

Dr hab. Anna Kiersztyn, prof. UW
Wydział Socjologii
Uniwersytet Warszawski

Recenzja rozprawy doktorskiej mgr Adama Rybaka pt. „, Praktyki sondażowe na świecie i w Polsce: dokumentacja, reprezentatywność, preferencja techniki”

Przedstawiona do recenzji rozprawa obejmuje spójny cykl czterech artykułów naukowych, których autorem w całości jest mgr Rybak:

1. „Analiza jakości dokumentacji metodologicznej sondaży: Studium przypadku inwentaryzacji 30 programów badań porównawczych z całego świata”, artykuł opublikowany w roku 2025 w piśmie Ruch Prawniczy, Ekonomiczny i Socjologiczny;
2. „Survey Mode and Nonresponse Bias: A Meta-Analysis Based on the Data from the International Social Survey Programme Waves 1996–2018 and the European Social Survey Rounds 1 to 9”, artykuł opublikowany w roku 2023 w PLOS ONE;
3. „Komunikacja oparta na kanale wizualnym i audialnym a jakość pomiaru sondażowego. W stronę nowego ujęcia hipotezy preferencji techniki przy wykorzystaniu koncepcji pamięci roboczej”, artykuł opublikowany w roku 2019 Przeglądzie Socjologicznym;
4. „Declarative and Behavioral Data in Predicting Respondents Survey Mode Preference”, artykuł opublikowany w roku 2020 w Polish Sociological Review.

W swoim doktoracie mgr Adam Rybak podejmuje ważną i aktualną problematykę z dziedziny metodologii surveyowej. Dotyczy ona kwestii związanych z jakością i wiarygodnością danych w obliczu współczesnego kryzysu badań sondażowych, przejawiającego się malejącymi response rates, przy jednoczesnym wzroście kosztów realizacji badań. Efektem tego stanu rzeczy jest stopniowe odchodzenie od badań z zastosowaniem najbardziej kosztownej metody wywiadu kwestionariuszowego z osobistą wizytą ankietera (w szczególności PAPI oraz CAPI), na rzecz ankiet pocztowych, internetowych czy metod mieszanych. Powyższe zmiany stanowią nie tylko wyzwanie organizacyjne, ale także rodzą konieczność systematycznych badań konsekwencji stosowania zróżnicowanych metod dla jakości pozyskanych danych – w szczególności możliwych błędów wynikających z niereprezentatywności badanych prób – a także możliwych czynników moderujących powyższe zależności. Ponadto rosnący koszt realizacji badań, przy jednoczesnym wzroście zróżnicowania stosowanych metod sprawia, że coraz większego znaczenia nabiera jakość dokumentacji poszczególnych sondaży. Jasna i kompletna informacja metodologiczna umożliwia bowiem wykorzystanie danych przez kolejnych badaczy i podnosi jakość analiz przeprowadzanych z użyciem tych danych.

Recenzowany cykl skupia się na trzech powiązanych ze sobą zagadnieniach, i dla każdego z nich wnosi nowy wkład do istniejącej wiedzy. Pierwsze pytanie badawcze dotyczy zróżnicowań praktyk związanych z dokumentacją sondaży porównawczych na świecie w ostatnich dziesięcioleciach. Uwaga Autora skupia się tu na różnicach w jakości i kompletności

dokumentacji dotyczącej poszczególnych charakterystyk sondażu w czasie i przestrzeni, a także między poszczególnymi programami sondażowymi. Drugie pytanie dotyczy sposobów przeprowadzania sondaży w różnych częściach świata – a dokładniej zróżnicowania technik zbierania danych. Powyższe zagadnienia są przedmiotem pierwszej publikacji; Autor odnosi się do nich ponadto w autoreferacie i aneksie do rozprawy. W odróżnieniu od dwóch pierwszych zagadnień, których wkład do literatury z zakresu metodologii badań surveyowych polega przede wszystkim na systematyzacji istniejących praktyk, trzecie ma charakter analityczny i skupia się na badaniu zależności między charakterystykami sondaży a jakością pomiaru, rozumianą przede wszystkim w kategoriach wewnętrznej reprezentatywności próby. Wybrane aspekty tego szerokiego zagadnienia są przedmiotem rozważań w kolejnych trzech publikacjach wchodzących w skład cyklu.

Prace w ramach cyklu są powiązane tematycznie, a zarazem wykorzystują zróżnicowane źródła danych i różne podejścia metodologiczne. Artykuły nie są uporządkowane chronologicznie, jako że dwie ostatnie publikacje z cyklu opublikowane zostały najwcześniej; niemniej kolejność ich ujęcia przez Autora odzwierciedla logikę jego analiz. Dwa pierwsze artykuły stanowią rezultat projektu PRELUDIUM zrealizowanego przez Doktoranta. W ramach projektu mgr Rybak podjął się ambitnego i pracochłonnego przedsięwzięcia „*inwentaryzacji praktyk sondażowych z programów porównawczych, które obejmowały wszystkie zamieszkane kontynenty i zostały przeprowadzone w ciągu ostatnich 65 lat*” w oparciu o publicznie dostępną dokumentację i bazy danych wynikowych. Samo to osiągnięcie zasługuje na uznanie, ponieważ jest to obecnie największa baza danych tego typu, jeśli chodzi o liczbę uwzględnionych badań (zawiera informację o 30 programach sondażowych obejmujących niemal 10,5 tys. sondaży realizowanych na całym świecie). Tym bardziej wartościowe jest to, iż baza została publicznie udostępniona online do użytku przez kolejnych badaczy. Informacje zgromadzone w wyniku inwentaryzacji Autor wykorzystał w dwojaki sposób. Po pierwsze, do oceny jakości i kompletności dokumentacji sondaży objętych badaniem i jej zgodności ze standardami *Data Documentation Initiative* (Publikacja nr 1). Po drugie, jako źródło danych do metaanalizy badającej związku między metodą badania a miarami reprezentatywności sondażu (Publikacja nr 2). Dodatkowe informacje o procedurze inwentaryzacji praktyk sondażowych na świecie przedstawione zostały w autoreferacie i aneksie, które w tym zakresie stanowią ważne uzupełnienie przedstawionego cyklu.

