *Total Survey Error*, błąd braku odpowiedzi, efekt techniki, *Mixed Mode Survey Design*, Europejski Sondaż Społeczny

* fsztabin@ifispan.waw.pl ** tzmijews@ifispan.waw.pl

UWAGI WSTĘPNE

Z założenia badania surveyowe uznawane są za pomiar, a tym samym mają umożliwiać ściśle oszacowanie liczbowe badanych zjawisk i uogólnianie wniosków na całą populację. Aby było to możliwe, muszą zostać spełnione trzy podstawowe warunki: masowość (chodzi o liczebność próby do badania), losowy charakter próby oraz standaryzacja narzędzia, czynności badawczych i warunków realizacji badania. Każdy z tych elementów procesu badawczego jest jednak obarczony ryzykiem wystąpienia błędu. Powszechnie przyjętym obecnie sposobem oceny błędu w badaniach surveyowych jest koncepcja *Total Survey Error* (TSE) [Weisberg 2005; de Leeuw, Hox, Dillman 2008a: 1–17; Groves 1991]¹. W klasycznym ujęciu, TSE obejmuje cztery rodzaje błędów: (i) błąd próby, (ii) pokrycia, (iii) błąd braku odpowiedzi oraz (iv) pomiaru. Błąd próby (tzw. *Sampling Error*) jest efektem tego, że badanie jest realizowane wśród wylosowanych osób, a więc na części, nie zaś na całej zbiorowości. Błąd pokrycia (tzw. *Coverage Error*) polega na tym, że operat losowania nie obejmuje wszystkich jednostek, wśród których chcemy zrealizować badanie, jest niekompletny bądź występują w nim „powtórzenia” (niektóre jednostki występują w nim więcej niż jeden raz). Z kolei błąd braku odpowiedzi (tzw. *Non-Response Error*) obejmuje dwie kategorie [Groves 1989; Grzeszkiewicz-Radulska 2009: 21–36]. Pierwsza z nich to nieuczestniczenie w badaniu wylosowanych jednostek, przede wszystkim ze względu na brak możliwości kontaktu z nimi oraz odmowy uczestnictwa w badaniu (tzw. *Unit Nonresponse*). Drugi rodzaj błędu braku odpowiedzi to brak odpowiedzi respondenta na dane, konkretne pytanie (tzw. *Item Nonresponse*). Ostatni rodzaj błędu to 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 respondenta².

¹ Koncepcja ta korzysta bezpośrednio z zaproponowanej przez Roberta Grovesa [1989] klasyfikacji błędów. Autor ten mówi o dwóch podstawowych rodzajach błędów: niezwiązanych z obserwacją oraz związanych z nią. Wśród tych pierwszych wyróżnia błąd: losowania próby, „pokrycia” oraz „braku odpowiedzi” (na poziomie całego wywiadu i poszczególnych pytań). Do drugiej kategorii błędów Groves zalicza błąd związany ze sposobem uzyskiwania danych (techniką); z pytaniem (określany także błędem narzędzia) lub z osobą ankietera oraz osobą respondenta.

² Rozwinięcie klasycznej koncepcji TSE przedstawiła m.in. Edith de Leeuw [de Leeuw et al. 2008a], wprowadzając piąty element – błąd specyfikacji (tzw. *specification error*) występujący w sytuacji, gdy pytanie mierzy różny od założonego koncept, np. gdy obserwujemy niską trafność konstrukcyjną (tzw. *construct validity*). Nie wspomina ona jednak o błędach opracowania (tzw. *processing errors*), a więc błędach: kodowania, wprowadzania, danych, edycji danych oraz

Obecnie przyjmuje się, że błąd braku odpowiedzi na poziomie całego wywiadu jest jednym z najpowszechniejszych i to właśnie on decyduje o rozmiarze TSE. Praktycznie od połowy lat osiemdziesiątych obserwuje się ciągły spadek tzw. odsetka realizacji (*Response Rate*)³. Tendencja ta występuje na całym świecie, choć jej skala jest różna w zależności od techniki i tematyki badania [Goyder 1987; de Leeuw, de Heer 2002: 41–54; Groves et al. 2004: 185–187; Stoop et al. 2010]. W Holandii odsetek realizacji w badaniach czytelnictwa spadł poniżej 50% w 1998 r., a w 2000 r. nie osiągnął nawet 40% [Bassler A., Tchaoussoglou C. 2001]. W Polsce, w projekcie Polski Generalny Sondaż Społeczny [Cichomski et al. 2009] odsetek realizacji zmniejszył się na przestrzeni ostatnich 16 lat (1992–2008) o 30 punktów procentowych (z 82 do 52%), co oznacza, że w 2008 r. nie zrealizowano 48% wylosowanej próby. Edith de Leeuw i Wim de Heer [de Leeuw, de Heer 2002], analizując badania realizowane w latach 1980–1989 w 16 krajach (w tym Stany Zjednoczone i Kanada), sformułowali prognozę, iż średni wzrost odsetka braku kontaktu (*Non-Contacts*) będzie postępował w tempie 0,2% na rok, zaś wskaźnik odmów (*Refusals*) – w tempie 0,3%. Nie należy się dziwić, że w 2005 r. ukazała się książka Ineke Stoop zatytułowana *The Hunt for the Last Respondent: Nonresponse in Sample Surveys* [Stoop 2005].

O wadze błędu braku odpowiedzi na poziomie wywiadu świadczy także to, że nie ma on charakteru losowego, jest błędem systematycznym i powoduje tendencyjność wyników: osoby niebiorące udziału w badaniu różnią się w sposób istotny – ze względu na cechy społeczno-demograficzne, psychologiczne oraz opinie – od osób, które w nim uczestniczyły [Sztabiński P.B., Sztabiński F., Przybysz 2007: 26–54].

Czy zatem w sytuacji narastającego problemu błędu braku odpowiedzi (*Non-Response Error*) istnieją sposoby jego ograniczenia? Zasadniczo tak, aczkolwiek winno się raczej mówić o sposobach „zapobiegania” powstawaniu tego błędu (*ex ante*) oraz sposobach jego „naprawiania” (*ex post*), czyli o dwu różnych, komplementarnych podejściach. W ramach pierwszego z nich wymienić należy wszystkie działania z zakresu Quality Control lub Quality Assurance, mające na celu optymalizację przebiegu procesu badawczego z punktu widzenia jakości uzyskiwanych danych (chodzi o właściwe przygotowanie narzędzia badawczego, szkolenie i monitoring ankietatorów itd. [Sztabiński F. 2011: 19–35]). W ramach drugiego podejścia wskazuje się przede wszystkim ważenie poststratyfikacyjne

uzupełnienia braków danych (tzw. *imputation errors*). [Biemer, Caspar 1994: 307–326; Weisberg 2005].

³ Mówimy tutaj oczywiście o badaniach, które są realizowane na próbach bez zastępowania (tzn. realizowanych „do wyczerpania”).

(uwzględniające rozkład cech respondentów w próbie populacji) oraz ważenie w oparciu o tzw. wskaźnik dostępności respondenta (*propensity score*), który określa prawdopodobieństwo udziału w badaniu w zależności od jego cech, nie tylko demograficznych [Stoop et al. 2010].

Zupełnie inne od wskazanych powyżej, aczkolwiek mieszczące się w ramach procedur *ex ante*, jest tzw. *Mixed Mode Survey Design*, a więc wykorzystanie różnych technik zbierania danych w ramach jednego projektu badawczego. Wiadomo, że właściwy dobór techniki badania może przynajmniej w pewnym zakresie ograniczyć błąd braku odpowiedzi. Chodzi tutaj o problem tzw. adekwatności techniki badania [Gostkowski 1965, 1966; Słomczyński 1968; Sztabiński F. 1997, 2003: 147–175]. W uproszczeniu sprowadza się on do tego, że różni ludzie preferują różne sposoby kontaktowania się i komunikowania. Dodatkowo, niezależnie od preferencji, nie ze wszystkimi można się skontaktować w dany sposób (np. młodzi ludzie kształcący się „poza domem” są trudniej dostępni pod adresem zameldowania, reprezentanci określonych grup zawodowych – głównie profesjonaliści i drobni przedsiębiorcy – są łatwiej dostępni „pod telefonem”).

Jeśli błąd braku odpowiedzi można ograniczyć, łącząc różne techniki zbierania danych w ramach jednego badania, to rodzi się pytanie, w jakim stopniu taki „cocktail” technik może jednocześnie zwiększać błąd pomiaru wynikający z efektu techniki (*Mode Effects*). Jest rzeczą oczywistą, że poszczególne techniki surveyowe mają sobie właściwe nie tylko zalety, ale i ograniczenia. Wiadomo, iż te same osoby, w zależności od formy kontaktu (osobisty lub pośredni, w tym telefoniczny, przez Internet itp.) oraz kanału komunikowania (komunikacja ustna lub pisemna) mogą w różny sposób reagować na ten sam bodziec, a więc udzielać różnych odpowiedzi na to samo pytanie⁴. Skoro tak, czy można je kwantyfikować, a następnie uogólniać na całą populację? Raczej nie, co przemawia za tym, aby nie „mieszać” technik. Na badanie należy jednak spojrzeć całościowo, z punktu widzenia „strat” i „korzyści”. *Przede wszystkim trzeba ocenić, czy wprowadzenie różnych technik (tzw. Mixed Mode) zmniejszy błąd braku odpowiedzi (Nonresponse Error), czy też zwiększy błąd pomiaru (Measurement Error) przy danych kosztach. Jest to swoisty trade-off: czy lepiej nie „mieszać” technik i uzyskać odpowiedzi od połowy lub mniejszego odsetka wylosowanych osób, czy też – mając świadomość błędu pomiaru związanego z efektem techniki – lepiej je łączyć i uzyskać odsetek realizacji na poziomie 70% lub powyżej⁵.*

⁴ Wiąże się to przede wszystkim z dystansem „sytuacyjnym” i „czasowym”, charakterystycznym dla komunikacji pisemnej [Sztabiński F. 1997: 41–47].

⁵ Odsetek taki stanowi minimum w badaniach europejskich [European Social Survey 2011].

W niniejszym artykule podejmujemy problem efektu techniki (*Mode Effects*) i jej wpływu na odpowiedzi respondentów, a tym samym rozmiar błędu pomiaru w ramach TSE. Ponieważ efekt techniki oznacza odmiennosc: (i) sformułowania pytania, (ii) sposobu kontaktowania oraz (iii) sposobu komunikowania się z respondentem, przedstawimy przede wszystkim zakres i rozmiar owych różnic. W analizach wykorzystane zostaną materiały zebrane w projekcie zrealizowanym techniką PAPI – ESS r. 3 oraz techniką CATI – ESS r. 3 zastosowaną w projekcie metodologicznym ESS Infrastructure: i3. Najpierw jednak doprecyzujemy termin *Mixed Mode Survey Design* oraz przedstawimy ogólne wnioski z badań nad różnymi technikami.

MIXED MODE SURVEY DESIGN

Mixed Mode Survey Design (*Mixed Mode Data Collecting*, *Mixed Mode Survey*, *Mixed Mode Design* itp., w skrócie *Mixed Mode*) jest procedurą badania, w której dopuszcza się łączenie nie tylko różnych technik wywiadu (np. PAPI i CATI), lecz także technik wywiadu z technikami ankiety (np. CATI, CAWI lub Ankieta Pocztaowa) [de Leeuw et al., 2008b: 299–316]. Z jednej strony, idea łączenia technik nie jest niczym nowym (np. zastosowanie ankiety wypełnianej pod kontrolą ankietera w trakcie lub po wywiadzie). Z drugiej jednak jest ona zupełnie nowa, ponieważ zakłada uzyskiwanie danych – w ramach tego samego badania – różnymi technikami.

