

HENRYK DOMAŃSKI
Instytut Filozofii i Socjologii PAN

WYBÓR MIĘDZY ALTERNATYWNYMI WERSJAMI PYTAŃ W KWESTIONARIUSZU WYWIADU

Zacznijmy od oczywistego stwierdzenia, że odpowiedzi respondentów służą do konstrukcji wskaźników. Doskonalenie ich jest przedmiotem eksperymentów metodologicznych, które koncentrują się – najogólniej rzecz biorąc – na dwóch etapach standardowego badania. Pierwszy z nich dotyczy etapu konstruowania pytań, drugi – tworzenia z odpowiedzi respondentów zmiennych, na etapie analiz. Ustalenia, które przedstawiam poniżej dotyczą tego pierwszego etapu.

Błędy związane z konstrukcją kwestionariuszy wywiadu od samego początku przyciągały uwagę badaczy. Wiadomo, że błąd jest immanentnym składnikiem pomiaru w każdej dziedzinie nauki. W odniesieniu do pytań zajmowano się, przykładowo, kwestiami: ich drażliwości, trudności, długości, kategoriami odpowiedzi, skłonność do potakiwania, tego czy mają być one otwarte czy zamknięte i umieszczenia ich we właściwym miejscu w kwestionariuszu wywiadu (Hyman 1954; Bradburn i Sudman 1979; Sztabiński 1997). Analizy te pozwoliły sformułować określone reguły, istnieje jednak świadomość, że nie są one czymś danym na zawsze i wymagają zmian w zależności od badanej populacji, kraju i rozpatrywanego tematu.

Skłania to do podjęcia tych zagadnień w odniesieniu do aktualnego kontekstu. Celem analiz, które tutaj przedstawiam, jest dostarczenie odpowiedzi w trzech kwestiach związanych z dążeniem do optymalnego sposobu formułowania pytań w kwestionariuszu wywiadu. Kwestie te dotyczą wyboru alternatywnych rozwiązań w zakresie: (i) liczby punktów na skalach, (ii) formułowania odpowiedzi, (iii) posługiwania się kartą respondenta. Empirycznego materiału do rozstrzygnięcia tych kwestii dostarczyły wyniki *Europejskiego Sondażu Społecznego* przeprowadzonego w 22 krajach. Ograniczę się tu do Polski.

DANE I SCHEMAT ANALIZY

Europejski Sondaż Społeczny jest stosunkowo najnowszym międzynarodowym projektem badawczym. Podjęto go w celu monitorowania zmian związanych z rozszerzeniem ram Unii Europejskiej, stąd w kwestionariuszu ESS wiele miejsca zajmują pytania na temat stosunku do imigrantów i mniejszości etnicznych, aktywności jednostek w życiu publicznym i postaw wobec gospodarki rynkowej. Autorzy badania poświęcili również osobną uwagę doskonaleniu narzędzi pomiaru tych zjawisk, przy czym szczególnie nacisk położono na wypróbowanie różnych wariantów pytań identyfikujących postawy jednostek. W tym celu, w kwestionariuszu ESS dwukrotnie powtórzono kilkanaście pytań dotyczących merytorycznie tych samych problemów. Za każdym razem zadawano je respondentom nieco inaczej – sytuując je na początku i na końcu kwestionariusza wywiadu. Porównanie odpowiedzi na te pytania miało dostarczyć danych do wniosków na temat tego, która forma jest lepsza.

Pierwszy blok, składający się z trzech pytań, identyfikuje przekonania jednostek dotyczące tego, czy warto ludziom ufać, postępować w życiu uczciwie i udzielać innym pomocy. Różnica polegała na zastosowaniu niejednakowych skal odpowiedzi. Za pierwszym razem respondentów proszono o udzielenie odpowiedzi na dziesięciopunktowych skalach przedstawionych w postaci podziałki na karcie, wręczanej im przez ankietatorów w trakcie wywiadu. Natomiast za drugim razem respondenci odpowiadali na te pytania „tak-nie”. Wszystkie analizowane pytania przytaczam w literalnym brzmieniu z kwestionariusza ESS w załączniku do tego artykułu zamieszczonym na końcu. Wymienione przed chwilą pytania oznaczone są numerami A8-A10 w wersji oryginalnej i HF 28-HF30 w wersji alternatywnej, zadawanej respondentom na końcu wywiadu.

Drugi blok składał się z trzech popularnych pytań dotyczących „osobistego” zaufania respondentów do parlamentu, policji i systemu prawnego. Badane osoby wyrażały swoje opinie na 10-punktowych skalach, z tym, że o ile za pierwszym razem posługiwały się kartą respondenta (B7-B9), to za drugim razem (HF31-HF33) tej możliwości nie było. Zadaniem trzeciego bloku było uchwycenie stosunku respondentów do istniejącego systemu. Respondentom zadano trzy pytania dotyczące zadowolenia ze stanu polskiej gospodarki, demokracji i rządu. Test metodologiczny polegał na wypróbowaniu skal odpowiedzi o różnej długości: za pierwszym razem była to skala 10-punktowa, a drugim 5-punktowa (B30-B32 i HF25-HF27). Na koniec, w bloku czwartym, znalazły się trzy pytania identyfikujące stosunek respondentów do aktywnej roli rządu i polityki socjalnej. W obydwu wersjach posłużono się ska-

lami pięciopunktowymi, chodziło zaś o przetestowanie wpływu sformułowania odpowiedzi: w pierwszej wersji (B43-B45) wyrażały one nieco bardziej osobisty stosunek respondentów (od „zdecydowanie zgadzam się” do „zdecydowanie nie zgadzam”), w drugiej (HF34-HF36) były one bardziej bezosobowe (od „na pewno tak” do „na pewno nie”).