Choć w ostatnich latach przeprowadzono szereg badań dotyczących wpływu metody realizacji badań sondażowych na *response rates* oraz miary reprezentatywności, literatura przedmiotu nadal posiada luki; brakuje w szczególności systematycznych analiz porównawczych w oparciu o dane z różnych krajów. Problemem jest ponadto niespójność wyników wcześniejszych analiz, utrudniająca formułowanie wiarygodnych rekomendacji metodologicznych. Tą ważną lukę wypełnia analiza mgr Rybaka, dzięki zastosowanej metaanalizie danych dla dwóch międzynarodowych programów badań surveyowych, w przypadku których rzetelność prowadzonej dokumentacji umożliwiła testowanie hipotez dotyczących związków między kluczowymi charakterystykami badania a miarami reprezentatywności próby. Na uznanie zasługuje przemyślany i dobrze uzasadniony dobór sondaży do analizy, w oparciu o wnioski uzyskane w wyniku inwentaryzacji praktyk sondażowych i przedstawione w pierwszej publikacji z cyklu. Analizy interakcji między

operatem doboru próby a techniką sondażu są nowością w literaturze, umożliwiającą sformułowanie oryginalnych konkluzji dotyczących warunków, w jakich odejście od wywiadów kwestionariuszowych na rzecz ankiet pocztowych bądź internetowych (bądź ich połączenia) nie będzie miało miejsca kosztem pogorszenia jakości uzyskanych danych.

O ile dwa pierwsze (i zarazem nowsze) artykuły w ramach cyklu skupiają się na analizach zinstytucjonalizowanych praktyk sondażowych z całego świata, o tyle analizy przedstawione w trzecim i czwartym artykule analizują zagadnienie jakości danych sondażowych z perspektywy indywidualnych respondentów, uwzględniając w szczególności kwestię ich preferencji wobec określonej techniki sondażu i jej znaczenia dla minimalizacji błędów braku odpowiedzi i błędów pomiaru. Pierwsza z tych publikacji ma charakter teoretyczny i proponuje nowatorską conceptualizację czynników determinujących oba rodzaje błędów w oparciu o przegląd literatury dotyczącej zagadnienia preferowanej techniki sondażu (*mode preference*), minimalizacji wysiłku poznawczego (*satisficing*) i zdolności przetwarzania informacji (psychologiczna koncepcja pamięci roboczej). Ważną zaletą tej analizy jest jej interdyscyplinarność: wypada zgodzić się bowiem z tezą Autora, że wyzwania stojące współcześnie przed badaniami sondażowymi wymagają sięgnięcia do dorobku nie tylko metodologii, ale także demografii i psychologii poznawczej. Ostatni artykuł prezentuje wyniki analiz ilościowych polskich danych z 7 i 8 rundy ESS, dotyczących społeczno-demograficznych korelatów preferencji wobec określonej techniki sondażu. Stanowi on kontynuację rozważań i realizację postulatów przedstawionych w publikacji nr 3. Obok walorów praktycznych, za istotną i wartościową poznawczo uważam przedstawioną w artykule analizę rozbieżności między miarami preferencji techniki sondażu opartymi na deklaracjach i faktycznych wyborach respondentów.

Podsumowując wyniki dysertacji, mgr Rybak zwraca uwagę na następujące kwestie (abstrakt, s. 7): *1) Potrzeba wypracowania jednolitego standardu dokumentacji metodologicznej pozwalającego na automatyzację procesu zbierania danych metodologicznych. 2) Znaczenie dostępności operatu imiennego w sytuacji stopniowego odchodzenia od wywiadów kwestionariuszowych. 3) Potencjał sondaży mixed mode do uzyskiwania prób o satysfakcjonującym poziomie reprezentatywności. 4) Zasadność badania źródeł i korelatów preferencji techniki sondażu w celu tworzenia modeli predykcyjnych pozwalających na planowanie przyszłych badań mixed mode w schemacie równoległym w poszczególnych krajach.* Prace przedstawione w ramach cyklu dostarczają zarówno ram pojęciowych, jak i materiału empirycznego, który pozwala przekonywująco argumentować na rzecz każdego z powyższych postulatów. Tego typu wiedza ma również znaczenie dla praktyki badań sondażowych – sprzyja podnoszeniu jakości i opracowaniu strategii, które pozwoliłyby lepiej dopasować stosowane metody i techniki surveyowe do zmieniających się realiów życia społecznego.

Na podkreślenie zasługuje ponadto, iż wszystkie publikacje włączone do cyklu ukazały się w renomowanych pismach naukowych o zasięgu ogólnopolskim oraz międzynarodowym; w tym dwie napisane zostały w języku angielskim. Są to pisma uwzględnione na liście czasopism Ministerstwa Nauki i Szkolnictwa Wyższego i z wysoką oceną punktową. Dotyczy to w szczególności PLOS ONE (140 punktów); z kolei pozostałe czasopisma, choć nieco niżej

punktowane, należą do czołówki polskich czasopism socjologicznych. Oznacza to zarazem, że każdy z elementów przedstawionego dorobku przeszedł już rygorystyczny proces recenzji. Przedstawione poniżej uwagi krytyczne nie podważają jakości ani wartości recenzowanej rozprawy – traktuję je jako dodatkowy materiał do przemyśleń i dalszych dyskusji oraz źródło wskazówek na przyszłość – w tym również przez zwrócenie uwagi na problemy, które mogą stanowić przedmiot kolejnych przedsięwzięć badawczych Doktoranta.