Czy jednak w obu przypadkach mówimy o tym samym? Raczej nie. W związku z tym celowe wydaje się doprecyzowanie terminu *Mixed Mode*. O zróżnicowaniu stosowanych technik można mówić na etapie kontaktowania się z respondentem oraz zbierania danych. W pierwszym przypadku wykorzystujemy różne sposoby, kanały komunikowania się (np. pisemny – listy zapowiednie, ustny – telefony do respondentów). Nie nazwiemy tego *Mixed Mode*. W drugim przypadku (etap zbierania danych) w wielu projektach przyjęty schemat badania zakłada wykorzystanie różnych technik zbierania danych. Jak wspomniano, w trakcie lub po wywiadzie (PAPI/CAPI) możemy wręczyć respondentowi ankietę do samodzielnego wypełnienia⁶. Tego badania również nie nazwiemy *Mixed Mode*, bowiem w przypadku każdego respondenta stosujemy ten sam schemat, a zatem możemy co najwyżej mówić o technice zasadniczej i pomocniczej (dodatkowej).

⁶ Taki schemat badania stosuje się w Polskim Generalnym Sondażu Społecznym (PGSS). Po zrealizowaniu wywiadu ankieter wręcza respondentowi ankietę ISSP do samodzielnego wypełnienia (w obecności ankietera).

Don Dillman [2000: 219–222] mówi o pięciu sytuacjach, które można określić mianem *Mixed Mode* [także: de Leeuw, Hox, Dillman, 2008b: 299–316]. Są one następujące:

– zbieranie tych samych informacji, od różnych osób przy zastosowaniu różnych technik; jest to najpowszechniej stosowana procedura i podstawowym jej celem jest ograniczenie błędu braku odpowiedzi, przy danych lub zredukowanych kosztach;

– zbieranie danych „panelowych” różnymi technikami od tych samych osób w kolejnych falach badania;

– zbieranie różnych informacji różnymi technikami od tych samych osób wylosowanych w trakcie „jednorazowego” badania (wspomniana sytuacja, kiedy w trakcie lub po wywiadzie zasadniczym pozostawia się respondentowi ankietę do wypełnienia);

– zbieranie tych samych informacji różnymi technikami w różnych populacjach (w projektach międzykrajowych, w poszczególnych krajach badanie może być realizowane różnymi technikami. Europejski Sondaż Społeczny w niektórych krajach realizuje się techniką PAPI, a w innych CAPI [European Social Survey 2011].

– wykorzystanie różnych technik „wspomagających” realizację badania prowadzonego inną techniką – mogą to być listy zapowiednie w przypadku badań *face-to-face* czy monity telefoniczne w przypadku badań realizowanych techniką ankiety pocztowej.

Przedstawione przez Dillmana rozumienie *Mixed Mode* należy uznać za zbyt szerokie, ponieważ większość badań należałoby zaliczyć do tej kategorii. Jeśli przed badaniem realizowanym techniką PAPI wysyłamy do respondentów listy zapowiedni, to należałoby uznać, że jest to badanie *Mixed Mode*⁷. Podobne, węższe rozumienie tego terminu przyjmują Edith de Leeuw [2009], Peter Lynn [2009] oraz Peter Martin [2010], mając na myśli wyłącznie pierwszą z wyróżnionych sytuacji, a więc zbieranie tych samych informacji przy zastosowaniu różnych technik od różnych jednostek stanowiących elementy danej próby. W pozostałych przypadkach należałoby mówić o *Multi(ple) Mode Design*.

Jeśli akceptujemy ideę *Mixed Mode*, rozumianą we wskazany sposób, musimy odpowiedzieć sobie na dwa podstawowe pytania: w jaki sposób łączyć techniki i jakie („How to Mix?”, „What to Mix?”). Zasadniczo istnieją *dwa sposoby implementacji Mixed Mode* [de Leeuw, Hox, Dillman 2008a: 306–307, 310–312; Martin

⁷ Warto tutaj zaznaczyć, iż w późniejszej pracy autor [Dillman et al. 2009] mówi już tylko o czterech typach *Mixed Mode* (pomija typ iv.).

2010]. Jeden z nich to sekwencyjne stosowanie różnych technik (tzw. *sequential*), które polega na tym, że określamy jedną podstawową (preferowaną) technikę oraz dodatkowe, uzupełniające, które będą stosowane w dalszej kolejności. Przykładowo, jeśli podstawową techniką jest CATI, to do grupy respondentów, z którymi nie udało się zrealizować wywiadu telefonicznie, możemy wysłać ankieterów, aby zrealizowali z nimi wywiad *face-to-face* (techniką PAPI lub CAPI). Celem sekwencyjnego *Mixed Mode* jest oczywiście redukcja błędu braku odpowiedzi (*unit nonresponse*), ale także kosztów badania. Zazwyczaj jako technikę podstawową wybiera się najtańszą, np. ankietę Web-ową lub tradycyjną ankietę pocztową. Kolejne techniki są relatywnie droższe w zastosowaniu (np. wywiad CATI, CAPI czy PAPI). Drugi rodzaj *Mixed Mode* polega na równoczesnym (tzw. *concurrent*) stosowaniu różnych technik zbierania informacji. Nie określa się wówczas techniki podstawowej, lecz traktuje wszystkie jak ekwiwalentne i wykorzystuje się je równocześnie, w tym samym czasie. Możemy wysłać do wylosowanych osób list z informacją o badaniu i propozycją wypełnienia ankiety „papierowej” (ankieta pocztowa), e-mailowej lub Web-owej. W zależności od preferencji wysyłamy respondentowi właściwą wersję ankiety lub też proponujemy rozmowę telefoniczną lub osobistą. Celem równoczesnego *Mixed Mode Design* jest, poza redukcją błędu braku odpowiedzi, przede wszystkim redukcja błędu pokrycia (np. osoby niemające dostępu do Internetu mogą wypełnić ankietę „papierową”), ale także minimalizacja kosztów.

Odpowiedź na pytanie o sposób administrowania poszczególnych technik (w sposób sekwencyjny czy równoczesny) zależy przede wszystkim od tego, jakiego rodzaju błąd chcemy minimalizować przy danych kosztach. Zgodnie z koncepcją *Total Design Method* (TDM) Dillmana [1978; 2000], opartą na teorii wymiany społecznej, model równoczesny *Mixed Mode* winien być optymalny i bardziej efektywny niż sekwencyjny. Daje respondentowi poczucie podmiotowości, „możliwość wyboru formy kontaktu”, stanowi wyraz „dyspozycyjności” badacza wobec respondenta, przez co „dowartościowuje” go, a tym samym stanowi element szeroko rozumianego systemu nagród. Jednocześnie model równoczesny eliminuje negatywne implikacje związane z poczuciem „konieczności subordynacji”, minimalizując tym samym ponoszone przez respondenta „koszty”. Praktyka wskazuje, że model sekwencyjny jest bardziej efektywny, przynajmniej w zakresie kosztów⁸.

⁸ Jaki więc model stosować? Wydaje się, iż zależy to przede wszystkim od badanej populacji. W badaniach *Mixed Mode* na populacji generalnej prawdopodobnie lepiej sprawdza się schemat sekwencyjny, natomiast w badaniach specyficznych kategorii respondentów chyba schemat równoczesny.

Właściwe stosowanie *Mixed Mode* z pewnością może redukować błąd braku odpowiedzi (na poziomie wywiadu), błąd pokrycia oraz koszty badania, komplikuje jednak schemat i sposób doboru próby (co może wpływać na wielkość błędu próby) oraz stwarza trudności, od strony logistycznej, w realizacji terenowej badania. Jeśli chodzi o błąd pomiaru, to zawsze należy liczyć się z efektem techniki (*Mode Effects*). I wreszcie koszty: zależą od tego, „co mieszamy i w jaki sposób”. W porównaniu z wywiadem PAPI (traktowanym jako technika podstawowa, *Single Mode*) zastosowanie sekwencyjne lub równoczesne *Mixed Mode*, w jakiegokolwiek konfiguracji, obniża koszty badania. Natomiast zastosowanie schematu sekwencyjnego, niezależnie od techniki podstawowej, podnosi koszty badania (rozpoczyna się od techniki „najtańszej”). Podsumowując, *Mixed Mode* odpowiednio stosowany powinien prowadzić do redukcji TSE, a w wielu przypadkach także kosztów.

EFEKT TECHNIKI (*MODE EFFECTS*)

Ogólnie biorąc, efekt techniki polega na tym, że zastosowanie różnych sposobów zbierania materiałów (technik badawczych) może pociągać za sobą ryzyko uzyskania odmiennych, różnych danych. Ponieważ wcześniej wspominaliśmy, iż wykorzystanie *Mixed Mode* stanowi zawsze swoisty *trade-off* między redukcją błędu braku odpowiedzi a efektem techniki, warto w tym miejscu przypomnieć podstawowe charakterystyki niektórych technik badawczych, aby mieć świadomość, „co zyskujemy, a co tracimy”. Ograniczymy się do podstawowych technik *opartych na komunikacji ustnej wspomaganych lub niewspomaganych komputerowo* (CAPI, CATI, PAPI) oraz *technik opartych na komunikacji pisemnej, także wspomaganych lub niewspomaganych komputerowo* (CASI – *Computer Assisted Self Interviewing* oraz PSAQ – *Paper Self Administered Questionnaires*) [Saris: 1991]⁹.

⁹ Ogólne rozważania na temat zalet i ograniczeń poszczególnych technik badawczych prowadzone były niemal „od zawsze”. W 1950 r. opublikowana została jedna z pierwszych prac poświęconych technikom surveyowym, oparta na ponad dwudziestoletnich doświadczeniach w dziedzinie badań tego typu [Parten, 1950]. Refleksja ta kontynuowana jest do dzisiaj, a problem określenia zalet i ograniczeń poszczególnych procedur badawczych rodzi się zawsze w związku z pojawieniem się nowych technik, np. CAPI, CATI czy CAWI [Bowling 2005: 281–291; de Leeuw et al. 2008a]. Na gruncie polskim tego typu rozważania prowadzone były m.in. przez J. Lutyńskiego [1968; 1972], F. Sztabińskiego [1997; 2009], P.B. Sztabińskiego [1999a–c; 2001: 65–89], F. Sztabińskiego i P.B. Sztabińskiego [1997: 73–92], T. Żmijewską-Jędrzejczyk [2004].

Zacznijmy od wywiadu kwestionariuszowego. Wybierając wywiad PAPI, który ma bardzo szeroki zakres zastosowania, mamy pewność, że w badaniu uczestniczy właściwa osoba, zapewniona jest standaryzacja sytuacji badania (chodzi o miejsce i warunki wywiadu). Dodatkowo, w przypadku pytań trudnych, niejasnych, istnieje możliwość pomocy i dodatkowych wyjaśnień ze strony ankietera (tym samym minimalizuje się ryzyko zbierania artefaktów). Zakłada się też, iż dzięki obecności ankietera zapewniona zostanie poprawność wypełnienia kwestionariusza. Jednocześnie ta sama technika ma określone ograniczenia. Przede wszystkim istnieje ryzyko wpływu ankietera na odpowiedzi respondenta [Lutyńska 1978, Sztabiński P.B. 1997] oraz, ze względu na ograniczone poczucie anonimowości badanych, ryzyko udzielania społecznie akceptowanych odpowiedzi. Jest to technika bardzo kosztowna, a długość okresu realizacji badania jest niemal wprost proporcjonalna do wielkości próby [Sztabiński F. 1997: 48–52].