Dane te pozwalają ustosunkować się do trzech problemów technicznych, bez których metodologia badań surveyowych nie może się obejść (Koniarek 1983; Sztabiński 1997). Pierwszy z nich dotyczy stopnia szczegółowości skali postaw – chodzi o wybór optymalnej liczby punktów na skali. Dane ESS dają możliwość rozstrzygnięcia, czy optymalnym wariantem jest wersja zawierająca 10 punktów, 5 punktów, czy też odpowiedź typu „tak-nie”. Po drugie, będziemy mogli stwierdzić, czy istnieje różnica między stosowaniem odpowiedzi typu „zgadzam się”, a odpowiedziami typu „na pewno”, a jeżeli tak, to które z tych rozwiązań jest lepsze. Wreszcie porównanie odpowiedzi na pytania, w których nie posłużono się kartą respondenta z odpowiedziami na identyczne pytania z zastosowaniem karty, dostarczy rekomendacji, które z tych rozwiązań dostarcza danych o wyższej jakości.

Pytania te porównamy ze względu na ich „operacyjność” i trafność. Pytanie jest bardziej „operacyjne”, gdy nie porusza ono drażliwych kwestii i nie powoduje zafalszowań i odmów. Charakterystyką operacyjności będzie tu odsetek braków danych i odpowiedzi typu „nie wiem”. Co do trafności, to – zgodnie z klasyczną definicją – pytanie jest trafne w tym stopniu, w jakim mierzy to co ma mierzyć, czyli na ile dobrze identyfikuje badane postawy jednostek. Zwykło się, najogólniej rzecz biorąc, przyjmować, że statystycznym miernikiem trafności wskaźnika jest siła jego korelacji z jakimś „zewnętrznym kryterium” – wyższy poziom korelacji informuje o większej trafności. Analizując trafność rozpatrywanych tu pytań będziemy porównywać siłę ich korelacji z wybranymi zmiennymi.

Empiryczną podstawą naszych analiz są dane z badania ESS zrealizowanego w listopadzie 2002 roku klasyczną metodą bezpośredniego wywiadu¹. Zrealizowana próba, dobrana metodą losową obejmuje 2110 respondentów reprezentujących dorosłych mieszkańców Polski zaczynając od 18 roku życia (Sztabiński 2001).

¹ Badanie to przeprowadził Instytut Filozofii i Socjologii PAN w ramach grantu KBN kierowanego przez Pawła Sztabińskiego. Szczegółowe informacje na temat możliwości korzystania ze zbioru ESS można znaleźć pod: www.europeansocialsurvey.org

OPERACYJNOŚĆ

Braki danych i odpowiedzi typu „nie wiem”, które traktujemy tu jako charakterystykę „operacyjności” wskaźników, są elementem każdego badania. Jeżeli chodzi o braki danych, to wynikają one z odmowy odpowiedzi, złej pracy ankierów i błędów przy kodowaniu danych. Można je traktować jako klasyczny błąd, który zwykle się eliminować ze zbioru.

W odróżnieniu od braków danych, odpowiedzi typu „nie wiem”, dostarczają informacji, które odpowiadają na pytania badawcze. Odpowiedź „nie wiem” może identyfikować różne zjawiska – od autentycznej „niewiedzy”, przez niezdecydowanie, do postawy neutralności wobec badanego obiektu. W zależności od preferencji badacza można te odpowiedzi traktować jako znaczące i uwzględniać je przy konstrukcji wskaźników i odwrotnie – można je traktować jako błąd i eliminować z analiz. Argumentem na rzecz tego drugiego rozwiązania jest podejrzenie, że pytanie zostało sformułowane niejasno. Natomiast jeżeli badaczowi zależy na zapewnieniu sobie wystarczającej liczby przypadków – a trzeba stwierdzić, że liczebność zbioru jest ważnym parametrem przy testowaniu hipotez – istnieją różne możliwości uwzględniania braków danych i odpowiedzi „nie wiem” na etapie analiz. Opracowano w tym celu specjalne techniki i programy statystyczne, których omówienie musimy tu z braku miejsca pominąć (zob. Ragunathan i in. 1997). W naszych analizach braki danych i odpowiedzi „nie wiem” potraktujemy jako błąd zmniejszający trafność wskaźnika.

T a b e l a 1. ODPOWIEDZI „NIE WIEM” I BRAKI DANYCH DLA WYBRANYCH PYTAŃ ZADANYCH W DWÓCH WERSJACH W *EUROPEJSKIM SONDAŻU SPOŁECZNYM*

Pytania zadawane w wersji: oryginalnej / testowej	Nie wiem (odsetki w %)		Brak danych (liczebności)	
	I wersja	Test	I wersja	Test
Czy uważa P., że większości ludzi można ufać, czy też, że w kontaktach z ludźmi ostrożności nigdy za wiele? (A8 / HF28)	0,6	4,2	1	9
Czy uważa P., że gdyby nadarzyła się okazja, większość ludzi starałaby się P. wykorzystać, czy też starałaby się postępować uczciwie? (A9 / HF29)	2,5	11,5	0	4
Czy uważa P., że ludzie starają się służyć pomocą innym, czy że przede wszystkim dbają o swój własny interes? (A10 / HF30)	0,9	6,1	0	3
Na ile osobiście ma P. zaufanie do polskiego parlamentu, tzn. sejmu i senatu? (B7 / HF31)	4,3	4,2	0	3