Uwagi do pytań badawczych

Na początek odnoszę się do sposobu sformułowania pytań badawczych w Autoreferacie, w szczególności listy szczegółowych zagadnień objętych analizą. Nie jest jasne, w jaki sposób i na jakiej podstawie zagadnienia te wywodzone są z pytań ogólnych. Trudno oprzeć się wrażeniu, że pytania zostały sformułowane *ad hoc*, tak aby pasowały do zastosowanych danych i tez postawionych w publikacjach wchodzących w skład cyklu. Lepsza byłaby tu odwrotna logika – sformułowanie pytań w oparciu o zidentyfikowane wcześniej luki badawcze, a następnie odniesienie wniosków z publikacji do tak sformułowanych pytań. Nie zakładam przy tym, że publikacje powinny w pełni odpowiadać na postawione pytania – takie oczekiwanie byłoby nierealistyczne choćby z uwagi na ograniczenia zastosowanych danych. Dzięki takiej konceptualizacji Autoreferat pozwoliłby nie tylko jaśniej podsumować, w jaki sposób wyniki badań wypełniają luki w literaturze, lecz także zidentyfikować zagadnienia wymagające dalszych analiz. Dobry przykład stanowi tu pytanie P3a, w którym mowa o *sondażach opartych o techniki ankiety i sondażach mixed-mode prowadzonych bez obecności ankietera* (s. 12). Sformułowanie to odzwierciedla wprost kategoryzację technik sondażowych przyjętą dla potrzeb analiz w publikacji nr 2, będącą z kolei pochodną sposobów gromadzenia danych występujących w badanych programach surveyowych (ESS i ISSP). Kategoryzacja ta nie uwzględnia wszystkich dostępnych opcji badań *single mode* (np. CATI czy ankiety internetowe) i grupuje razem różne podejścia w ramach strategii *mixed-mode*. Powyższe braki nie przekreślają znaczenia wyników przedstawionych w publikacji, które przynoszą częściową odpowiedź na pytania o związki między techniką sondażu a realizacją próby i jej reprezentatywnością.

Druga kwestia dotyczy sposobu porządkowania pytań szczegółowych. Ich lista jest długa, ale łączy pytania opisowe lub wartościujące z takimi, które dotyczą zależności poddających się empirycznej weryfikacji. Obok pytań, które wpisują się w dążenie do ewaluacji praktyk sondażowych (np. P1a: *Jakie charakterystyki sondaży są najlepiej i najslabiej udokumentowane w raportach z projektów międzykrajowych?*; P1b: *Czy rozwój standardów dokumentowania sondaży jest adekwatny do poziomu możliwości technicznych?*; P2b: *Które programy międzykrajowe zapewniają najlepszą jakość pomiaru sondażowego?*) pojawiają się pytania czysto opisowe (np. P1d: *Jakie są międzykrajowe zróżnicowania kompletności dokumentacji?*; P1e: *Jakie są różnice w jakości dokumentacji pomiędzy programami sondażowymi?*), po których z kolei następują pytania o konkretne zależności (np. wspomniane wyżej P3a: *Czy sondaże oparte o techniki ankiety i sondaże mixed-mode prowadzone bez obecności ankietera uzyskują niższe wskaźniki realizacji próby, niż wywiady kwestionariuszowe?*), przeplatające się z pytaniami o charakterze teoretycznym (np. P3d: *Czy koncepcja satisficingu wyjaśnia problemy związane z preferencją techniki sondażu?*). Brak

rozdzielenia pytań odwołujących się do różnych porządków analizy nie stanowi sam w sobie błędu, ale wprowadza niepotrzebny bałagan do opisu przeprowadzonych analiz.

Ponadto same pytania są sformułowane mało precyzyjnie. Przykładowo, w P1a nie wiadomo dokładnie, co oznaczają pojęcia „najlepiej” i „najsłabiej” w odniesieniu do dokumentacji praktyk surveyowych – zaś zastosowanie w pytaniu dychotomicznego rozróżnienia może sprawiać wrażenie upraszczającego podejścia do analizowanych zagadnień. Z kolei w P1b nie jest jasne, o jakie możliwości techniczne chodzi, jakie standardy można uznać za spełniające warunek adekwatności i dlaczego. W pytaniu P1c występuje odwołanie do analizowanego obszaru w momencie, w którym czytelnik nie wie jeszcze, jaki obszar jest poddany analizie. Pytania P1d i P1e częściowo nakładają się na siebie.

Jak słusznie zauważa Autor, zagadnienie ujęte w pytaniu P2 (o sposoby prowadzenia sondaży w różnych krajach na przestrzeni lat) nie zostało w pełni poddane analizie w publikacjach uwzględnionych w cyklu. Tu też wątpliwości budzi przede wszystkim dobór pytań szczegółowych, z których pierwsze (P2a: *Które sondażowe programy międzynarodowe zapewniają zróżnicowanie charakterystyk umożliwiające pożądane analizy metodologiczne?*) wydaje się bardziej odnosić do uzasadnienia doboru programów do analiz metodologicznych z artykułu 2, zaś drugie (P2b: *Które programy międzynarodowe zapewniają najlepszą jakość pomiaru sondażowego?*) nie jest bezpośrednio powiązane z kwestią praktyk. Te ostatnie bowiem różnią się w ramach jednego programu w czasie i przestrzeni, co ma znaczenie dla jakości pomiaru – i co zresztą Autor przekonywująco demonstruje w swoim doktoracie. Zawarte w P2b odniesienie do jakości pomiaru bardziej pasuje do pytania P3 niż P2.

W przypadku zagadnień zgrupowanych w pytaniu P3 moje zastrzeżenia dotyczą zwłaszcza pytań P3d i P3e, które według mnie są sformułowane zbyt ogólnikowo. W przypadku pierwszego z tych pytań (Czy koncepcja *satisficingu* wyjaśnia problemy związane z preferencją techniki sondażu?) nie jest jasne, o jakie wyjaśnienie i jakie problemy chodzi. Dopiero po lekturze artykułu 3 staje się jasne, że *satisficing* może stanowić jedną z potencjalnych konsekwencji braku dopasowania techniki do preferencji. Drugie pytanie (P3e: *Czy kognitywistyczna koncepcja pamięci roboczej może tłumaczyć zmienność preferencji techniki sondażu wśród respondentów?*) jest ciekawe, ale nie zostało poddane empirycznej weryfikacji w pracy. Na poziomie teoretycznym odpowiedź sprowadza się do truizmu – koncepcja w jakimś stopniu „może tłumaczyć”, ale trudno jest bez empirycznych analiz ustalić względne znaczenie pamięci roboczej na tle innych czynników (więcej na ten temat w uwagach do publikacji 3).