Wywiad osobisty ze wspomaganiami komputerowym (CAPI), oprócz wspomnianych zalet wywiadu PAPI, ma jeszcze kilka dodatkowych. Przede wszystkim zastosowanie komputera stwarza możliwości wykorzystania stymulacyjnych materiałów wizualnych (np. przy testowaniu reklam), co w konsekwencji osłabia efekt znużenia respondenta. Umożliwia rotację kafeterii zawartej w pytaniu czy nawet całych bloków pytań, co niweluje efekt porządku. Technika CAPI daje ankieterowi poczucie profesjonalizmu i w znacznym stopniu standaryzuje jego zachowania poprzez ograniczenie błędów w prowadzeniu wywiadu, a także rozszerza zakres możliwej kontroli pracy ankieterów (chodzi m.in. o automatyczny zapis tzw. „międzyczasów” dla poszczególnych sekcji wywiadu). Co więcej, wykorzystanie komputera znacząco wpływa na obniżenie kosztów realizacji badania (eliminacja kosztów związanych z drukiem kwestionariuszy, zdjęć itd.) oraz skrócenie okresu realizacji badania (pominięcie fazy transportu wypełnionych kwestionariuszy ankiet i wprowadzania danych). Niemniej technika ta, obok zalet, ma także pewne ograniczenia. Oprócz analogicznych, jak dla wywiadu PAPI, istnieją jeszcze dwa, z których pierwsze może rzutować na wartość uzyskiwanych danych. Chodzi o problem redukcji wymiaru społecznego interakcji (koncentracja ankieterów przede wszystkim na komputerze i monitorze, nie zaś na interlokutorze) oraz poczucie respondentów zupełnego „prześwietlenia” i depriwatywacji (mówi się nawet o efekcie tzw. *Big Brother*) [Sztabiński P.B. 2001: 65–89]. Drugie ograniczenie ma charakter techniczny. Chodzi o takie sprawy, jak np. problemy z uruchomieniem czy podłączeniem komputera, błędy w skrypcie. Nie sposób nie wspomnieć także o poczuciu zagrożenia bezpieczeństwa osobistego ankieterów.

Kolejna technika – wywiad CATI – ma wiele analogicznych zalet jak wywiady PAPI i CAPI: bardzo szeroki zakres zastosowania, pewność, że w badaniu uczestniczy właściwa osoba, standaryzacja sytuacji badania (miejsca, warunków wywiadu, ograniczony zostaje czynny wpływ osób trzecich) oraz standaryzacja zachowań ankieterskich (poprzez bieżący monitoring zostaje ograniczone prawdopodobieństwo wystąpienia błędów ankieterskich, niezwiązanych z zadawaniem pytań, te „kontroluje” program). Istnieje także możliwość pomocy ze strony ankietera w przypadku pytań trudnych, niejasnych, a ze względu na większe poczucie anonimowości występuje słabsza (w porównaniu z innymi technikami) skłonność do udzielania społecznie akceptowanych odpowiedzi. Ze względu na charakter rozmowy telefonicznej (szybkie tempo) uzyskuje się „szybkie”, skrótowe i schematyczne, a w konsekwencji mało przemyślane i niepogłębione odpowiedzi (np. respondent przerywa ankieterowi odczytywanie kafeterii, gdy usłyszy pierwszą „satysfakcjonującą” go odpowiedź). Dodatkowo, ze względu na specyfikę kontaktu, występuje silniejsza, w porównaniu z wywiadem PAPI/CAPI, skłonność do „zgadzania się” (tzw. „potakiwania”) oraz wyboru skrajnych punktów skali, a przy skalach liczbowych 100-punktowych – częstszy wybór „pełnych” wartości. Wreszcie jedno z podstawowych ograniczeń wywiadu CATI: nie mamy możliwości wykorzystywania wizualnych materiałów stymulujących (przede wszystkim Kart Respondenta), aczkolwiek istnieje możliwość odtworzenia materiału dźwiękowego. Sposób kontaktowania się narzuca także pewne ograniczenia czasowe: wywiad nie może być zbyt długi, gdyż istnieje ryzyko przerwania rozmowy przez respondenta). W rezultacie techniki oparte na tym samym kanale komunikowania (komunikacja ustna), ale odmiennym sposobie kontaktowania (kontakt naoczny vs kontakt telefoniczny) mogą przynosić zupełnie różne wyniki. Przykładowo, dla pytań otwartych o znajomość firm ubezpieczeniowych różnice rozkładów odpowiedzi w badaniu CATI i PAPI sięgają 8–9 punktów procentowych (odsetki są niższe w CATI). Z kolei w przypadku pytań zamkniętych różnice rozkładów sięgają 20–30 punktów procentowych. W tym przypadku odsetek „znajomości” firm jest wyższy w CATI [Sztabiński P.B. 1999a]. Technika ta zapewnia także krótki czas realizacji badania i jest najtańszą techniką opartą na komunikacji ustnej. Jednocześnie stwarza ryzyko wypaczenia próby ze względu na sposób konstruowania próby do badania. Chodzi tutaj przede wszystkim o błąd pokrycia, wynikający z nierównomiernego wyposażenia gospodarstw domowych w telefony¹⁰.

¹⁰ Obecnie, w związku ze wzrostem wyposażenia gospodarstw domowych w telefony komórkowe, obniża się poziom wyposażenia w telefony stacjonarne. Ponieważ do tej pory nie wypracowano

Jakie są zalety i ograniczenia ankiet samodzielnie wypełnianych przez respondentów w porównaniu z różnymi formami wywiadu? Pierwszą i podstawową jest to, że uzyskany materiał badawczy, w przeciwieństwie do technik wywiadu PAPI, CAPI i CATI, jest tożsamy ze źródłem informacji. Chodzi o to, że – ze względu na brak pośrednictwa ankietera – informacje, których udzielił respondent (sam zapisał) są identyczne z tymi, których chciał udzielić (w wywiadzie są „przefiltrowane” przez ankietera). Inne zalety ankiet to bardzo szeroki zakres zastosowania, poczucie anonimowości badanych oraz niski koszt realizacji. Jednakże, tak jak w przypadku innych technik, ankiety samodzielnie wypełniane przez respondenta mają pewne ograniczenia. Należy tutaj wymienić choćby to, iż nigdy nie ma pewności, kto faktycznie wypełnia ankietę (poza przypadkami ankiety pod kontrolą oraz audytoryjnej). Co więcej, respondenci zazwyczaj zapoznają się z treścią pytań ankiety przed jej wypełnieniem i mogą konsultować swoje odpowiedzi z innymi osobami (tzw. konsultacja „wewnątrzdomowa”), a w efekcie ich odpowiedzi są zakotwiczone społecznie, nie są zaś własne, indywidualne¹¹. Inne ograniczenia sprowadzają się do znacznej pracochłonności jej przygotowania, dość długiego okresu realizacji badania (z monitami ok. 5–7 tygodni, przy czym długość okresu realizacji nie jest praktycznie zależna od wielkości próby), nieco niższego odsetka realizacji w porównaniu z wywiadem [Dillman 1978; Sztabiński F. 1997] oraz tego, że część ankiet wraca niewypełniona lub wypełniona niekompletnie. W efekcie wszystkich tych odmienności odpowiedzi uzyskiwane na te same pytania przynoszą inne rezultaty niż wywiad [Sztabiński F. 1997]. Dla pytań o opinie różnice rozkładów w ankiecie pocztowej i wywiadzie PAPI sięgają 17 punktów procentowych. W przypadku pytań faktualnych różnice zasadniczo nie występują.

Techniki ankiety ze wspomaganiami komputerowymi, a więc przede wszystkim CAWI. Łączą one wszystkie zalety technik ankiety i technik wspomaganymi komputerowo. Oznacza to szybkość realizacji badania, niskie koszty przygotowania narzędzia i realizacji oraz możliwość wielokrotnego monitowania. Podstawowym ograniczeniem jest jednak relatywnie niski i bardzo zróżnicowany geograficznie poziom wyposażenia gospodarstw domowych w Internet (tzw. poziom penetracji, który w Polsce wynosi ok. 50%) i związany z tym błąd pokrycia [Dillman 2000; Sztabiński F. 2009; Żmijewska-Jędrzejczyk 2004].

efektywnej metodologii „łączenia” prób telefonów stacjonarnych i komórkowych, realizacja badań CATI bazuje na próbach telefonów stacjonarnych [Haeder et al. 2010: 3–9].

¹¹ Ostatnia cecha ankiety – społeczne zakotwiczenie opinii – nie dotyczy oczywiście ankiety pod kontrolą oraz audytoryjnej. Dyskusyjne jest to, czy owo „zakotwiczenie” społeczne jest ograniczeniem, czy też zaletą ankiet.

Znajomość zalet, ograniczeń i specyfiki poszczególnych technik jest niezbędna na etapie conceptualizacji badania, kiedy podejmujemy decyzję dotyczącą wyboru technik, które zamierzamy łączyć. Generalnie wydaje się, że należy dążyć do łączenia technik maksymalnie do siebie zbliżonych ze względu na sposób dotarcia i komunikowania, a unikać łączenia technik całkowicie odmiennych. Jednocześnie musimy uwzględnić problematykę badania, charakter badanej zbiorowości (chodzi tutaj o wspomniany problem społecznej i psychologicznej adekwatności techniki) oraz koszty badania. Wybierając sposób „mieszania technik”, musimy przede wszystkim odwoływać się do własnej wiedzy, doświadczenia badawczego oraz intuicji badawczej.

Ponieważ, jak wspomniano, efekt techniki wiąże się z odmiennością sformułowania pytań, odmiennością sposobu kontaktowania i komunikowania się z respondentem, rozważymy, jakie są konsekwencje owych odmienności i możliwości ograniczenia ich skutków. Przedtem jednak opiszemy projekt, którego rezultaty będziemy prezentować.

OPIS PROJEKTU

Dane, na podstawie których będziemy prowadzić dalsze rozważania dotyczące efektu techniki zostały zebrane w ramach projektu JRA1 „Strategie zbierania danych: możliwości łączenia różnych technik zbierania danych w kolejnych rundach ESS”, który był częścią metodologicznego grantu ESS Infrastructure: i3. Projekt ten miał na celu m.in. podjęcie próby stworzenia narzędzia „odpornego” na efekt techniki oraz ocenę błędu braku odpowiedzi i błędu pomiaru związanego z daną techniką. Chodziło zatem o porównanie poszczególnych technik na dwóch płaszczyznach – realizacyjnej (jaki jest faktyczny *response rate* przy zastosowaniu różnych technik, przy zachowaniu porównywalności prób i narzędzi) i merytorycznej (w jakim stopniu dane uzyskane różnymi technikami są zbieżne).

Mimo że projekt JRA1 był realizowany od roku 2006, już wcześniej, bo w roku 2003 we współpracy z Gallup Europe przeprowadzone zostały (na Węgrzech, a w 2005 r. na Węgrzech i w Portugalii) wstępne badania metodologiczne dotyczące *Mixed Mode*. Na ich podstawie można stwierdzić, iż techniką najbardziej różniącą się od technik wywiadu osobistego (PAPI i CAPI) jest CATI, pomijając ankietę pocztową [por. Sztabiński P.B. 1999b: 51–66], jednocześnie żadna pojedyncza technika nie może być konkurencyjna wobec wywiadu PAPI [Eva, Widdop 2007]. Stąd też w niniejszym projekcie podjęto próbę udzielenia odpowiedzi na pytania w odniesieniu do techniki PAPI, CAPI oraz CATI.

Badanie ESS-CATI przebiegało wg schematu eksperymentalnego opracowanego przez Caroline Roberts i Petera Lynna, przy współpracy Gillian Eva. W projekcie porównano trzy techniki: CATI, CAPI i PAPI. Było to możliwe, ponieważ w projekcie uczestniczyło 5 krajów: Cypr, Niemcy, Polska, Szwajcaria oraz Węgry, z których niektóre realizowały badanie zasadnicze ESS r. 3 techniką PAPI, inne zaś CAPI. Porównanie dotyczyło ESS-CAPI i ESS-CATI (Niemcy, Szwajcaria) oraz ESS-PAPI i ESS-CATI (Cypr, Polska, Węgry).