Pytania zadawane w wersji: oryginalnej / testowej	Nie wiem (odsetki w %)		Brak danych (liczebności)	
	I wersja	Test	I wersja	Test
Na ile osobiście ma P. zaufanie do systemu prawnego? (B8 / HF32)	4,5	4,4	0	3
Na ile osobiście ma P. zaufanie do policji? (B9 / HF33)	1,9	2,5	1	3
Na ile jest P. zadowolony/a z obecnego stanu polskiej gospodarki? (B30 / HF25)	3,1	2,5	1	4
Na ile jest P. zadowolony/a ze sposobu działania obecnego polskiego rządu? (B31 / HF26)	4,0	2,8	1	3
W jakim stopniu jest P. zadowolony/a z funkcjonowania demokracji w Polsce? (B32 / HF27)	6,4	5,0	1	9
W jakim stopniu zgadza się P. lub nie zgadza ze stwierdzeniem: im mniej rząd angażuje się w sferę gospodarczą, tym lepiej dla Polski? (B43 / HF34)	9,2	10,1	2	5
W jakim stopniu zgadza się P. lub nie zgadza ze stwierdzeniem: rząd powinien podjąć działania zmierzające do zmniejszenia różnic w dochodach? (B44 / HF35)	2,7	2,5	2	6
W jakim stopniu zgadza się P. lub nie zgadza ze stwierdzeniem: pracownicy potrzebują silnych związków zawodowych dla ochrony warunków pracy i poziomu płac? (B45 / HF36)	4,2	5,2	1	4

W tabeli 1 zamieściłem odsetki odpowiedzi typu „nie wiem” i liczbę braków danych. Wielkości w pierwszej i trzeciej kolumnie odnoszą się do odpowiedzi na pytania zadane na początku wywiadu, natomiast w kolumnie drugiej i czwartej – do pytań zadanych w alternatywnej wersji na końcu. Zaczniemy od odpowiedzi kodowanych jako „brak danych”. Było ich stosunkowo niedużo i – jak łatwo stwierdzić – wystąpiły one częściej w odpowiedziach na pytania zadane na końcu wywiadu. Ze względu na ich „śladową” obecność ograniczyłem się do prezentacji ich liczebności bez podawania ich udziału wśród odpowiedzi ogółem. Wynik wydaje się być jednoznaczny. Można bez większego ryzyka stwierdzić, że pojawienie się braków danych w pytaniach zadawanych na końcu jest rezultatem dobrze znanego efektu, związanego z określonym usytuowaniem pytań w kwestionariuszu wywiadu. Im bliżej końca, tym większe jest prawdopodobieństwo odmowy, ankieterzy i respondenci popełniają więcej błędów, a uzyskanie trafnych odpowiedzi staje się coraz trudniejsze. Rzecz jasna, wzrost liczby braków danych nie pozo-

staje w żadnym związku z odmiennym sformułowaniem pytania, nie stosowaniem karty respondenta, czy posługiwaniem się określoną liczbą punktów na skali.

W przypadku odpowiedzi typu „nie wiem” nie było żadnych oznak występowania tego efektu. Natomiast odsetek tych odpowiedzi zdecydowanie wzrasta przy zastosowaniu skali dychotomicznej „tak-nie”. Pytania te dotyczyły zaufania, uczciwego postępowania z ludźmi i tego, czy warto ludziom pomagać. Najwięcej „nie wiem” padło w odpowiedzi na pytanie: „Czy uważa P., że gdyby nadarzyła się okazja, większość ludzi starałaby się P. wykorzystać, czy też starałaby się postępować uczciwie?”. O ile przy odpowiedzi „tak-nie” kategoria „nie wiem” liczyła 11,5%, to przy zastosowaniu skali dziesięciopunktowej wyniosła ona tylko 2,5%. Z kolei, w pytaniu „czy uważa P., że ludzie przede wszystkim starają się służyć pomocą innym, czy też, że przede wszystkim dbają o własny interes?” zastosowanie dychotomicznego podziału pociągnęło za sobą wzrost z 0,9 do 6,1%. Wyraźny wzrost – z 0,6 do 4,2% – wystąpił również dla pytania „czy ogólnie biorąc, uważa P., że większości ludzi można ufać, czy też w kontaktach z ludźmi ostrożności nigdy za wiele?”.

Wynika stąd, że skala dychotomiczna jest gorsza. Nasuwa się pytanie, jaka długość skali generuje najmniej odpowiedzi „nie wiem”, co oznaczałoby, że jest z tego punktu widzenia najlepsza. Porównajmy różnice między skalą dziesięcio i pięciopunktową w pytaniach dotyczących polskiej demokracji, gospodarki i rządu. Różnice okazują się tu być stosunkowo nieduże, niemniej rysuje się tendencja na korzyść pytań zawierających pięć punktów. W porównaniu ze skalami dziesięciopunktowymi dostarczają one nieco mniej odpowiedzi „nie wiem”. Dla ilustracji spójrzmy na pytanie „na ile jest P. zadowolony/a z obecnego stanu polskiej gospodarki?”. O ile dla wersji z podziałką dziesięciopunktową w kategorii tej znalazło się 3,1% respondentów, to dla skali pięciopunktowej odsetek ten wyniósł 2,5%. Analogiczny spadek „nie wiem” wystąpił w pytaniach dotyczących zadowolenia z polskiego rządu i funkcjonowania demokracji. Ustalenia te prowadzą do wniosku, że skala pięciopunktowa dostarcza najmniej odpowiedzi „nie wiem”. Potraktujmy to jako hipotezę, dopiero bowiem bezpośrednie porównanie wyników dla wszystkich trzech skal pozwoliłoby na bardziej jednoznaczne stwierdzenia.