Uwagi do poszczególnych publikacji w ramach cyklu

Jak wspomniałam wyżej, **artykuł 1** ma przede wszystkim charakter opisowy i skupia się na ocenie kompletności dokumentacji międzynarodowych programów surveyowych. Choć publikacja ta nie budzi większych zastrzeżeń merytorycznych, uważam, że Autor powinien w niej bardziej systematycznej refleksji poddać kwestię doboru studiów do analiz, a także konsekwencji wyboru określonego zestawu surveyów dla ogólnych wyników porównań. Skoro przedmiotem rozważań jest „inwentaryzacja praktyk sondażowych”, to pewne wątpliwości może budzić skupienie się wyłącznie na międzynarodowych programach badawczych, z pominięciem uznanych krajowych programów sondażowych realizowanych na próbach

reprezentatywnych. Tym bardziej, że uwzględnienie badań tego typu mogłoby znacząco zmienić obraz zróżnicowań między kontynentami.

Sam dobór studiów do badań dokonany przez Autora został jasno opisany w pracy i jest w moim przekonaniu uzasadniony. Przemawiają za nim dość oczywiste względy praktyczne (uwzględnienie programów krajowych znacząco skomplikowałoby już i tak złożone i ambitne przedsięwzięcie, a także wymagałoby nakładów przekraczających możliwości pojedynczego grantu PRELUDIUM). Ponadto baza danych będąca podstawą prezentowanych w artykule rozważań powstała jako materiał do dalszych analiz metodologicznych; jej głównym celem nie była inwentaryzacja jako taka, co także zostało przez Autora podkreślone w artykule. Ponieważ jednak sam Doktorant za główny wkład artykułu do literatury przedmiotu uznaje stwierdzenie niższej jakości dokumentacji sondaży w Europie, w porównaniu z pozostałymi kontynentami – obserwacja odwrotna do opisanej we wcześniejszych badaniach – jestem zdania, że bardziej pogłębiony namysł nad efektem doboru badań do zestawienia dla wyników prostych porównań uczyniłby artykuł bardziej przekonującym.

W krótkim omówieniu prezentowanych porównań międzykontynentalnych mgr Rybak słusznie zwraca uwagę, że na gorszy wynik dla Europy rzutuje przede wszystkim uwzględnienie w zestawieniu badań Eurobarometru. Na podkreślenie zasługuje jednak skala tego efektu. Z tabeli zamieszczonej w artykule wynika, że Eurobarometr jest programem dominującym w badanym zestawieniu: inwentaryzacja obejmuje aż 4604 badania zrealizowane w ramach tego jednego programu na 10493 (stanowiąc zatem 44% wszystkich badań i zdecydowaną większość badań europejskich). Drugi co do wielkości program uwzględniony w zestawieniu, obejmujący aż 2280 badań (22% wszystkich rekordów), to sondaże Gallupa. Dla porównania, choć we wcześniejszej publikacji cytowanej w artykule (Kołczyńska i Schoene 2019), stawiającej tezę o lepszej jakości dokumentacji w Europie, Eurobarometr jest również uwzględniony, w zestawieniu znalazło się jedynie ok. 150 surveyów zrealizowanych w ramach tego programu. Tak wyraźna dominacja Eurobarometru w zestawieniu rodzi dalsze pytanie, o sposób rozumienia zastosowanego w doborze przypadków do inwentaryzacji kryterium niekomercyjnego charakteru badania. Samo to kryterium nie jest stosowane w sposób konsekwentny – nie dotyczyło ono wspomnianych sondaży Gallupa, przy czym Autor nie uzasadnia w tekście tego wyjątku. W przypadku Eurobarometru, choć sam program nie ma charakteru komercyjnego, badania w poszczególnych krajach realizowane były na przestrzeni lat przez komercyjne firmy badawcze zajmujące się badaniami rynku i opinii publicznej. Stoi to w kontraście z takimi programami, jak ESS czy ISSP, realizowanymi przez instytucje akademickie i naukowe, co znalazło przełożenie na wyższe standardy gromadzonej dokumentacji.

Choć nie uważam, aby uwzględnienie sondaży realizowanych – choćby w części – przez komercyjne podmioty było w artykule błędem, w świetle powyższych rozważań jestem zdania, że ogólna konkluzja o niższej jakości dokumentacji sondażowej w Europie została sformułowana zbyt mocno i sama w sobie niewiele wnosi do naszego rozumienia czynników kształtujących jakość dokumentacji. Geograficzne porównania nie wydają się tutaj kwestią najważniejszą, także z uwagi na ich podatność na to, co zostanie uwzględnione w porównaniach; mogą stanowić natomiast punkt wyjścia do dalszych, bardziej pogłębionych

rozważań. Ciekawą poznawczo i potencjalnie ważną dla poprawy standardów praktyk sondażowych ścieżką dalszych analiz byłoby zastanowienie się, jakie czynniki instytucjonalne charakteryzujące poszczególne programy (organizacja badania i poziom jego złożoności, stopień centralizacji, charakter instytucji odpowiedzialnych za realizację badań w krajach objętych projektem, strategię zarządzania projektem, procedury kontroli jakości, sposób finansowania badań i dostęp do zasobów, etc.) kształtują standardy dokumentacji.

Dwie ostatnie uwagi mają charakter bardziej techniczny – w artykule nie został jasno zdefiniowany wskaźnik kompletności dokumentacji: czy jest to udział surveyów, których dokumentacja zawiera odpowiednie informacje? Pewne zastrzeżenia budzi także decyzja o losowym wyborze informacji w sytuacji, gdy w dokumentacji pojawiają się sprzeczne dane – według mnie lepszym rozwiązaniem – także pod kątem wykorzystania bazy danych pochodzących z inwentaryzacji przez kolejnych badaczy – byłoby dodanie zmiennej dychotomicznej (flagi) identyfikującej sondaże, dla których takie niezgodności wystąpiły.