Procedura badania ESS-CATI zakładała realizację trzech różnych wersji kwestionariusza ESS r. 3: A, B i C, które systematycznie alokowano w trakcie realizacji do wylosowanej próby. W grupie A, która stanowiła 40% próby (400 przypadków) realizowana była pełna wersja wywiadu ESS (czas trwania szacowany na ok. 60 min.); w grupie B, która stanowiła 20% próby (200 przypadków) realizowana była skrócona wersja wywiadu (czas trwania szacowany na ok. 45 min.); w grupie C, która stanowiła 40% próby (400 przypadków) realizowana była także pełna wersja wywiadu ESS, ale podzielona na dwie części (z każdą osobą realizowane były dwa wywiady po ok. 30 min. (zob. Aneks: Schemat badania ESS-CATI r. 3).

W Polsce badanie ESS-CATI przeprowadzał Instytut MillwardBrown SMG/KRC od grudnia 2006 do lutego 2007, pod kierownictwem i przy nadzorze metodologicznym IFiS PAN. Poszczególne fazy przygotowania i realizacji badania były następujące:

1. Przygotowanie „papierowych” wersji kwestionariuszy ESS-CATI oraz napisanie i testowanie skryptów trwało relatywnie długo (około 1 miesiąca). Po ostatecznym sprawdzeniu skryptu testowano go w terenie (przeprowadzono ponad 140 wywiadów pilotażowych).

2. Opracowanie założeń doboru próby. Próba do badania ESS-CATI z założenia miała być porównywalna z próbą w badaniu zasadniczym (ESS-PAPI). Ponieważ było to niemożliwe (badanie zasadnicze realizowano na próbie losowanej z operatu PESEL, w którym brak informacji o numerach telefonów), podjęto decyzję o doborze próby metodą List-Assisted RDD (Random Digit Dialing), a więc losowania końcówek numeru do istniejących i sprawdzonych prefiksów. Sposób warstwowania próby był identyczny jak w przypadku badania zasadniczego. Na etapie losowania respondenta zastosowano procedurę doboru Kisha. Badanie realizowano na 1000-elementowej próbie dorosłej ludności Polski w wieku 15 i więcej lat (bez górnej granicy), zamieszkującej w indywidualnych gospodarstwach domowych. Była to próba realizowana „do wyczerpania” (zastępowano tylko numery spoza próby – telefony nieistniejące oraz należące do

instytucji). Poszczególne wersje kwestionariusza ESS-CATI (A, B i C) były systematycznie alokowane do próby.

3. Szkolenie ankierów do badania ESS-CATI odbywało się w ramach czterech etapów. Pierwsze szkolenie przeprowadzono 3 tygodnie przed rozpoczęciem badania i poświęcono prezentacji projektu ESS (w tym „sposobom” zwiększania odsetka realizacji i zasadom konwersji miękkich odmów – *Refusal Aversion Training*, RAT) [Czaja, Blair 2005; Groves, Couper 1998; Groves, McGonagle 2001: 249–265]. Drugie szkolenie przeprowadzono 2 tygodnie przed rozpoczęciem realizacji badania. Zostały omówione poszczególne wersje kwestionariusza (skryptów), a następnie „pytanie po pytaniu” przeanalizowano skrypt w wersji A (jako „bazowy”). Trzecie szkolenie odbyło się w przeddzień rozpoczęcia terenowej realizacji projektu. Jego przedmiotem była powtórna, szczegółowa analiza kodów opisu rezultatu kontaktu, połączona z dyskusją dotyczącą ich „użyteczności” (ankierzy przeprowadzili ponad 140 wywiadów pilotażowych i mogli przedstawić własne doświadczenia). Ostatnie, czwarte z kolei szkolenie odbyło się w połowie okresu realizacji badania (4 tygodnie po rozpoczęciu). Miało ono na celu wymianę doświadczeń z realizacji badania oraz, co było ważniejsze, pełniło funkcje motywacyjne wobec ankierów.

4. Realizacja badania ESS-CATI trwała niemal 2 miesiące (od 14.12.2006 do 10.02.2007). Do jego realizacji wyznaczono 27. najbardziej doświadczonych ankierów, tzn. realizujących wywiady od kilku lat. Niezależnie od ich doświadczeń poddano ich wspomnianemu 4-etapowemu szkoleniu. W sumie podjęto ponad 14 000 kontaktów telefonicznych z ponad liczbą abonentów przekraczającą 1 400. Alokacja wywiadów dla ankierów w poszczególnych wersjach była losowa przy założeniu, iż jeden ankier nie może zrealizować więcej niż 35 wywiadów.

5. Nadzór i monitoring nad realizacją badania przebiegał dwutorowo. Był to wewnętrzny, bieżący nadzór sprawowany przez Instytut MillwardBrown SMG/KRC (odsłuchiwanie rozmów, obserwacja poprawności przebiegu sesji, zarówno od strony technicznej, jak i merytorycznej). Szczegółowej analizie poddano ponad 20% wywiadów zrealizowanych przez 23 ankierów (spośród 27). Kontrola każdego z przypadków obejmowała przynajmniej 75% całkowitego czasu trwania wywiadu. Instytut MillwardBrown SMG/KRC w odstępach tygodniowych miał obowiązek raportowania do IFiS PAN statystyk realizacyjnych. Drugi typ – nadzór zewnętrzny, sprawowany przez IFiS PAN – polegał na bieżącej, wrywkowej kontroli pracy ankierów (odsłuchiwanie rozmów, obserwacja poprawności przebiegu sesji itp.).

6. Niezależna kontrola terenowa (telefoniczna) przeprowadzona została przez IFiS PAN w lutym 2007 r. i objęła 10% przypadków odmów. Do kontroli

dobrano przypadki realizowane przez 3 ankieterów, którzy uzyskali najniższe wskaźniki realizacji, tzn. mieli relatywnie największą liczbę odmów.

Dość szczegółowo omówiono realizację terenową projektu JRA1. w Polsce. Wydaje się to konieczne, ponieważ analizujemy tutaj m.in. możliwości ograniczenia wpływu różnych sposobów kontaktowania się z respondentem na wynik badania.

EFEKT TECHNIKI (NA PRZYKŁADZIE ESS-PAPI I ESS-CATI)

Jak wspomnieliśmy, różnice między techniką PAPI i CATI sprowadzają się do odmienności związanej z formułowaniem pytań, sposobem kontaktowania się z respondentem oraz sposobem komunikowania się z nim. Wszystko to skutkuje odmiennością efektów komunikacyjnych, a więc wpływa na uzyskiwane wyniki. Aby dowiedzieć się, jakie są różnice oraz czy istnieją możliwości redukcji efektu techniki i jakie one są, należy kolejno przeanalizować wszystkie trzy rodzaje odmienności.

Różnice w sposobie sformułowania pytań

Mówiąc o redukowaniu różnic dotyczących sposobu formułowania pytań, mamy na myśli podjęcie próby stworzenia narzędzia, a więc pytań kwestionariuszowych oraz instrukcji dla ankietera „odpornych” na efekt techniki. Największym „wyzwaniem” jest oczywiście przygotowanie kwestionariusza, różniącego się w zależności od technik, które zamierzamy wykorzystać w *Mixed Mode*, ale także ze względu na model *Mixed Mode* (sekwencyjny lub równoczesny). W przypadku modelu sekwencyjnego, po określeniu metody podstawowej („wyjściowej”) należy przygotować narzędzie dla tej techniki, a w kolejnym kroku stworzyć jego ekwiwalentne wersje dla innych technik. W drugim przypadku (równoczesny model *Mixed Mode*) brak jednej, „preferowanej” (podstawowej) metody. Wszystkie techniki traktowane są jako ekwiwalentne. Czym zatem się kierować i w jakiej kolejności przygotować taki kwestionariusz?. Możliwe są trzy podejścia [de Leeuw, Hox, Dillman 2008b: 312–313]:

1) pierwsze z nich zakłada optymalizację narzędzia dla danej techniki (tzw. *Mode Specific Design*) – oznacza to, że dla każdej techniki należy przygotować „osobne”, równoważne (niekoniecznie identyczne) narzędzie;

2) drugie umożliwia maksymalne upodobnienie każdego narzędzia na poziomie bodźca, tj. sformułowania pytania (tzw. *Uni-Mode Design*) [Dillman 2000: 232–240] oraz unifikację formatu pytań, odpowiedzi oraz instrukcji. Unifikacja

ta polega m.in. na: (a) zachowaniu jednolitej struktury pytań dla poszczególnych technik (zmiana struktury może zmienić bodziec); (b) redukowaniu liczby możliwych odpowiedzi (winny być takie same i włączone w treść pytania, także odpowiedź „Trudno powiedzieć”); (c) ustaleniu ekwiwalentnych instrukcji i reguł przejścia (jeśli byłyby bardzo skomplikowane, należy rozważyć restrukturyzację kwestionariusza w celu ich uproszczenia);

3) trzecie stawia na ekwiwalentność pytań na poziomie kognitywnym (tzw. *Generalized Mode Design*) [de Leeuw, Hox, Dillman, 2008b: 312–313]. Odchodzi się tutaj od literalnej unifikacji bodźca na poziomie formalnym, ponieważ identyczne sformułowanie nie zawsze zapewnia jego równoważność. Zakłada się natomiast, iż różne bodźce (pytania) mogą przynosić ekwiwalentne odpowiedzi na poziomie zapotrzebowania badawczego. Jeśli chcemy się dowiedzieć, czy skontaktowaliśmy się z właściwą osobą, to stosując technikę CATI zadamy pytanie: „Czy rozmawiam z.....? (imię i nazwisko)”; z kolei stosując ankietę Web-ową, postawimy pytanie: „Czy nazywasz się.....? (imię i nazwisko)” [de Leeuw 2009].

Wydaje się jednak słuszne, aby przy wyborze technik, które mają być łączone, kierować się jeszcze jedną zasadą. Gdy któraś z nich stawia przed nami większe w porównaniu z innymi „wymagania” w stosunku do kwestionariusza, sformułowania pytań, grafiki bądź języka programowania, to należy je uwzględnić w przypadku pozostałych technik. Jeśli w badaniach internetowych stosujemy technikę e-mailingu oraz ankietę Web-ową (bardziej „zaawansowaną”), to ta pierwsza, na poziomie formalnym, nie powinna być bardziej skomplikowana niż ankietę Web-owa – winny być przygotowane w podobny sposób, w podobnych „językach”.

W omawianym projekcie, podczas konstrukcji kwestionariusza ESS-CATI przyjęto perspektywę kognitywną, tj. zasadę unifikacji na poziomie bodźca, a więc ekwiwalentności koncepcyjnej, nie zaś unifikacji na poziomie formalnym. Oznacza to, iż dla „pierwotnego” kwestionariusza PAPI został stworzony jego „odpowiednik” dla techniki CATI. Poniżej prezentujemy przykład pytania źródłowego (wersja PAPI) i pytania adaptowanego do CATI (takiego, jakie widział ankietę na ekranie komputera).