Jeżeli chodzi o pozostałe dwa testy, to należy stwierdzić, że posłużenie się kartą respondenta nie spowodowało zmniejszenia relatywnego udziału odpowiedzi „nie wiem”. Nic nie wskazuje żeby zadziałał tu – związany z kartą – jakiś efekt percepcji ułatwiający zadeklarowanie postawy. Nie ma tu jednoznacznego wzoru różnic między dwiema wersjami, a same różnice są małe.

Czwarty test polegał na porównaniu 5-punktowych skal różniących się sformułowaniem odpowiedzi. W pierwszej wersji były to odpowiedzi „zgadzam się – nie zgadzam”, w drugiej „na pewno tak – na pewno nie”. Zastosowano je w pytaniach dotyczących stosunku do zaangażowania się rządu w sferę gospodarczą, jego działań na rzecz zmniejszenia różnic w dochodach i potrzeby istnienia silnych związków zawodowych. Zastosowanie formy „zgadzam się” dostarczyło nieco mniej „nie wiem” w pytaniach o ocenę aktywności gospodarczej rządu i działalności związków zawodowych. Odwrotnie jest w pytaniu na temat nierówności dochodów, jednak była to różnica nieduża.

TRAFNOŚĆ

Zasadniczym kryterium oceny użyteczności wskaźników jest ich „siła wyjaśniająca”, którą zwykle się rozpatrywać w terminach siły związku między wskaźnikiem a identyfikowanym przez niego zjawiskiem. Jest to precyzyjny miernik trafności, który pozwala dokonać wyboru w sytuacji, gdy dysponujemy zestawem alternatywnych wskaźników.

Zachowując w pamięci rozkłady braków danych i odpowiedzi „nie wiem”, przejdźmy do porównania trafności pytań. Zaczniemy od ustalenia, w jakim stopniu alternatywne wersje każdego z tych pytań „mierzą” te same zjawiska. Bezpośredniego rozstrzygnięcia tej kwestii dostarcza porównanie współczynników korelacji między wersjami. Będzie to pierwszy krok analizy trafności. Wszystkie pytania mierzone są na skalach porządkowych, stąd też najbardziej odpowiednim miernikiem korelacji wydaje się być w tym przypadku współczynnik *Gamma*, stosowany w analizie zmiennych reprezentujących ten poziom pomiaru. Wielkości *Gamma* zawierają się w granicach od -1 do 1 .

Tabela 2. WSPÓLCZYNNIKI KORELACJI MIĘDZY PYTANIAMI ZADANYMI W DWÓCH WERSJACH W EUROPEJSKIM SONDAŻU SPOŁECZNYM

Pytania zadawane w wersji: oryginalnej / testowej	Współczynniki korelacji Gamma
Czy uważa P., że większości ludzi można ufać, czy też, że w kontaktach z ludźmi ostrożności nigdy za wiele? (A8 / HF28)	0,75
Czy uważa P., że gdyby nadarzyła się okazja, większość ludzi starałaby się P. wykorzystać, czy też starałaby się postępować uczciwie? (A9 / HF29)	0,75
Czy uważa P., że ludzie starają się służyć pomocą innym, czy że przede wszystkim dbają o swój własny interes? (A10 / HF30)	0,78
Na ile osobiście ma P. zaufanie do polskiego parlamentu, tzn. sejmu i senatu? (B7 / HF31)	0,74
Na ile osobiście ma P. zaufanie do systemu prawnego? (B8 / HF32)	0,73