Odnosząc się do treści **Artykułu 2**, zdaniem Autora najważniejsza konkluzja analiz dotyczy większej reprezentatywności badań realizowanych w oparciu o wywiady pocztowe bądź techniką mieszaną bez udziału ankietera, w porównaniu z wywiadem kwestionariuszowym, o ile w badaniu wykorzystany zostanie imienny operat losowania próby. Zarówno w artykule, jak i w autoreferacie mgr Rybak podkreśla znaczenie tych wniosków dla perspektyw rozwoju badań sondażowych, w szczególności możliwości uzyskania reprezentatywnych danych przy niższych kosztach realizacji dzięki odejściu od techniki wywiadu bezpośredniego. Jak pisze w autoreferacie, *wyniki prezentowane w Publikacji 2 mogą napawać optymizmem. Chociaż istnieje względnie rozpowszechnione przekonanie na temat techniki sondażu, że „im drożej, tym lepiej”, prezentowane modele wykazały, że (...) w przypadku stosowania próby imiennej zarówno sondaże pocztowe, jak i mixed mode miały niższe wskaźniki błędu braku odpowiedzi* (s. 44), co skłania go do wniosku, że *alternatywne techniki sondażu mogą dostarczać danych bardzo dobrej jakości i warto z nich korzystać* (s. 45). Nie w pełni podzielam optymizm Doktoranta – w moim przekonaniu do możliwości generalizacji wyników analiz, skądinąd ciekawych i wartościowych, należałoby podejść z większą ostrożnością.

Po pierwsze, opisywany rezultat uzyskany został w oparciu o dane, gdzie duża część sondaży realizowanych techniką ankiety pocztowej i zdecydowana większość realizowanych za pomocą schematu *mixed mode* obejmującego inne techniki niż wywiad face-to-face pochodzi z krajów skandynawskich. Choć tej informacji nie ma bezpośrednio w tekście, w przypadku badań, gdzie dodatkowo zastosowano imienny operat losowania, ta nadreprezentacja jest zapewne jeszcze większa. Zaobserwowana zależność może być zatem przynajmniej w części pochodną specyfiki tych krajów, ich zasobów instytucjonalnych i kultury. Charakterystyki, które hipotetycznie mogą przekładać się na większą reprezentatywność ankiet realizowanych bez udziału ankietera w krajach nordyckich, to wyższy poziom partycypacji obywatelskiej i zaufania do instytucji publicznych, lepsza jakość dostępnych operatów imiennych (więcej na ten temat poniżej), jak również niższy poziom nierówności ze względu na płeć (potencjalnie istotny zwłaszcza w badaniu wewnętrznej reprezentatywności próby w oparciu o tę właśnie cechę).

Choć zastosowane w publikacji modele statystyczne uwzględniają różnice między poszczególnymi krajami, taka specyfikacja wydaje się niewystarczająca w sytuacji, gdy niemal wszystkie kraje, w których występuje dana technika, zaliczają się do specyficznej grupy. Modele same w sobie nie dają zatem powodów, aby zakładać, że zaobserwowane zależności wystąpią w krajach spoza tej grupy – a dokładniej, że odejście od wywiadu kwestionariuszowego na rzecz innych technik, nawet przy zastosowaniu operatu imiennego, w innych krajach da równie pozytywny wynik. Wprawdzie wcześniejsza literatura (przywołana także w artykule) także wskazuje na istnienie sondaży stosujących inne techniki niż wywiad kwestionariuszowy, które osiągają wyższe wskaźniki reprezentatywności w porównaniu z tym ostatnim pomimo stosunkowo niskiej *response rate*, jednak badania, w których uzyskano taki wynik, również nie obejmują dużej liczby krajów (w artykule przywołano głównie badania z Holandii i Finlandii). Z kolei inna cytowana w pracy analiza oparta na danych porównawczych dla większej liczby krajów, zgromadzonych w ramach programu ESS, wskazuje na wyższą reprezentatywność sondaży zrealizowanych techniką wywiadu kwestionariuszowego, co nasila wątpliwości związane z możliwością generalizacji wyników analiz Doktoranta.

Po drugie, w swoich rozważaniach mgr Rybak w dużej mierze pomija kwestię jakości imiennego operatu możliwego do zastosowania w poszczególnych krajach. Rodzi to pytanie o to, na ile pozytywne rezultaty osiągnięte w krajach skandynawskich są pochodną istnienia lepszych baz danych umożliwiających skuteczny kontakt z wylosowanymi osobami. W innych krajach taka sytuacja niekoniecznie musi mieć miejsce, nawet przy zastosowaniu operatu imiennego – np. stosowany w Polsce rejestr PESEL zawiera jedynie dane o adresie zameldowania (nie faktycznego zamieszkania), co w praktyce prowadzi do problemów z dotarciem do wylosowanych osób, zwłaszcza w przypadku ankiet realizowanych tańszym kosztem i bez udziału ankietera. Co więcej, problem ten może wystąpić w różnym nasileniu w przypadku różnych kategorii populacji, zwiększając ryzyko *nonresponse bias*. To może stanowić jedno z wyjaśnień problemów z realizacją badania ESS techniką mieszaną w Polsce, opisanych w cytowanych przez Autora publikacjach P. Sztabińskiego. W autoreferacie Autor wspomina także o problemach związanych z jakością operatu imiennego, w kontekście dostępu do adresów e-mail obywateli, który to problem został rozwiązany poprzez wykorzystanie oficjalnych cyfrowych skrzynek pocztowych obywateli – z takiego operatu skorzystano w rundach 2018, 2019 i 2020 w ISSP w Danii (s. 31). Kwestię charakterystyk operatu imiennego podkreślam nie po to, aby podważać konkluzje Autora o jego znaczeniu dla zapewnienia jakości badań zwłaszcza realizowanych technikami niebezpośrednimi. Z tymi konkluzjami się zgadzam, sądzę jednak, że rekomendacje dla badaczy powinny tu zostać uzupełnione o lobbowanie na rzecz wprowadzenia operatów zawierających dane umożliwiające skuteczne dotarcie do wylosowanych osób.