Przykład pytania: ESS-PAPI

B24. Biorąc wszystko pod uwagę, ogólnie, na ile jest P. zadowolony/-a ze swojego obecnego życia? Odpowiadając, proszę <u>posłużyć się kartą [KARTA 10]</u> , gdzie 0 oznacza, że jest P. zupełnie niezadowolony/-a, a 10 oznacza całkowicie zadowolony/-a.											
KARTA 10. POZOSTAWIĆ DO B27											
Zupełnie niezadowolony/-a						Całkowicie zadowolony/-a				(Trudno powiedzieć)	
00	01	02	03	04	05	06	07	08	09	10	88

Przykład pytania: ESS-CATI

TB24. Biorąc wszystko pod uwagę, ogólnie, na ile jest P. zadowolony/-a ze swojego obecnego życia? Odpowiadając, proszę <u>wskazać liczbę od 0 do 10</u> , gdzie 0 oznacza, że jest P. zupełnie niezadowolony/-a, a 10 oznacza, że całkowicie zadowolony/-a. <u>Pozostałe liczby służą do wyrażenia opinii pośrednich.</u>											
Zupełnie niezadowolony/-a						Całkowicie zadowolony/a				(Trudno powiedzieć)	
00	01	02	03	04	05	06	07	08	09	10	88

Powyższy przykład stanowi dobrą ilustrację odmienności sformułowania pytania w PAPI i CATI oraz instrukcji do pytania, w tym wykorzystania materiałów pomocniczych (chodzi o Kartę Respondenta).

Różnice w sposobie kontaktowania się i ich efekty

Różnice w sposobie kontaktowania się najlepiej obrazuje odsetek realizacji próby wylosowanej przy zastosowaniu wspomnianych dwóch technik: wywiadu osobistego (PAPI/CAPI) oraz telefonicznego (CATI). Dla porównania, dla poszczególnych krajów biorących udział w projekcie JRA 1, w poniższych tabelach 1. i 2.) zamieszczono także odsetek realizacji próby wylosowanej (*Response Rate: RR*) w badaniu zasadniczym¹².

¹² Odsetek ten został wyliczony zgodnie z definicją tzw. Response Rate (RR2), przyjętą przez American Association for Public Opinion Research [2008]. Jest on definiowany jako liczba zrealizowanych wywiadów (kompletnych lub częściowo zrealizowanych), dzielona przez liczebność wylosowanej próby, pomniejszoną o liczbę przypadków, które nie powinny znaleźć się w próbie, np. błędny adres/numer telefonu; adres/numer telefonu instytucji itp.

TABELA 1. Odsetek realizacji próby (AAPOR RR2) w poszczególnych krajach biorących udział w projekcie JRA1 dla badania realizowanego techniką PAPI/CAPI oraz CATI

Kraj	Próba	Technika ESS r. 3: (RR%)	ESS r. 3 CATI: (RR%)	Różnica
Cypr	Gospodarstw	PAPI: 67,3	8,3	-59,0
Niemcy	Indywidualna	CAPI: 52,9	24,3	-28,6
Węgry	Adresowa	PAPI: 65,9	25,2	-40,7
Polska	Indywidualna	PAPI: 70,1	35,3	-34,8
Szwajcaria	Gospodarstw	CAPI: 50,0	41,5	-8,6

Źródło: Roberts et al.: 2010.

Dla poszczególnych krajów ogólny odsetek realizacji próby wylosowanej jest, co oczywiste, różny – w zależności od wersji kwestionariusza ESS-CATI (A, B, C), wyróżnionej ze względu na przewidywany czas trwania wywiadu. Ilustruje to tabela 2.

TABELA 2. Odsetek realizacji próby wylosowanej w zależności od wersji kwestionariusza ESS-CATI

Kraj	Technika ESS r. 3 (RR%)	ESS r. 3 CATI (RR%)	Wersja A (%)	Wersja B (%)	Wersja C (%)
Cypr	PAPI: 67.3	8,3	8,8	6,2	9,0
Niemcy	CAPI: 52.9	24,3	21,8	26,3	25,7
Węgry	PAPI: 65.9	25,2	19,3	24,5	31,5
Polska	PAPI: 70.1	35,3	34,1	38,0	35,3
Szwajcaria	CAPI: 50.0	41,5	37,9	39,5	50,2

Źródło: Roberts et al. 2010.

Jak można zauważyć, w każdym z krajów uczestniczących w badaniu ESS-CATI uzyskano znacząco niższe odsetki realizacji niż w badaniu zasadniczym (PAPI lub CAPI). Największe różnice – blisko 60 punktów procentowych – odnotowano dla Cypru (PAPI vs CATI), najmniejsze dla Szwajcarii, niespełna 9 punktów procentowych (CAPI vs CATI). Co ciekawe, uzyskano tam – w porównaniu z innymi krajami – najniższy odsetek realizacji w badaniu zasadniczym, a najwyższy w badaniu przy zastosowaniu techniki CATI. Zwraca także uwagę większa rozpiętość odsetka realizacji w badaniu CATI niż w badaniu zasadniczym: o ile dla technik *face-to-face* wynosi ona 20 punktów procentowych, o tyle dla CATI osiąga ponad 33 punkty procentowe. Nie podejmujemy się interpretacji tak

dużych rozpiętości wyników, ponieważ wymagają one znajomości kulturowych i społecznych uwarunkowań realizacji badań w każdym z krajów¹³.

W Polsce zrealizowano 353 wywiady, przy czym w 301 przypadkach w całości, a w 52 częściowo (zostały przerwane przez respondenta lub respondenci nie wyrazili zgody na realizację drugiej części wywiadu: wersja C). Spośród wszystkich wywiadów w wersji A udało się przeprowadzić 136, w wersji B – 76, a w wersji C – 141. Pamiętając, iż proporcje w próbie wynosiły odpowiednio: 400, 200 i 400, zauważyć można, że najwięcej wywiadów udało się zrealizować w wersji B (najkrótszej – założony czas trwania 45 min.), następnie w wersji C (2×30 min.), natomiast najmniej w wersji pełnej A (60 min.). To, że w wersji C zrealizowano relatywnie mało wywiadów (tylko o 5 więcej, niż w wersji A) wydawać się może zaskakujące, gdyż w założeniu miała to być najbardziej „przyjazna” dla respondentów opcja. Jednocześnie respondenci aż w 41 przypadkach po zakończeniu realizacji pierwszej części wywiadu nie wyrazili zgody na jego kontynuację. Wyjaśnić to można, jak się wydaje, błędną aranżacją do tej wersji kwestionariusza. W projekcie przyjęto, że w trakcie aranżacji wstępnej do pierwszej części wywiadu (część a.) respondent jest informowany, że wywiad będzie trwał ok. 30 min., a dopiero po jego zakończeniu, w aranżacji końcowej, proszony jest o zgodę na drugi wywiad, także trwający 30 min. (część b.). W takiej sytuacji część respondentów czuła się „oszukana” i nie wyrażała zgody na realizację drugiej części wywiadu.

Warto wspomnieć, że rozpiętość czasów trwania wywiadów była dość duża. Dla poszczególnych wersji maksymalna i minimalna długość wywiadu (podczas jednego kontaktu) wynosiła odpowiednio: wersja A – 150 min. i 41 min., wersja B – 75 min. i 32 min., wersja C (część a. i b.) – 132 min. i 44 min.

Jeśli chodzi o wywiady niezrealizowane, to bardzo liczną kategorię stanowiły przypadki „braku odpowiedzi po 7 sygnałach”, a więc te, w których nikt nie odbierał telefonu (120 na ogółem 131 przypadków „nienawiązania kontaktu”). Wydaje się, że było to związane z terminem realizacji badania (14.12.2006–10.02.2007). Był to okres wyjazdów „świętecznych” i „międzyświętecznych”, a potem „ferii zimowych” w szkołach.

Z kolei w sytuacji, gdy ankieterowi udało się nawiązać kontakt, najczęstsze były przypadki, w których nie zakończył procedury doboru respondenta i umawiał się na powtórny telefon w innym terminie (25 przypadków na 59). Sytuację tę traktować można jako „pośrednią odmowę”, podobnie 13 przypadków, w któ-

¹³ Zgodnie z interpretacją Dominique Joye z Université de Lausanne, narodowego koordynatora projektu ESS, w Szwajcarii „większość spraw załatwia się przez telefon”. Stąd relatywnie niski odsetek realizacji wywiadu techniką CAPI, a wysoki – CATI.

rych wylosowano respondenta, ale nie można było się z nim skontaktować lub wielokrotnie „przekładał” terminy wywiadu.

Bardzo liczną kategorię stanowiły „odmowy” udzielenia wywiadu (45,7% przypadków, w których nawiązano kontakt z numerem należącym do próby). Należy jednak zwrócić uwagę, że na 368 takich przypadków jedynie w 63. zgody na wywiad nie wyraził wylosowany respondent (w 305 przypadkach „odmowa” miała miejsce przed wylosowaniem respondenta). Oznacza to, że kluczową sprawą dla powodzenia badania jest pozyskanie kooperacji osoby, która odbiera telefon.

Z kolei zasadniczą przyczyną konieczności włączenia dodatkowych numerów do pierwotnie wylosowanej próby był „nieistniejący numer” (ponad 70% wszystkich dodatkowych numerów) oraz „numery instytucji” (24,3%).

Mówiąc o różnicach w sposobie kontaktowania się (kontakt telefoniczny – kontakt *face-to-face*), kolejną ważną kwestią jest pytanie o podobieństwo struktury próby zrealizowanej w badaniu ESS-PAPI i ESS-CATI. Podstawowe charakterystyki społeczno-demograficzne dla obu prób są zawarte w tabeli 3.

TABELA 3. Porównanie próby zrealizowanej w ESS-PAPI i ESS-CATI

Polska: ESS r. 3	PAPI (%)	CATI (%)	Różnica
N	1718 (100%)	353 (100%)	
Mężczyźni [w %]	47,4	40,1	7,3 *
Średnia wieku [w latach]	43,7	47,7	4,0
Obecnie wykonuje pracę odpłatną [w %]	48,0	51,0	3,0
Miejsce zamieszkania			
- duże miasto	24,9	16,6	8,3
- przedmieścia lub obrzeża dużego miasta	5,3	9,6	4,3
- średnie lub małe miasto	31,3	32,1	0,8
- wieś	37,6	33,4	4,2
- pojedyncze gospodarstwo lub dom na wsi	0,9	8,3	7,4
Główne zajęcie w ciągu ostatnich 7 dni			
- praca odpłatna	35,8	45,2	9,4 *
- nauka	14,5	6,6	7,9
- bezrobotny, poszukujący pracy	3,9	0,6	3,3
- bezrobotny, nie poszukujący pracy	2,8	1,4	1,4
- trwała choroba, niepełnosprawność	1,6	0,3	1,3
- emerytura/renta	25,3	32,6	7,3
- opieka nad domem	14,9	10,1	4,8
- inne	1,1	3,2	2,1
Lata skolaryzacji (średnia)	12,4	13,0	0,6

* $p \leq 0,05$

Źródło: Roberts et al. 2010 i obliczenia własne

Podsumowując, różnice w efektach kontaktowania się *face-to-face* oraz telefonicznego są znaczące. Czy zatem, w badaniu ESS w Polsce technika CATI może stanowić alternatywę dla PAPI? Tak, jeśli uznamy za satysfakcjonujące, że: (i) ogólny odsetek zrealizowanych wywiadów nie jest zbyt niski (35%); (ii) w warunkach polskich koszty uzyskania 1 przypadku w badaniu CATI są niewiele niższe niż PAPI i wyrażają się stosunkiem 9: 10. Należy jednak pamiętać, że odpowiedź ta jest warunkowa i stanowi odpowiedź częściową. Nie wiemy, jaki jest wpływ zastosowanej techniki (chodzi o tzw. *mode effects*) na wartość uzyskanych danych, a więc efekty komunikacyjne. Czy i w jakim stopniu uzyskane wyniki w ESS-CATI różnią się od wyników ESS-PAPI spróbujemy odpowiedzieć w dalszej części opracowania.

Różnice w sposobie komunikowania się i ich efekty

Obecnie na przykładzie kilku wybranych pytań, różnych ze względu na przedmiot (postawy/opinie i zachowania) oraz strukturę (dychotomiczne oraz skale semantyczne i liczbowe) przeanalizujemy, czy istnieją różnice efektów komunikacyjnych, a więc dotyczące udzielanych odpowiedzi w wywiadzie ESS-PAPI i ESS-CATI, a jeśli występują, to czy są znaczące?