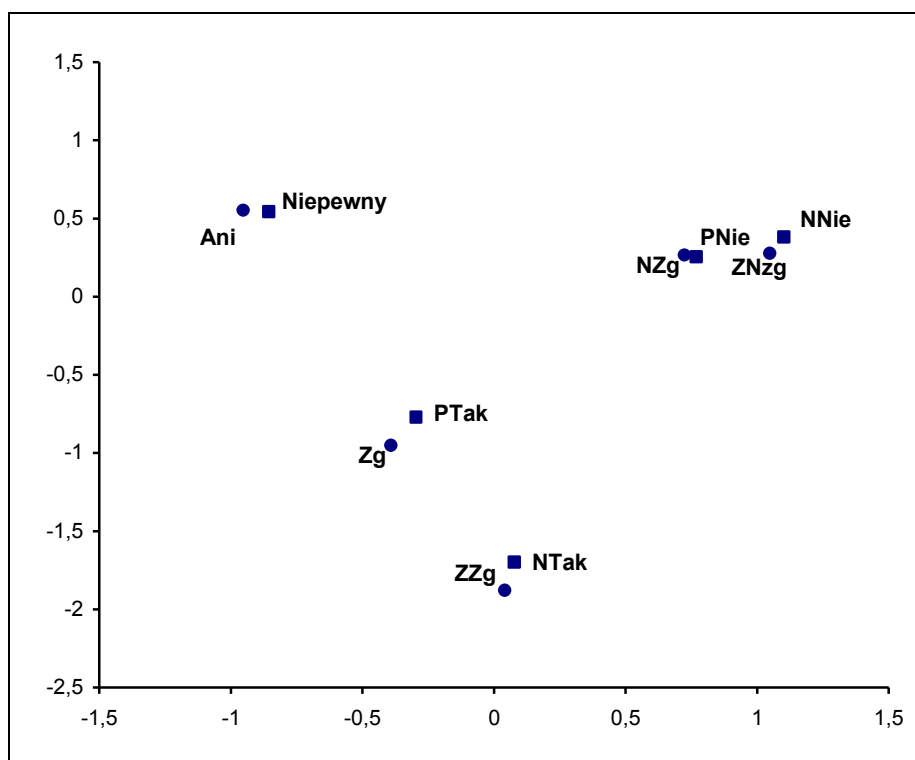
Pytania zadawane w wersji: oryginalnej / testowej	Współczynniki korelacji Gamma
Na ile osobiście ma P. zaufanie do policji? (B9 / HF33)	0,79
Na ile jest P. zadowolony/a z obecnego stanu polskiej gospodarki? (B30 / HF25)	0,64
Na ile jest P. zadowolony/a ze sposobu działania obecnego polskiego rządu? (B31 / HF26)	0,72
W jakim stopniu jest P. zadowolony/a z funkcjonowania demokracji w Polsce? (B32) / HF27	0,69
W jakim stopniu zgadza się P. lub nie zgadza ze stwierdzeniem: im mniej rząd angażuje się w sferę gospodarczą, tym lepiej dla Polski? (B43 / HF34)	0,60
W jakim stopniu zgadza się P. lub nie zgadza ze stwierdzeniem: rząd powinien podjąć działania zmierzające do zmniejszenia różnic w dochodach? (B44 / HF35)	0,74
W jakim stopniu zgadza się P. lub nie zgadza ze stwierdzeniem: pracownicy potrzebują silnych związków zawodowych dla ochrony warunków pracy i poziomu płac? (B45 / HF36)	0,70

Wartości tych korelacji przedstawione są w tabeli 2. Największe podobieństwo występuje między alternatywnymi wersjami pytania dotyczącego zaufania respondentów do policji (B9 i HF33). Był to test na porównanie pytań: z kartą i bez karty. Wielkość współczynnika korelacji między tymi skalami wynosi 0,79. Z kolei najsłabszy związek (0,60) występuje między wersjami odpowiedzi dotyczących stosunku do angażowania się rządu w sferę gospodarczą (B43 i HF34). Respondenci formułowali je w dwóch wersjach, posługując się za pierwszym razem kategoriami odpowiedzi typu „zgadzam się”, a za drugim – „na pewno”.

Korelacje między wszystkimi wersjami zawierają się więc w granicach od 0,60 do 0,79. Wolno twierdzić, że są one raczej wysokie. Prawidłowości społeczne nie są bezwyjątkowe i realizują się tylko z pewnym prawdopodobieństwem, stąd też w analizach socjologicznych nie uzyskuje się korelacji zbliżonych do 1. Z danych tych można by wysnuć wniosek, że każda z dwóch wersji pytania mierzy to samo zjawisko, ale nieco inaczej. Aby zilustrować empiryczny sens tego stwierdzenia posłużyłem się analizą korespondencyjną, która pozwala ustalić bliskość między kategoriami odpowiedzi z dwóch wersji. Analiza korespondencyjna jest techniką, której celem jest ustalenie struktury zależności między dwiema zmiennymi przedstawionymi w dwuzmiennej tabeli. Dokonuje się to drogą dekompozycji tego rozkładu na pewną liczbę wymiarów. Każdej kategorii przypisywane są współrzędne, które identyfikują ich usytuowanie w wielowymiarowej przestrzeni (zob. Gór-

niak 2000). Zamieszczony poniżej wykres identyfikuje ich usytuowanie w dwóch pierwszych, najbardziej znaczących wymiarach. Jest to graficzna postać związku między alternatywnymi odpowiedziami na pytanie: „im mniej rząd angażuje się w sferę gospodarczą, tym lepiej dla Polski”, który to związek był stosunkowo najsłabszy.

RYSUNEK. 1. Struktura zależności między dwiema alternatywnymi kategoriami odpowiedzi na pytanie „w jakim stopniu zgadza się P. albo nie zgadza z następującym stwierdzeniem: im mniej rząd angażuje się w sferę gospodarczą, tym lepiej dla Polski”. wynik analizy korespondencyjnej.^a



ZZg – Zdecydowanie się zgadzam	NTak – Na pewno tak
Zg – Zgadzam się	PTak – Prawdopodobnie tak
Ani – Ani się zgadzam, ani nie zgadzam	Niepewny – Nie jestem pewny/a czy tak, czy nie
NZg – Nie zgadzam się	Pnie – Prawdopodobnie nie
ZNzg – Zdecydowanie się nie zgadzam	NNie – Na pewno nie

^a Usytuowanie kategorii odpowiedzi jest wypadkową ich wartości dla dwóch wymiarów otrzymanych w analizie korespondencyjnej. Można je traktować jako współrzędne w dwuwymiarowej przestrzeni, która definiuje strukturę tej zależności.