Po trzecie, w interpretacji uzyskanych wyników warto byłoby zwrócić dokładniejszą uwagę na techniki sondażu uwzględnione w analizie. Z uwagi na charakterystyki występujące w badanych programach surveyowych (ESS oraz ISSP), Autor stosuje rozróżnienie na cztery kategorie: dwie pierwsze grupują badania z wykorzystaniem pojedynczej techniki: wywiad kwestionariuszowy oraz ankietę pocztową, dwie pozostałe to techniki mieszane, w podziale na takie, które obejmuje kontakt twarzą w twarz z ankieterem, oraz takie, które łączą ze sobą – w różnych konfiguracjach – ankietę pocztową, telefoniczną, bądź internetową (*mixed mode* bez

obecności ankietera). Ta ostatnia kategoria jest zróżnicowana wewnętrznie, i czytelnik niewiele dowiaduje się o tych zróżnicowaniach. Zgadza się, iż uwzględnienie charakterystyk poszczególnych strategii *mixed mode* w analizach statystycznych jest niemożliwe ze względu na ograniczoną liczbę badań zrealizowanych za ich pomocą. Uważam zarazem, że dla uniknięcia ogólnikowości oraz z punktu widzenia rekomendacji metodologicznych cennym uzupełnieniem rezultatów statystycznych byłoby dokładniejsze omówienie specyfiki podejść stosowanych w krajach i programach objętych analizą i zastanowienie się nad mechanizmami, które w ich przypadku mogły przełożyć się na większą reprezentatywność.

Warto wreszcie zwrócić uwagę na fakt, iż opisane wyżej zależności okazały się silniejsze po włączeniu do modelu zmiennej identyfikującej sondaże zrealizowane w ramach ESS. Generalnie te ostatnie uzyskały niższe wskaźniki błędu związanego z *non-response bias* oraz wyższe wskaźniki realizacji próby, w porównaniu z sondażami ISSP, co Autor słusznie tłumaczy wskazując na wysoką rzetelność procedur stosowanych w tym badaniu. Powyższy rezultat przemawia zarazem przeciwko pochoptemu odrzuceniu przekonania, że w dziedzinie terenowej realizacji sondaży „im drożej tym lepiej”.

Pewne wątpliwości może budzić także interpretacja moderatorów zastosowanych w metaanalizie. Przykładowo, włączenie efektów losowych dla krajów traktowane jest jako strategia częściowej kontroli efektu tzw. *survey climate* – co wydaje się uproszczeniem, jako że *survey climate* to tylko jeden z aspektów mogących składać się na różnice między jakością sondaży realizowanych w różnych krajach (co najmniej równie ważne mogą być tu czynniki instytucjonalne, o których pisałam powyżej). Drugi przykład to miara mająca uchwycić doświadczenie związane z realizacją badań surveyowych przez krajowe instytucje. Choć sam pomysł uwzględnienia doświadczenia instytucjonalnego w analizach uważam za bardzo cenny, zastosowana miara wydaje się nieco uproszczona i nie została do końca jasno opisana. W przypadku ISSP pojawia się pytanie, od jakiego momentu liczono wcześniejsze fale badania zrealizowanego w ramach programu w danym kraju (artykuł obejmował sondaże przeprowadzone od 1996, ale sam ISSP rozpoczął się znacznie wcześniej). W przypadku ESS, którego pierwsza fala przypadła na rok 2002, zaś badania realizowane są w cyklach dwuletnich, liczba wcześniejszych badań w ramach programu jest niższa w porównaniu do ISSP z definicji, co niekoniecznie oznacza mniejsze doświadczenie instytucji realizujących to badanie w poszczególnych krajach. Jakimś rozwiązaniem mogłoby być sprawdzenie, czy uwzględnienie w modelu interakcji między programem badawczym a liczbą wcześniejszych edycji sondażu ma wpływ na wyniki. Taki dodatkowy test byłby interesujący, niemniej należy podkreślić, że nie jest to najistotniejszy wątek opisywanej publikacji, zaś jego brak nie podważa wiarygodności uzyskanych wyników.

Ostatnia uwaga do publikacji 2 ma charakter bardziej ogólny. Choć skupienie się na reprezentatywności zrealizowanych prób badawczych w analizach uważam za uprawomocniony wybór przedmiotu badania, w dyskusji wyników (zwłaszcza tej szerszej zamieszczonej w autoreferacie) warto jaśniej podkreślić, że przedstawione analizy dają jedynie częściowy obraz reprezentatywności; w szczególności nic nam nie mówią o *nonresponse bias* w odpowiedziach na konkretne pytania. Zasadność stosowania różnych technik sondażu bardzo zależy od zagadnień, których sondaż dotyczy, długości sondażu, stopnia złożoności pytań.

Obecność ankietera, zwłaszcza w przypadku dłuższych, trudniejszych sondaży, stanowić może czynnik sprzyjający pełnej realizacji, bez pomijania pytań czy przerywania w trakcie. Warto analizy uzupełnić o takie, które sprawdziłyby, na ile uzyskany rezultat uległby zmianie, gdyby np. uwzględnić tylko respondentów, którzy wypełnili ankiety do końca (pełną realizację badania; odrzucając ankiety wypełnione częściowo). Jak jednak słusznie zauważa Doktorant, przeprowadzenie takich bardziej kompleksowych analiz wymagałoby zaadresowania kwestii braku ujednoliconych definicji ankiety wypełnionej częściowo oraz braku odpowiednich danych w dokumentacji wielu badań.