Porównanie wyników tych dwóch badań prowadzić będziemy na próbach „matchowanych”, tzn. porównywalnych ze względu na kryteria warstwowania (płeć, wiek respondenta oraz wielkość miejscowości zamieszkania). W celu zrównoważenia struktur obu prób przeważaliśmy próbę zrealizowaną w badaniu ESS-PAPI (znacząco większą) do struktury próby zrealizowanej w badaniu ESS-CATI. Dalsze próby zrównoważenia struktur porównywanych prób ograniczała stosunkowo mała liczebność próby zrealizowanej w badaniu ESS-CATI.

Jako pierwsze przeanalizujemy pytanie dychotomiczne, o zachowania, związane z udziałem w wyborach do sejmiku we wrześniu 2005 r. W tabeli 4 przedstawiono rozkład odpowiedzi.

TABELA 4. Rozkład odpowiedzi na pytanie o udział w wyborach do sejmiku w 2005 r.

TB11/B11. Obecnie wielu ludzi z różnych przyczyn nie bierze udziału w wyborach. Czy brał/-a P. udział w ostatnich wyborach do sejmiku we wrześniu 2005 r.?	PAPI (%)	CATI (%)	Różnica
1. Tak	64,3	74,7	10,4
2. Nie	30,8	24,0	6,8
3. Nie byłem/-am uprawniony/-a do głosowania	4,4	1,3	3,1
8. trudno powiedzieć	0,4	0,0	0,4
9. brak odpowiedzi	0,1	0,0	0,1

$$\text{Chi}^2 = 7,58; p \leq 0,01; \text{df} = 1$$

Różnice rozkładów odpowiedzi (z wyłączeniem odpowiedzi: „nie byłem uprawniony”/„trudno powiedzieć”)¹⁴ są istotne w teście Chi². Należy zwrócić uwagę na ponad 10-punktową różnicę w przypadku odpowiedzi potwierdzającej udział w wyborach: w wywiadzie CATI respondenci częściej udzielali takich odpowiedzi niż w PAPI.

Warto przypomnieć, że frekwencja wyborcza w 2005 r. wyniosła 40,6%, a więc także w wywiadzie PAPI odsetek odpowiedzi twierdzących jest zawyżony w porównaniu ze stanem faktycznym. Można wobec tego przypuszczać, że jest to pytanie drażliwe, skłaniające do udzielania społecznie akceptowanych odpowiedzi, choć – co zaskakujące – efekt ten ujawnił się mocniej w CATI. Być może taki wynik jest powiązany ze wspomnianą wcześniej tendencją do „potakiwania”, a może jest to efekt różnic doboru próby i domniemanego jej „wypaczenia” w ESS-CATI.

W następnej kolejności przeanalizujemy pytanie dychotomiczne o postawy dotyczące poczucia związku z partią polityczną. W tabeli 5. ujęto rozkład odpowiedzi na to pytanie.

TABELA 5. Rozkład odpowiedzi na pytanie o poczucie związku z jakąś partią polityczną

TB20a/B20a. Czy istnieje jakaś partia polityczna, z którą czuje się P. bardziej związany/-a niż z innymi partiami, jest P. jej sympatykiem?	PAPI (%)	CATI (%)	Różnica
1. Tak	26,7	47,0	20,3
2. Nie	71,9	53,0	18,9
8. trudno powiedzieć	1,3	0,0	1,3
9. brak odpowiedzi	0,2	0,0	0,2

$$\text{Chi}^2 = 47,96; p \leq 0,05; \text{df} = 1$$

Wynik testu Chi² także wskazuje na istotne różnice w rozkładach odpowiedzi na to pytanie. Analogicznie jak w poprzednim pytaniu, w wywiadzie CATI występuje znacząca przewaga odpowiedzi twierdzących (w porównaniu z wywiadem PAPI). Mimo iż pytania te różni przedmiot – jedno z nich dotyczy zachowań, drugie opinii – w obu prawdopodobnie ujawnia się efekt, wspomnianej wcześniej, tendencji do „potakiwania”.

¹⁴ Za wyłączeniem tych kategorii przemawia to, że zwiększanie liczby kategorii obniża wiarygodność testu dla rozkładów (tym bardziej, że liczebności dla tych kategorii są bardzo niskie). Powyższa uwaga dotyczy również kolejnych tabel.

Pytanie trzecie o opinie (postawy), dotyczy zaufania do innych ludzi. Wykorzystano w nim 11-punktową skalę liczbową z opisanymi krańcami. W tabeli 6. zilustrowano rozkład odpowiedzi na to pytanie.

TABELA 6. Rozkład odpowiedzi na pytanie o poczucie zaufania

TA8/A8. Czy, ogólnie biorąc, uważa P., że większości ludzi można ufać, czy też, że w kontaktach z ludźmi ostrożności nigdy za wiele? Odpowiadając proszę wskazać liczbę od 0 do 10, gdzie 0 oznacza, że ostrożności nigdy za wiele, a 10 oznacza, że większości ludzi można ufać. Pozostałe liczby służą do wyrażenia opinii pośrednich.	PAPI (%)	CATI (%)	Różnica
00. Ostrożności nigdy za wiele	10,3	20,0	9,7
01.	6,6	3,0	3,6
02.	11,6	7,7	3,9
03.	14,7	7,3	7,4
04.	8,8	10,3	1,5
05.	24,9	26,0	1,1
06.	6,7	7,3	0,7
07.	7,2	7,0	0,2
08.	5,8	8,3	2,5
09.	1,1	1,7	0,6
10. Większości ludzi można ufać	1,8	1,3	0,5
88 trudno powiedzieć	0,5	0,0	0,5

$\chi^2 = 44,34; p \leq 0,01; df = 10$

Różnice rozkładów odpowiedzi dla tego pytania także są istotne w teście χ^2 . Taki wynik można wiązać z trudnością odbioru skali liczbowej w wywiadzie CATI. W wywiadzie PAPI, po odczytaniu treści pytania, ankieter wręczał respondentowi kartę, na której zamieszczono wizualizację skali, tzn. 11 punktów, oznaczonych liczbami od „00” do „10”, z odpowiednim opisem jej krańców: „Ostrożności nigdy za wiele” oraz „Większości ludzi można ufać”. Respondent widział skalę, na której ma wskazać odpowiedni punkt oraz opis jej krańców. W wywiadzie CATI natomiast respondent musiał „wyobrazić sobie” taką skalę jedynie na podstawie odczytanego przez ankietera opisu. Bez wątpienia musiało być to dla niego trudne zadanie. Analizując rozkłady odpowiedzi w wywiadzie CATI, zaobserwować można, że respondenci „skracali” skalę i wskazywali odczytywany jako pierwszy punkt skrajny („00. Ostrożności nigdy za wiele”)

oraz środkowy („05”) [Sztabiński F. 2003: 147–175]. W sumie te dwa punkty wskazało 46% respondentów¹⁵. O występowaniu tendencji do „spłaszczania” skali w wywiadzie CATI świadczy także niska częstość wskazań punktów „pośrednich”, w szczególności punktów: „01”, „02” i „03”. Jeden z tych punktów w wywiadzie PAPI wskazała prawie 1/3 respondentów (32,9%), zaś w wywiadzie CATI tylko 18%¹⁶.

Ostatnie pytanie o postawy wobec osób mających orientację homoseksualną, z wykorzystaniem 5-punktowej skali semantycznej. W tabeli 7 zilustrowano rozkład odpowiedzi na to pytanie.

TABELA 7. Rozkład odpowiedzi na pytanie o stosunek do osób o orientacji homoseksualnej

TB31/B31. Proszę powiedzieć, w jakim stopniu zgadza się P. albo nie zgadza z następującym stwierdzeniem: „Osoby o orientacji homoseksualnej, geje i lesbijki, powinni mieć swobodę układania sobie życia wg własnych przekonań”. Czy P. ... <i>ODCZYTAĆ</i>	PAPI (%)	CATI (%)	Różnica
1. Zdecydowanie się zgadzam	8,9	6,7	2,2
2. Zgadzam się	32,6	37,0	4,4
3. Ani się zgadzam, ani nie zgadzam	20,1	23,0	2,9
4. Nie zgadzam się	20,9	19,7	1,2
5. Zdecydowanie się nie zgadzam	10,3	12,3	2,0
7. odmowa odpowiedzi	0,1	0,0	0,1
8. trudno powiedzieć	7,1	0,0	7,1
9. brak odpowiedzi	0,1	1,3	1,2

$$\text{Chi}^2 = 4,04, p < 0,05 (0,4008122); \text{df} = 4$$

Generalnie różnice rozkładów odpowiedzi nie są duże (maksymalna wynosi |4,4| punktu procentowego). Wynik testu Chi^2 – przy założeniu, że $p < 0,05$ – nie wykazał istotnych różnic w rozkładach odpowiedzi. Jest to zupełnie inna sytuacja niż w pytaniu poprzednim, mimo że obydwa dotyczyły opinii. Wydaje się,

¹⁵ Ankieterzy realizujący badanie zwracali uwagę na to, iż skale te sprawiały respondentom wiele trudności: prosili oni o powtarzanie opisów ich krańców lub dodatkowe wyjaśnienia. Zaobserwowali także tendencję do ich „skracania” i „przekładania” na skale słowną (tzn. łączenia poszczególnych punktów w większe całości i wskazywania określonych punktów, odpowiadających w ich odczuciu skali słownej).

¹⁶ Tendencji takiej nie zaobserwowano w przypadku „drugiej połówki” skali (punkty: od „06” do „10. Większości ludzi można ufać”). Jest to prawdopodobnie powiązane ze skośnością rozkładu odpowiedzi – punkty te wskazywane były relatywnie rzadziej – w sumie wskazało je 22% respondentów w wywiadzie PAPI i 25% w CATI.

iż można to wyjaśnić dwojako. Po pierwsze, pięciopunktowa skala semantyczna była dla respondentów z pewnością „łatwiejsza” niż 11-punktowa skala liczbowa. Po drugie, powyższe pytanie dotyczyło drażliwej i bardzo konkretnej kwestii: akceptacji bądź jej braku dla związków homoseksualnych, natomiast poprzednie sprawy bardzo ogólnej, związanej z indywidualnymi doświadczeniami – zaufania do ludzi. W wywiadzie PAPI respondenci mieli możliwość zastanowienia się, „głośnego” myślenia i dopiero potem wskazania właściwego punktu na skali. W wywiadzie CATI, ze względu na charakter i szybkie tempo rozmowy telefonicznej, byli oni pozbawieni takiej możliwości.

Analizując efekt odmienności sposobów komunikowania się w PAPI i CATI, z pewnością można stwierdzić, iż wystąpił on w analizowanych pytaniach. Zaskoczenie stanowi to, że szczególnie silnie wystąpił on w przypadku pytań dychotomicznych (różnice sięgają 20 punktów procentowych), przy czym, co oczywiste, jest on słabszy w pytaniach dotyczących zachowań, silniejszy natomiast w pytaniach o opinie. W przypadku pytań dychotomicznych różnice w odpowiedziach wynikają najprawdopodobniej ze wspomnianej tendencji do „potakiwania” w sytuacji wywiadu CATI.

W przypadku pytań ze skalami nie można sformułować tak jednoznacznego wniosku. Przede wszystkim zależy to od rodzaju zastosowanej skali. Stosując abstrakcyjne, długie i z pewnością trudne dla respondentów skale (np. liczbowe) musimy liczyć się z możliwością wystąpienia efektu techniki. W przypadku prostych skal semantycznych ma on raczej ograniczone znaczenie. Jak wskazaliśmy, czynnikiem „interweniującym” może być także tematyka pytania.