Na wykresie tym rzuca się w oczy pięć skupień, układających się w pary odpowiedzi według wzoru znamionującego ich bliskość. Idąc z dołu do góry stwierdzamy, że blisko siebie – a daleko od innych – sytuują się odpowiedzi wyrażające najwyższą aprobatę dla zaangażowania rządu w sferę gospodarczą i to samo dotyczy pięciu pozostałych par punktów. Przykładowo więc, odpowiedź „zdecydowanie się zgadzam „sytuuje się najbliżej odpowiedzi „na pewno tak”, „zgadzam się” jest tuż obok „tak”, a z drugiej strony – najbliżej siebie położone są kategorie „zdecydowanie się nie zgadzam” i „na pewno nie”, które wyznaczają przeciwny biegun obu tych skal. Warto jest, tytułem dygresji, zwrócić uwagę na nieregularny kształt tego wykresu. Odpowiedzi na to pytanie nie tworzą jednego wymiaru. Rozrzut punktów wskazuje, że jest ich dwa. Jeden przebiegający na wykresie ukośnie, układa się wzdłuż osi łączącej odpowiedzi „Na pewno nie / zdecydowanie się nie zgadzam” i „Zdecydowanie się zgadzam / Na pewno tak”. Natomiast drugi kontrastuje te ostatnie z odpowiedziami wyrażającymi neutralne postawy „Nie jestem pewny/Ani się zgadzam, ani nie zgadzam”.

Nawet silna korelacja między pytaniami nie musi znamionować trafności, czyli tego, że mierzą one to, co powinny – zgodnie z zamierzeniami badacza. W naszym przypadku korelacja jest umiarkowana, co pozwala wnioskować, że alternatywne wersje charakteryzują się niejednakową trafnością. Spróbujemy to obecnie ustalić.

W metodologii przedstawiono wiele sposobów analizowania problemu trafności. Jeżeli chodzi o dokładność pomiaru, to stosunkowo najbardziej zaawansowane podejście przedstawił Kenneth Bollen (1989). Nawiązuje ono do schematu określanego w literaturze socjologicznej mianem *multitrait-multimethod*, którego idea opiera się na porównywaniu zależności między zmiennymi „ukrytymi” definiowanymi za pomocą różnych baterii wskaźników (Sullivan i Feldman 1979). Zastosowanie tego schematu wymagałoby uwzględnienia zmiennych, których nie ma w ESS, zresztą nie uwzględnia się ich w normalnych badaniach. Z tego też względu odwołamy się do bardziej tradycyjnego podejścia, które polega na porównaniu siły korelacji alternatywnych wskaźników z jakimś „zewnętrznym kryterium”. Wybór kryterium powinien wynikać z teorii: zakłada się, że „zewnętrzne kryterium” identyfikuje zjawisko wchodzące w związek przyczynowy ze zjawiskiem identyfikowanym przez rozpatrywane wskaźniki. Założenie to pozwala wnioskować, że wyższa korelacja stanowi charakterystykę wyższej trafności.

Odwołując się do tego schematu przyjmujemy założenie, że rozpatrywane przez nas pytania mierzą cztery zjawiska. Blok pytań dotyczących „osobistego” zaufania respondentów do parlamentu, policji i systemu prawnego identy-

fikuje postawę, która różnicuje respondentów ze względu na „zaufanie do instytucji”. Chcemy rozstrzygnąć, czy trafniejszym narzędziem jej pomiaru jest zadanie pytania z kartą czy bez karty. Zmienna „Karta” mierzona jest przez odpowiedzi na pytania B7, B8 i B9. Porównamy jej trafność ze zmienną „Bez karty”, mierzona przez odpowiedzi na pytania HF31 HF32 i HF33. „Zewnętrzny” kryterium trafności będzie tu zmienna „skłonność do populizmu” – zakładamy, że jest to silny, negatywny korelat postaw wyrażających stopień „zaufania do instytucji”, tzn. spadek zaufania do parlamentu, policji i prawa sprzyja wzrostowi populizmu, podczas gdy większemu zaufaniu towarzyszy osłabienie populistycznych nastrojów. „Skłonność do populizmu” definiujemy w postaci skali postaw, która jest wypadkową odpowiedzi na cztery pytania. Respondentów proszono o stwierdzenie na ile zgadzają się z tym, że: (i) „Prezydent nie zastanawiając się, co robi, podpisuje ustawy ułatwiające wyrzucanie ludzi na bruk”, (ii) „Prywatyzacja doprowadziła do tego, że Polacy są dziś jedynym narodem w Europie bez własnego przemysłu i handlu”, (iii) „Rząd mydli społeczeństwu oczy używając gładkich słów”, (iv) „Ludzie biedni i bezrobotni powinni odważnie walczyć o swoje prawa z rządem, nawet gdyby to było związane z łamaniem prawa”. Pytania te, skonstruowane na podstawie publicznych wypowiedzi przywódcy partii Samoobrona Andrzeja Leppera, włączyliśmy do polskiego kwestionariusza w *European Social Survey*. Odpowiadając na nie respondenci posługiwali się pięciopunktową skalą – od „zdecydowanie się zgadzam” do „zdecydowanie się nie zgadzam”.