W **artykule 3** Doktorant przedstawia – w oparciu o przegląd literatury – ciekawą i nowatorską koncepcję mechanizmów kształtujących zależności między dopasowaniem techniki badania do preferencji respondenta a jego skłonnością do udziału w badaniu i jakością udzielanych odpowiedzi. Nowością jest, po pierwsze, zastosowanie jednolitej ramy pojęciowej do zrozumienia źródeł błędów braku odpowiedzi (w tej mierze, w jakiej brak odpowiedzi stanowi efekt dążenia do minimalizacji wysiłku) i błędów pomiaru (udzielanie niedokładnych, niezgodnych z prawdą odpowiedzi). Aby wyjaśnić oba zjawiska, Autor odwołuje się to koncepcji *satisficingu*. Po drugie, cenna i warta podkreślenia jest próba włączenia do proponowanego modelu wiedzy pochodzącej z różnych dyscyplin nauki – zwłaszcza w tej mierze, w jakiej do analiz czynników obniżających jakość i reprezentatywność danych będących przedmiotem literatury metodologicznej, włączyć można dorobek psychologii poznawczej (koncepcję pamięci roboczej). Po trzecie, ważnym aspektem publikacji jest to, iż problematyzuje kwestię czynników mogących determinować zmienność preferencji techniki sondażu na przestrzeni życia jednostek, co stanowi postęp względem przyjętego w literaturze upraszczającego założenia o stałości tych preferencji.

Proponowana konceptualizacja wydaje mi się jednak nie do końca jasna i miejscami niepełna. Przede wszystkim szkoda, iż Autor nie pokusił się o przedstawienie zakładanych w proponowanym modelu zależności przyczynowych np. za pomocą diagramu. Ułatwiłoby to znacznie zrozumienie sposobu, w jaki poszczególne czynniki mogą oddziaływać na siebie nawzajem, a także uwzględnienie w modelu innych zmiennych mogących interweniować w analizowane zależności. Jeśli dobrze zrozumiałam, artykuł formułuje następujące tezy dotyczące tych oddziaływań:

- 1) ograniczenia pamięci roboczej zwiększają skłonność do *satisficingu* wśród respondentów badań sondażowych
- 2) pamięć robocza stanowi jedną z determinant preferencji techniki sondażu zgodnie z założeniem, że preferowana będzie technika mniej obciążająca poznawczo
- 3) niezgodność zastosowanej techniki sondażu z preferowaną zwiększa prawdopodobieństwo *satisficingu*
- 4) *satisficing* przekłada się na wyższy odsetek odmów i niższą jakość odpowiedzi
- 5) pamięć robocza zmienia się na przestrzeni życia.

Powyższe tezy nie wykluczają się nawzajem, niemniej zestawienie ich ze sobą ukazuje złożoność podjętej tu problematyki, zwłaszcza jeśli uwzględnić dodatkowo to, iż na jakość udzielanych odpowiedzi wpływ mają zarówno możliwości poznawcze, jak i poziom trudności oraz motywacje (s. 125). Przede wszystkim nie jest jasne, co jest głównym czynnikiem

generującym *satisficing* w przypadku braku zgodności techniki sondażu z preferencją respondenta – czy jest to przede wszystkim brak możliwości (tj. respondent chce udzielić poprawnej odpowiedzi, ale z powodu trudności z przetwarzaniem informacji nie jest w stanie tego zrobić), czy brak motywacji (przekładający się na mniej lub bardziej świadome dążenie do minimalizacji wysiłku, a w skrajnej formie odmowa udziału w badaniu). Na to nakłada się złożoność czynników mogących kształtować preferencje techniki sondażu, z których część nie musi się wiązać z zasobami pamięci roboczej: według literatury cytowanej w artykule do takich czynników można zaliczyć m.in. samotność, problemy psychiczne, bądź obawy o anonimowość (s. 120). Znaczenie mogą mieć również czynniki kulturowe i związane z trybem życia danej osoby (np. to, na ile jest w stanie poświęcić czas na spotkanie z ankierem). W dodatku zarówno pamięć robocza, jak i inne determinanty preferencji mogą nie tylko zmieniać się w ciągu życia, lecz także różnić pomiędzy jednostkami. W zależności od tego, jakie czynniki mają większe znaczenie w kształtowaniu preferencji, efektem może być albo większy odsetek odmów udziału w badaniu (lub odpowiedzi „na chybcika”) – wskutek niskiej motywacji, albo większy odsetek niezamierzonych błędnych odpowiedzi – efekt trudności poznawczych. Istnienie powyższej komplikacji dobrze ilustrują wyniki badań przywołanych przez Autora, w których jedna z porównywanych technik sondażu (dokładniej: internetowa) z jednej strony poprawiła jakość pomiaru, ale z drugiej zwiększyła liczbę braków odpowiedzi (chodzi o badania Sakshauga, Tinga i Tourangeau z 2010; s. 124). W praktyce komplikacja ta oznacza, że dopasowanie techniki do preferencji nie zawsze prowadzi będzie do poprawy wiarygodności danych. Mam zatem poczucie, że przedstawiona koncepcja to nie tyle źródło gotowych odpowiedzi, ile punkt wyjścia do postawienia nowych, fascynujących pytań badawczych, co w gruncie rzeczy stanowi jej zaletę, i co warto by wyraźniej podkreślić.

Idąc dalej tym tropem, można zadać pytanie o i interpretację obserwowanych przez Autora (także w artykule 4) krzywoliniowych zależności między wiekiem a preferencją techniki sondażu (*Najwyższe prawdopodobieństwo wyboru techniki pocztowej występowało w kategorii osób w wieku 45–64 lat, natomiast powyżej tego wieku gwałtownie spadało (na korzyść przede wszystkim wywiadu kwestionariuszowego...)*; s. 127-128). Doktorant interpretuje tę zależność jako świadczącą o znaczeniu pamięci roboczej, której sprawność zmienia się wraz z wiekiem. Teoretycznie możliwe jest jednak alternatywne wyjaśnienie, w duchu *life-course studies*: możliwość swobodnego wyboru czasu wypełnienia ankiety może mieć większe znaczenie dla pracujących osób w średnim wieku niż dla będących na emeryturze. Bez empirycznej weryfikacji nie wiemy, która z tych hipotez jest bliższa prawdzie (lub który z tych mechanizmów ma większe znaczenie).