Mówiąc o efekcie techniki, warto zwrócić uwagę na jeszcze jedną kwestię. Otóż w żadnym z analizowanych pytań w wywiadzie CATI nie pojawiły się odpowiedzi typu: „nie wiem”/”trudno powiedzieć”, spotykane w wywiadach PAPI. Ich odsetek nie jest wysoki, oscyluje wokół 0,5–1%. Wyjątek stanowi zadane w wywiadzie PAPI pytanie dotyczące związków homoseksualnych, w którym odsetek tego typu odpowiedzi wyniósł ponad 7% (przypomnijmy, w CATI: 0,0%, a różnica ta jest istotna statystycznie). Interpretując ten wynik, znowu należy odwołać się do specyfiki rozmowy telefonicznej, w której odpowiedź „nie wiem” może zostać odebrana jako wyraz zniechęcenia czy znudzenia. W bezpośredniej rozmowie taka odpowiedź po „chwili namysłu”, z ewentualnym uzasadnieniem, nie zostanie zinterpretowana w ten sposób.

Wpływ różnic związanych z odmiennością sposobu komunikowania się na związek między zmiennymi

Różnice rozkładów odpowiedzi na poszczególne pytania nie przesądzają o sile związków i ewentualnych zależnościach z innymi zmiennymi. Stąd też należy postawić pytanie o wpływ różnic techniki zbierania danych (odmienności sposobu komunikowania się) na związki z innymi zmiennymi, a więc wynik. Gdyby okazało się, że korelacje między odpowiedziami respondentów na pytanie TX a ich odpowiedziami na pytanie TY w wywiadzie CATI oraz analogicznymi uzyskanymi w wywiadzie PAPI (tzn. X i Y) różnią się między sobą w istotny sposób, to wówczas można zakwestionować trafność wniosków odwołujących się do zależności między tymi zmiennymi.

Dla pytań dychotomicznych wyznaczyliśmy równania regresji logistycznej [deMaris, 1992] i porównaliśmy wartości współczynnika „b”. Dla pytania o udział w wyborach parlamentarnych wprowadziliśmy zestaw 4 zmiennych niezależnych: płeć, wiek, miejsce zamieszkania oraz wykształcenie. Ze względu na niskie liczebności zmiennych niezależnych wprowadzanych do modelu dla danych z badania ESS-CATI, zrekodowaliśmy je do zmiennych dwuwartościowych w następujący sposób: wiek (15–44; 45+ lat), miejsce zamieszkania (miasto; wieś) i wykształcenie (dwie zmienne: (i) średnie i poniżej średniego vs pozostali; (ii) powyżej średniego vs pozostali). W analogiczny sposób postąpiliśmy w odniesieniu do pytania o poczucie związku z partią polityczną, tzn. wprowadziliśmy te same zestawy zmiennych niezależnych: płeć, wiek i miejsce zamieszkania.

Dla pytań o poczucie zaufania i stosunek do osób o orientacji homoseksualnej wyznaczyliśmy równania regresji liniowej i porównaliśmy wartości współczynnika „b”. Dla obu pytań wprowadziliśmy te same zestawy 4 zmiennych: płeć, wiek, miejsce zamieszkania oraz wykształcenie¹⁷. Są one zgodne pod względem konstrukcji, tzn. zmienne zostały odpowiednio zrekodowane, tak, by w obu zbiorach przyjmowały te same wartości. Wyniki analizy zamieszczono w tabeli 8.

¹⁷ Mówiąc o wykształceniu, mamy na myśli poziom wykształcenia określony wg kategorii ISCED, ustalony na podstawie pytania F6/TF6 (9 kategorii).

TABELA 8. Wartości metrycznych współczynników regresji logistycznej dla pytań TB11/B11 i TB20a/B20a oraz regresji liniowej dla pytań TA8/A8 i TB3 /B31

Pytanie	Zmienna wyjaśniająca	ESS-PAPI		ESS-CATI	
		„b”	istotność	„b”	istotność
Pyt. TB11/B11. Udział w wyborach	płeć	-.019	.865	-.558	.064
	wiek	.535	.000	.905	.002
	miejsce zamieszkania	-.094	.402	.176	.545
	wykształcenie	.889	.000	.538	.064
Pyt. TB20a/B20a. Poczucie związku z partią	płeć	-.513	.000	-.365	.131
	wiek	.257	.031	.647	.010
	miejsce zamieszkania	-.366	.002	-.255	.298
Pyt. TA8/A8. Poczucie zaufania	płeć	-.039	.743	.617	.055
	wiek	-.062	.064	-.440	.174
	miejsce zamieszkania	-.011	.920	.053	.601
	wykształcenie	.293	.000	.158	.105
Pyt. TB31 /B31. Stosunek do osób o orientacji homoseksualnej	płeć	-.017	.766	-.135	.327
	wiek	.120	.000	.532	.000
	miejsce zamieszkania	.213	.000	.019	.657
	wykształcenie	-.138	.000	.018	.667

Porównanie wartości niestandardyzowanych metrycznych współczynników b regresji logistycznej i liniowej wskazuje jedynie, że dla analizowanych pytań odpowiedzi respondentów są inaczej powiązane z podstawowymi zmiennymi społeczno-demograficznymi w badaniu dla ESS-PAPI, a inaczej w badaniu ESS-CATI. Niewątpliwie taki wynik jest konsekwencją dysproporcji wielkości prób w badaniu ESS-PAPI ($N = 1721$) oraz ESS-CATI ($N = 300$) i może mieć wpływ na to co jest istotne a co nie jest. Obie próby są zatem różne i to w takim stopniu, że inne zmienne pozwalają przewidywać wartości zmiennej zależnej. Innymi słowy, w oparciu o wyniki tak samo skonstruowanych równań regresji w każdym z badań otrzymujemy odmienne układy zależności i w konsekwencji uzyskujemy różne wyniki.

ESS-PAPI/CATI MIXED MODE: ZA I PRZECIW

Możliwości i warunki prowadzenia badań surveyowych na próbach losowych ulegają coraz większym ograniczeniom, stają się coraz trudniejsze. Związane jest to przede wszystkim ze wzrostem wiedzy, a właściwie uproszczonych wyobrażeń o badaniach, wzrostem podmiotowości i poczucia prywatności respondentów, ale także wzrostem poczucia zagrożenia bezpieczeństwa osobistego. Nie bez

znaczenia jest także obiektywny wzrost „obciążeń i zobowiązań” zawodowych [Sztabiński F., Sztabiński P.B. 1994: 141–164, Sztabiński P.B. et al. 2008: 39–84]. Wszystko to skutkuje obniżającym się odsetkiem realizacji próby i błędem braku odpowiedzi (*unit nonresponse*). Obecnie, w standardowych badaniach konsumenckich i opinii publicznej, współczynnik realizacji (przy zastosowaniu PAPI/CAPI) kształtuje się w granicach 20–40%. W tej sytuacji *Mixed Mode Survey Design* (rozumiany jako procedura zbierania tych samych informacji różnymi technikami od różnych osób) może się stać swoistym remedium. Zakłada się tutaj, że wylosowane osoby mają pewne preferencje dotyczące formy kontaktu i sposobu komunikowania, pewne techniki są – o czym wspominaliśmy – bardziej lub mniej adekwatne ze względu na społeczne i psychologiczne charakterystyki badanych.

Jeśli przyjmiemy, że *Mixed Mode Survey Design* może redukować błąd braku odpowiedzi i błąd pokrycia, to musimy także przyjąć, iż może generować błąd pomiaru, związany z techniką (*Mode Effects*). Analizując wyniki badania ESS-PAPI oraz ESS-CATI r.3 dla Polski (ale także dla innych krajów), potwierdzamy, że tak jest.

Po pierwsze, odsetek realizacji wywiadu ESS-CATI (niezależnie od wersji) w niektórych krajach jest blisko o połowę niższy niż przy zastosowaniu PAPI (Polska: 35 vs 70%) czy CAPI (Niemcy: 53 vs 24%). W przypadku innych krajów jest on znacznie niższy: Węgry, Cypr (odpowiednio: 65,9 vs 25,2% oraz: 67,3 vs 8,3%). Wyjątek stanowi Szwajcaria, która w ESS-CAPI uzyskała odsetek realizacji 50%, a w ESS-CATI nieco niższy – 41,5%. Wynik ten zdaje się wskazywać na problem różnego poziomu akulturacji technik. Pierwszym podstawowym wnioskiem jest zatem to, iż CATI, biorąc pod uwagę sposób kontaktowania się z respondentem, nie stanowi „konkurencji” dla PAPI (przynajmniej w Polsce), ale także dla CAPI (np. Niemcy), co potwierdza występowanie efektu techniki, przynajmniej w sferze kontaktu.

Po drugie, analizując uzyskane rezultaty (rozkłady odpowiedzi) zauważyć można, iż różnią się one między sobą. W większości przypadków są to różnice istotne statystycznie i zależą nie tylko od struktury pytania, lecz także od jego przedmiotu. W tym przypadku nie sposób określić, które z uzyskanych wyników (CATI czy PAPI) są „prawdziwe”, nie o to zresztą chodzi. Następny podstawowy wniosek jest taki, że CATI przynosi odmienne wyniki (chodzi o odpowiedzi respondentów) w porównaniu do PAPI, przynajmniej w Polsce. Potwierdza to wystąpienie efektu techniki w sferze komunikacji.

Po trzecie, porównując wnioski, które można sformułować na podstawie badania PAPI i CATI (chodzi o związki między zmiennymi), znowu trzeba uznać,

iż są one odmienne. Jeśli bierzemy pod uwagę poszczególne cechy społeczno-demograficzne (płeć, wiek itd.), to ich związek z niektórymi zmiennymi zależnymi jest podobny. Nie zmienia to ogólnego wniosku dotyczącego wystąpienia efektu techniki.

Sformułowane wnioski mają oczywiście bardzo ograniczony charakter, głównie ze względu na specyfikę zebranego materiału, który jest pochodną celów i założeń przyjętych w projekcie JRA1. Ograniczenia dotyczą przede wszystkim wielkości próby założonej do badania ESS-CATI (próba N = 1000 realizowana do wyczerpania, w podziale na 3 podpróby, w wyniku czego podstawy procentowania były bardzo niskie).

Czy zatem zastosowanie *Mixed Mode Survey Design* redukuje błędy pokrycia, braku odpowiedzi i zwiększa błąd pomiaru, głównie za sprawą efektu techniki? Wydaje się, że procedura ta może redukować te błędy, ale . . . W chwili obecnej wszystkie pytania dotyczące *Mixed Mode Survey Design* wydają się pozostawać bez jednoznacznej odpowiedzi. Nie jest on gotowym „przepisem”, umożliwiającym redukcję błędów braku odpowiedzi. Musimy zawsze brać pod uwagę czasowe modyfikacje społecznej sytuacji badania (choćby wspomniany poziom penetracji telefonów stacjonarnych i komórkowych).

Na obecnym etapie refleksji *Mixed Mode Survey Design* wydaje się raczej zbiorem racjonalnych postulatów, a nie dopracowaną procedurą badawczą. Nie jest to prosty *trade-off*, lecz raczej *multiple trade-off*. Co więcej, za celowe można uznać rozszerzenie o wymiar „czasowy” „paradygmatu” myślenia o *Mixed Mode Survey Design*. Czy „czas” (długość okresu realizacji) stanowi ograniczenie i jak znaczące w odniesieniu do kosztów badania? Warto postawić dodatkowe pytanie: czy badanie sekwencyjne *Mixed Mode* rozpoczynać od techniki „najszybszej” czy „najtańszej”? Ale, to wszystko zależy . . .