W celu porównania trafności skal dziesięcio- i pięciopunktowych skonstruowaliśmy dwa sumaryczne wskaźniki odwołując się do pytań dotyczących zadowolenia ze stanu polskiej gospodarki, demokracji i rządu. Pierwszy z nich został utworzony z trzech skal 10-punktowych (B30, B31 i B32), drugi – ze skal pięciopunktowych HF25 HF26 i HF27. Nazwijmy je odpowiednio „Skala10” i „Skala5”. Zakładamy, że identyfikują one „zadowolenie z systemem”. Aby rozstrzygnąć, który z nich jest trafniejszym wskaźnikiem tych postaw porównamy siłę ich związku z zewnętrznym kryterium, którym będzie „skłonność do populizmu”, mierzona na skali zdefiniowanej powyżej.

To samo kryterium zastosujemy do porównania trafności pytań różniących się odpowiedziami, jeżeli chodzi o stosunek do aktywnej roli rządu i polityki socjalnej. Wybór jest między kategoriami odpowiedzi – od „zdecydowanie zgadzam się” do „zdecydowanie nie zgadzam” (zastosowanymi w pytaniach B43 B44 i B45) – a kategoriami, od „na pewno tak” do „na pewno nie” (w pytaniach HF34, HF35 i HF36). Wolno założyć, że poparcie dla interwencjonizmu państwowego i aktywnej polityki socjalnej łączy się z większą podatnością na populizm, podczas gdy krytyczny stosunek do interwencjoni-

zmu go zmniejsza. Spróbujemy rozstrzygnąć, jaka forma pytania jest trafniejszym narzędziem pomiaru przez porównanie związku między „skłonnością do populizmu” a sumarycznymi wskaźnikami „Zgadzam się” i „Tak”.

W ramach czwartego testu porównamy trafność skal dychotomicznych (pytania A8 A9, A10 – identyfikujących sumaryczną zmienną, którą nazwiemy „Skala2”) i skal dziesięciopunktowych (HF28 HF29 HF30, które są wskaźnikami sumarycznego konstruktów „Skala10bis”). Posługując się tymi skalami respondenci wyrażali swój pogląd na temat tego, czy warto ludziom ufać, postępować w życiu uczciwie i udzielać innym pomocy. Odpowiedzi te identyfikują postawę, która różnicuje jednostki ze względu na „zaufanie do innych”. Jeżeli chodzi o wybór kryterium trafności, to trudno znaleźć przesłanki do wysunięcia tezy, że zróżnicowanie na skali populizmu związane jest w znaczącym stopniu z zaufaniem i chęcią udzielania pomocy. Odwołam się do innego kryterium, a mianowicie do postawy wyrażającej stosunek do rzetelności polityków. Wolno założyć, że im bardziej się wierzy w to, że warto jest postępować uczciwie z innymi, tym lepiej ocenia się świat polityki, mimo jego nieprzejrzystości i stosunkowo niskiego prestiżu. Zmienna „ocena polityków” skonstruowana została na podstawie odpowiedzi na dwa pytania: (i) „Czy uważa P., że ogólnie rzecz biorąc politycy przywiązują wagę do tego, co myślą ludzie tacy, jak P.?” Respondenci dokonywali oceny na skali pięciopunktowej – od „prawie żaden polityk nie przywiązuje wagi do tego, co myślą tacy ludzie jak ja” do „większość polityków przywiązuje do tego wagę”; (ii) Czy uważa P., że polityków interesuje tylko uzyskanie głosów wyborców, czy raczej poznanie ich opinii?”. W tym przypadku zastosowano pięciopunktową skalę odpowiedzi od „Prawie wszystkich polityków interesuje pozyskanie głosów” do „prawie wszystkich polityków interesuje pozyskanie opinii wyborców”.

Schemat analizy polega na porównaniu siły związku między alternatywnymi wersjami pytań a przyjętymi tu kryteriami trafności. Fakt zdefiniowania zmiennych identyfikujących wersje i kryteria trafności w terminach zmiennych składowych, pociąga za sobą konieczność zastosowania technik wielowskaźnikowego pomiaru. Posłużyłem się modelami równań strukturalnych dostępnych w programie LISREL 8.3. W tabeli 3 zamieszczone są wielkości parametrów charakteryzujących siłę tych zależności – wszystkie zostały wyspecyfikowane w ramach identycznego modelu. Zmiennymi wyjaśnianymi są „skłonność do populizmu”, lub „oceny polityków”, które (przypominam) traktujemy jako kryteria trafności. Po stronie zmiennych wyjaśniających mamy za każdym razem dwie zmienne identyfikujące dwie wersje pytań, z których każda zdefiniowana jest w postaci trzech pytań zadawanych respondentom.

Tabela 3. PORÓWNANIE TRAFNOŚCI DWÓCH WERSJI PYTAŃ. WSPÓLCZYNNIKI GAMMA W MODELACH RÓWNAŃ STRUKTURALNYCH

Zmienne wyjaśniające (numery identyfikują pytania z kwestionariusza ESS)	Zmienne wyjaśniane	
	Skłonność do populizmu	Ocena polityków
Karta (B7-B9)	0,27	-
Bez karty (HF31-HF33)	0,29	-
Skala 10 (B30-B32)	0,44	-
Skala 5 (HF25-HF27)	0,04	-
Zgadzam się (B43-B45)	-0,38	-
Tak (HF34-HF36)	-0,23	-
Skala 2 (A8-A10)	-	0,39
Skala 10bis (HF28-HF30)	-	0,11