W **artykule 4**, ostatnim z przedstawionego cyklu, Doktorant podjął się empirycznego zbadania demograficznych korelatów preferencji techniki sondażu wśród polskich respondentów dwóch edycji badań ESS, przy jednoczesnym porównaniu wyników uzyskanych dzięki zastosowaniu dwóch alternatywnych sposobów pomiaru preferencji: na podstawie deklaracji badanych oraz obserwacji faktycznych wyborów. Podstawową słabością metodologiczną tych analiz jest, na co słusznie zwraca uwagę Autor, ograniczenie wyboru techniki w przypadku behawioralnego pomiaru preferencji do ankiety pocztowej i internetowej. Stanowi to w oczywisty sposób przynajmniej częściowe wyjaśnienie obserwowanych w modelach regresji wielomianowej zależności między formą pomiaru preferencji a

preferowaniem wywiadu kwestionariuszowego. Zdaniem Doktoranta powyższe różnice nie wyjaśniają jednak niezgodności między odsetkiem preferujących ankiety pocztowe i internetowe, gdyż obie opcje były tak samo dostępne, a zarazem nie mogły podlegać zniekształceniom związanym z deklarowaniem preferencji dla techniki, w której się samemu wzięło udział. Ta ostatnia teza budzi moje wątpliwości. Teoretycznie nie można wykluczyć sytuacji, w której większość respondentów na pierwszym miejscu w hierarchii preferencji stawia wywiad kwestionariuszowy, zaś w dalszej kolejności ankiety pocztowe, następnie internetowe. Taki rozkład preferencji może tłumaczyć zaobserwowane różnice w sytuacji, gdy w jednym badaniu respondenci mają opcję wyboru wywiadu kwestionariuszowego, zaś w drugim ci, którzy preferowaliby wywiad kwestionariuszowy, z braku takiej opcji częściej wybierają ankietę pocztową. Powyższy przykład ilustruje główny problem przeprowadzonych analiz: trudno o dokładne badanie zgodności między deklarowanymi preferencjami a faktycznym wyborem w sytuacji, gdy każdy z tych wskaźników mierzony jest na odrębnej próbie, zaś badanym w obu próbach przedstawiono odmienny zestaw opcji. Prezentowane wyniki mogą co najwyżej posłużyć jako podstawa hipotez do dalszych analiz.

Mam także wątpliwości dotyczące doboru zmiennych niezależnych w analizach, w szczególności ograniczenie ich do podstawowych cech demograficznych. Z jednej strony rozumiem uzasadnienie Autora, który dobór predyktorów podporządkował praktycznemu celowi analiz – mających zmierzać do wypracowania modelu pozwalającego przewidywać preferencje techniki sondażu do zastosowania w projektowaniu przyszłych badań sondażowych realizowanych techniką *mixed mode*. Stąd uwzględnienie jedynie tych charakterystyk, które dostępne są w imiennych operatach losowania (płeć, wiek, miejsce zamieszkania / zameldowania). Z drugiej strony, w świetle moich uwag do publikacji 3 przedstawionych wyżej, dla lepszego zrozumienia czynników kształtujących preferencje techniki sondażu warto byłoby w przyszłych analizach tego typu uzupełnić modele o dodatkowe społeczno-ekonomiczne charakterystyki respondentów (np. poziom wykształcenia, miary aktywności ekonomicznej) oraz stanu ich zdrowia. Zwłaszcza, że zastosowanie danych pochodzących z badań ESS umożliwia takie pogłębione analizy.

Uwagi techniczne

Na koniec chciałam zwrócić uwagę na drobne niedociągnięcia natury formalnej, które nieco utrudniły mi lekturę autoreferatu rozprawy mgr Rybaka. Pierwsza kwestia dotyczy sposobu porządkowania treści w analizie literatury przedmiotu: fragment o wskaźnikach błędu braku odpowiedzi lepiej byłoby zamieścić na początku części II.2; ułatwiłoby to zrozumienie dalszego wywodu (w pierwszej części przeglądu Autor odwołuje się do miar, które zostały wyjaśnione dopiero w części kolejnej). Druga kwestia dotyczy braku opisów zmiennych pod niektórymi tabelami (tabele 6-8). W tabeli 9 brakuje informacji o kategoriach referencyjnych zmiennych uwzględnionych w modelu; w przypadku technik sondażu; czytelnik musi się domyślić (lub sprawdzić w publikacji 2), że jest to wywiad kwestionariuszowy. Kolejnym niedociągnięciem było odwołanie do nieprzetłumaczonej dokumentacji (rys. 1, s. 41), bez dokładnego wyjaśnienia, w jaki sposób schemat wyboru respondenta w obrębie gospodarstwa generuje błędy. W tekście rozprawy zabrakło wreszcie tabeli S1 z publikacji 2 (jest dostępna online i dzięki temu udało mi się z nią zapoznać).

Podsumowanie

Niezależnie od szczegółowych uwag krytycznych zamieszczonych powyżej, moja ocena przedstawionej do recenzji pracy doktorskiej jest pozytywna. Doktorant w swojej pracy odwołuje się do światowej literatury przedmiotu, wykraczającej poza zasięg wąsko rozumianej dyscypliny naukowej. Widać zarazem, iż świetnie orientuje się w zagadnieniach współczesnej metodologii badań surveyowych. Zarówno teoretyczne rozważania, jak i zastosowane strategie badawcze i interpretacja wyników dowodzą wysokich kompetencji analitycznych mgr Rybaka oraz umiejętności samodzielnego rozwiązywania problemów badawczych. Uzyskane wyniki dostarczają nowej wiedzy o znaczeniu wykraczającym poza kontekst ogólnopolski.

Po zapoznaniu się z treścią autoreferatu i cyklu publikacji składających się na rozprawę doktorską mgr Adama Rybaka oraz po przeanalizowaniu zawartych w nich wyników analiz i wniosków, stwierdzam zatem, iż praca **stanowi oryginalne rozwiązanie problemu naukowego o istotnym znaczeniu dla dyscypliny nauki socjologiczne**. Tym samym spełnia wymagania określone w art. 187 ust. 1 ustawy z dnia 20 lipca 2018 r. – **Prawo o szkolnictwie wyższym i nauce**. Mając na uwadze powyższe, wnioskuję o dopuszczenie mgr Adama Rybaka do dalszych etapów przewodu doktorskiego, w tym do publicznej obrony rozprawy doktorskiej.