BIBLIOGRAFIA

- American Association for Public Opinion Research [2011], <http://aapor.org/Content/StandardDefinitions.pdf> [dostęp 5.08.2011].
- Bassler A., Tchaoussoglou C. [2001], *Searching for best CASI practices: experiments for a new data collection method*, 10th Worldwide Readership Research Symposium, Venice.
- Biemer P., Caspar R. [1994], *Continuous Quality Improvement for Survey Operations: Some General Principles and Applications*, „Journal of Official Statistics”, vol. 10, no. 3.
- Bowling A. [2005], *Mode of questionnaire administration can have serious effects on data quality*, „Journal of Public Health”, vol. 27.

- Cichomski B., Jerzyński T., Zieliński M. [2009], *Polskie Generalne Sondaże Społeczne: struktura skumulowanych wyników badań 1992–2008*, Warszawa: Wydawnictwo ISS UW.
- Czaja R., Blair J. [2005], *Designing surveys: A guide to decisions and procedures*, Thousand Oaks, CA: Sage Publications.
- de Leeuw E.D. [2008], *Choosing the method of data collection*, [w:] E.D. de Leeuw, J. Hox, D.A. Dillman (eds), *International Handbook of Survey Methodology*, New York–London: Lawrence Erlbaum Associates.
- de Leeuw E.D. [2009], *Mixed Mode Surveys – 6th ESS Training Course, 30 April–1 May*, Ljubljana, Slovenia.
- de Leeuw E.D., de Heer W. [2002], *Trends in Household Survey Nonresponse: A Longitudinal and International Comparison*, [w:] R.M. Groves, D.A. Dillman, J. L. Eltinge, R.J.A. Little (eds), *Survey Nonresponse*, New York: John Wiley & Sons.
- de Leeuw E.D., Hox J.J., Dillman D.A., [2008a], *The Cornerstones of Survey Research*, [w:] E.D. de Leeuw, J.J. Hox, D.A. Dillman (eds), *International Handbook of Survey Methodology*, New York–London: Lawrence Erlbaum Associates.
- de Leeuw E.D., Hox J.J., Dillman D.A., [2008b], *Mixed Mode Surveys: When and Why*, [w:] E.D. de Leeuw, J.J. Hox, D.A. Dillman (eds), *International Handbook of Survey Methodology*, New York–London: Lawrence Erlbaum Associates.
- de Maris A. [1992], *Logit Modeling. Practical Applications*, Thousand Oaks, CA: Sage Publications.
- Dillman D.A. [1978], *Mail and Telephone Surveys: The Total Design Method*, New York: John Wiley & Sons.
- Dillman D.A. [2000], *Mail and Internet Surveys: The Tailored Design Method*, New York: John Wiley & Sons.
- Dillman D.A., Smith J., Christian L. [2009], *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method*, New York: John Wiley & Sons.
- European Social Survey, Round 5. Specification for participating countries [2011], <http://www.europeansocialsurvey.org> [dostęp 5.08.2011].
- Eva G., Widdop S. [2007], *Mixed Mode Data Collection in Europe*, 2nd European Survey Research Association Conference, Prague, June.
- Gostkowski Z. [1965], *Badania adekwatności technik badawczych w krajach na drodze rozwoju*, „Przegląd Socjologiczny”, t. XIX/2.
- Gostkowski Z. [1966], *O założeniach i potrzebie empirycznych badań nad technikami i procedurami badawczymi w socjologii*, [w:] Z. Gostkowski (red.), *Analizy i próby technik badawczych w socjologii*, t. 1, Wrocław: Zakład Narodowy im. Ossolińskich.
- Goyder J.C. [1987], *The Silent Minority*, Boulder, CO: Westview Press.
- Groves R.M. [1989], *Survey Errors and Survey Costs*, New York: John Wiley & Sons.
- Groves R.M. [1991], *Measurement Error Across the Disciplines*, [w:] P.P. Biemer, R.M. Goves, L.E. Lyberg, N. A. Mathiowetz, S. Sudman (eds), *Measurement Errors in Surveys*, New York: John Wiley & Sons.
- Groves R.M., Couper M. P. [1998], *Nonresponse in household interview surveys*, New York: John Wiley & Sons.
- Groves R.M., McGonagle K. [2001], *A theory-guided interviewer training protocol regarding survey participation*, “Journal of Official Statistics”, vol. 17.

- Groves R.M., Fowler F.J. Jr., Couper M.P., Lepkowski J.M., Singer E., Tourangeau R. [2004], *Survey Methodology*, Hoboken, New Jersey: John Wiley & Sons, Inc.
- Grzeszkiewicz-Radulska K. [2009], *Respondenci niedostępni w badaniach sondażowych*, [w:] Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. 12, Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Haeder S., Lehnhoff I., Mardian I. [2010], *Mobile Phone Surveys: Empirical Findings from a Research Project*, "ASK. Research and Methods", vol. 19.
- Lutyńska K. [1978], *Ankieterzy i badacze. Z badań nad wpływem ankieterskim*, „Przegląd Socjologiczny”, t. XXX.
- Lutyński J. [1968], *Ankieta i jej rodzaje na tle podziału technik otrzymywania materiałów*, [w:] Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. 2, Wrocław: Zakład Narodowy im. Ossolińskich.
- Lutyński J. [1972], *Analiza procesu otrzymywania informacji w badaniach z zastosowaniem wywiadu kwestionariuszowego*, [w:] Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. 4, Wrocław: Zakład Narodowy im. Ossolińskich.
- Lynn P. [2009], *Mixed or Muddled? Combining Survey Modes in The 21st Century*, *Keynote Presentation*, 3rd European Survey Research Association Conference, Warsaw, June.
- Martin P. [2010], *Podejście mieszane (Mixed Mode) – możliwości i ograniczenia*, <http://www.konferencja-ewaluacja.pl/2010/materialy.php> [dostęp 5.08.2011].
- Parten M. [1950], *Surveys, Polls and Samples*, New York: Harper & Brothers.
- Roberts C. [2007], *Mixing Modes of Data Collection in Surveys*, 2nd European Survey Research Association Conference, Prague, June.
- Roberts C., Eva G., Lynn P., Johnson J. [2010], *Measuring the effect of interview length on response propensity and response quality in a telephone survey – Final report of the ESS CATI experiment*, Centre for Comparative Social Surveys, City University, London.
- Saris W.E. [1991], *Computer-Assisted Interviewing*, Newbury Park: Sage Publications.
- Słomczyński K.M. [1968], *Granice stosowalności ankiety audytoryjnej w środowisku robotniczym*, [w:] Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. 2, Wrocław: Zakład Narodowy im. Ossolińskich.
- Stoop I. [2005], *The Hunt for the Last Respondent: Nonresponse in Sample Surveys*, The Hague: Social and Cultural Planning Office of the Netherlands.
- Stoop I., Billiet J., Koch A., Fitzgerald R. [2010], *Improving Survey Response. Lessons learned from the European Social Survey*, New York: John Wiley & Sons.
- Sztabiński F. [1997], *Ankieta pocztowa i wywiad kwestionariuszowy*, Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński F. [2003], *Logika badacza i logika respondenta. Problem adekwatności narzędzia badawczego*, „ASK. Społeczeństwo, Badania, Metody”, vol. 12.
- Sztabiński F. [2009], *Ankieta internetowa (CAWI): czy rzeczywiście idealna technika?*, [w:] A. Haber, M. Szałaj (red.), *Ewaluacja wobec wyzwań stojących przed sektorem finansów publicznych*, Warszawa: PARP.
- Sztabiński F. [2011], *Ocena jakości danych w badaniach surveyowych*, Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński F., Sztabiński P.B. [1994], *Polscy respondenci początku lat 90. Stosunek do badań i jego uwarunkowanie*, „Studia Socjologiczne”, nr 2.

- Sztabiński F., Sztabiński P.B. [1997], *Wartość odpowiedzi w wywiadach telefonicznych ze wspomaganie komputerowym (CATI)*, „ASK. Społeczeństwo, Badania, Metody”, vol. 1–2 (5–6).
- Sztabiński P.B. [1997], *Ankieterzy i ich respondenci. Od kogo zależą wyniki badań ankietowych?*, Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński P.B. [1999a], *Wywiad telefoniczny ze wspomaganie 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*, Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński P.B. [1999b], *Wywiad telefoniczny ze wspomaganie komputerowym (CATI): co zyskujemy, co tracimy?*, „ASK. Społeczeństwo, Badania, Metody”, vol. 8.
- Sztabiński P.B. [2001], *Wywiad osobisty ze wspomaganie komputerowym (CAPI): czy rzeczywistość idealna technika?*, „ASK. Społeczeństwo, Badania, Metody”, vol. 10.
- Sztabiński P.B., Dyjas-Pokorska A., Żmijewska-Jędrzejczyk T. [2008], *Understanding Refusals*, „ASK. Research and Methods”, vol. 17.
- Sztabiński P.B., Sztabiński F., Przybysz D. [2007], *Are non-respondents similar to respondents? Findings from the ESS-2004 in Poland*. „ASK. Społeczeństwo, Badania, Metody”, vol. 16.
- Weisberg H.F. [2005], *The Total Survey Error Approach. A Guide to the New Science of Survey Research*, Chicago–London: The University of Chicago Press.
- Żmijewska-Jędrzejczyk T. [2004], *Badania internetowe*, [w:] P.B. Sztabiński, F. Sztabiński, Z. Sawiński (red.), *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Warszawa: Wydawnictwo IFiS PAN.

Aneks

TABELA I. Schemat badania ESS-CATI r. 3

Grupa	Wersja kwestionariusza/moduły						Czas	Próba (%)		
Grupa A	Pełna wersja kwestionariusza ESS (wersja A)						≈ 60'	n = 400/40		
	moduły:	A	B	C	D	E			F	
Grupa B	Skrócona wersja kwestionariusza ESS (wersja B)						≈ 45'	n = 200/20		
	moduły:	A	B	C		E			F	
Grupa C	Pełna wersja kwestionariusza ESS (wersja C) realizowana w dwóch częściach: a i b						2 × ≈ 30'	n = 400/40		
	Część a. moduły:	A	B		D				Fx	Moduł Fx (24 pyt.)
	Część b. moduły:			C		E			Fy	Moduł Fy (52 pyt.)

Moduły:

- A – Media, zaufanie do ludzi
- B – Zainteresowania polityką, zaufanie do instytucji, udział w wyborach, orientacje polityczne i życiowe

- C – Poczucie dobrobytu i wykluczenia społecznego, przynależność religijna, narodowa i etniczna
- D – Kluczowe wydarzenia życiowe, postawy wobec najniższego i najwyższego progu wydarzeń życiowych
- E – Dobrobyt osobisty i społeczny, pomoc innym, odczucia z ubiegłego tygodnia, zadowolenie z życia i pracy
- F (x, y) – Charakterystyki społeczno-demograficzne respondenta i partnera.

Franciszek Sztabiński
Teresa Żmijewska-Jędrzejczyk

MIXED MODE SURVEY DESIGN: MODE EFFECT PROBLEM

Abstract

The fieldwork conditions for surveys becomes progressively worse so methods are being sought to reduce the *total survey error* arising from *non-response error*. One solution is the *mixed mode survey design* procedure whereby the same information is obtained from different respondents, using different data collecting modes (e.g. PAPI, CATI, Web survey). However, this procedure may produce the so-called *mode effect*. In this paper, we analyse the mode effect from the following perspectives: (i) wording of questions, (ii) contacting mode, and (iii) mode of communication with the respondent. We assess response rates and compare respondents' answers to the same questions obtained through different techniques.

Our analysis is based on materials collected in research projects employing two modes PAPI (the European Social Survey, Round 3) and CATI (ESS Infrastructure: i3, a study funded under a methodological grant). The findings from our analysis confirm the existence of significant differences related to the *mode effect* (in the case of PAPI and CAPI). Those findings challenge the idea of combining those modes to reduce *non-response error*.

Key words: Total Survey Error, Non-Response Error, Mode Effect, Mixed Mode Survey Design, European Social Survey