Wielkości parametrów można interpretować, tak jak interpretuje się standaryzowane współczynniki regresji. Zgodnie z terminologią LISREL będziemy je określać mianem współczynników gamma (które nie mają nic wspólnego z współczynnikami *Gamma* dla zmiennych porządkowych). Analiza ich pozwala wysnuć trzy wnioski. Pierwszy z nich rzuca światło na kwestię optymalnej długości skal. Skala dziesięciopunktowa okazuje się być trafniejszym miernikiem analizowanych tu postaw w porównaniu ze skalą pięciopunktową i odpowiedziami „tak-nie”. Stopień „zadowolenia z systemu” silniej różnicuje skłonności do populizmu, gdy operacjonalizowany jest w postaci skali dziesięcio- niż pięciopunktowej. Wielkość współczynnika gamma dla tej pierwszej wynosi 0,44, dla drugiej 0,04. To samo dotyczy „zaufania do innych” w odniesieniu do „oceny polityków”, gdzie gamma dla skali dziesięciopunktowej kształtuje się na poziomie 0,39, a dla skali dychotomicznej 0,11. Wynika stąd, że względnie trafność przemawiałaby za wyborem skal dłuższych, chociaż dokonuje się to – jak należałoby sądzić na podstawie dokonanych tu wcześniej ustaleń – kosztem generowania przez nie większej liczby braków danych i odpowiedzi „nie wiem”.

Poza zakresem naszych analiz jest socjologiczna interpretacja tych danych. Warto jest jednak uzupełnić ten metodologiczny wątek stwierdzeniem, że większe zadowolenie z istniejącego systemu przyczynia się do osłabienia populizmu, czego wskaźnikiem są dodatnie wielkości współczynników gamma. Z kolei przekonanie, że warto ludziom ufać wzmacnia pozytywny stosunek do odpowiedzialności polityków, tak jak zakładaliśmy przy formułowaniu hipotez.

Drugi wniosek dotyczy wyboru między formą odpowiedzi typu „zgadzam się” i „tak”. Z przedstawionych tu analiz wynika, że większemu poparciu dla aktywnej roli rządu towarzyszy większa skłonność do ulegania populizmowi. Przy zastosowaniu obydwu rodzajów pytań zależność ta okazuje się być znacząca, jednak silniejszym wyznacznikiem populizmu okazują się pytania

z odpowiedziami od „zdecydowanie się zgadzam” do „zdecydowanie nie zgadzam”. Uzyskana dla zmiennej „Zgadzam się” wartość współczynnika gamma wynosi $-0,38$, podczas gdy ten sam związek, mierzony przy pomocy pytań ze skalą od „na pewno tak” do „na pewno nie” był słabszy $(-0,23)^2$. Trudno formułować na tej podstawie dalej idące wnioski. Taki jest wynik dla danych ESS, natomiast niewykluczone, że przy zastosowaniu tych alternatywnych odpowiedzi do innych pytań zależności te mogłyby się kształtować odwrotnie.

Trzeci wniosek koncentruje się wokół problemu, czy posługiwanie się kartą respondenta zawsze dostarcza odpowiedzi trafniejszych? Metodologia badań terenowych każe nam wierzyć, że tak być powinno. Okazuje się, że nie zawsze tak jest. Nie miało to żadnego znaczenia dla analizowanych przez nas pytań identyfikujących zaufanie do policji, prawa i rządu. W obydwu przypadkach im większe zaufanie dla tych instytucji, tym słabszy populizm, natomiast nie było znaczącej różnicy między wersją z zastosowaniem karty respondenta i bez niej, o czym świadczy brak istotnej różnicy między wielkościami współczynników gamma $(0,27-0,29)$, które traktujemy tu jako kryterium trafności.

UWAGI KOŃCOWE

Przedstawione tu analizy można traktować jako próbę rekomendacji w kilku kwestiach związanych z konstruowaniem pytań w kwestionariuszu wywiadu. Staraliśmy się rozstrzygnąć 3 kwestie, rozpatrując je pod kątem trafności pytań i ich „operacyjności”, definiowanej przez generowanie braków danych i odpowiedzi typu „nie wiem”.

Pierwsza z tych kwestii dotyczyła wyboru między skalami odpowiedzi o różnej długości. Porównaliśmy trzy skale: dychotomiczną, pięcio- i dziesięciopunktową – ta ostatnia okazała się być stosunkowo najbardziej trafnym narzędziem pomiaru postaw dotyczących zadowolenia z systemu polityczno-ekonomicznego w Polsce i przekonania, że zaufanie do ludzi jest właściwą postawą życiową. Wzgląd na trafność przemawiałby więc za wyborem skal dłuższych, chociaż stwierdziliśmy też, że odpowiedzi na pytanie ze skalą dziesięciopunktową generowały więcej braków danych i odpowiedzi typu „nie wiem”.

² Współczynniki gamma są ujemne, ponieważ odpowiednio większe wartości na czterech zmiennych składowych mierzących „skłonności do populizmu” identyfikują rosnącą niechęć do populizmu. Tak więc malejącemu populizmowi towarzyszy spadek poparcia dla aktywnej roli rządu.

Druga z rozpatrywanych tu kwestii dotyczyła wyboru między odpowiedziami typu „zdecydowanie zgadzam się” – „zdecydowanie nie zgadzam”, a kategoriami typu „na pewno tak” – „na pewno nie”. Wystąpiły one w dwóch alternatywnych wersjach pytania dotyczącego oceny interwencjonizmu państwowego i aktywnej polityki socjalnej. Z przeprowadzonej tu analizy trafności wynika, że rozwiązaniem trafniejszym było pytanie z odpowiedziami „zdecydowanie zgadzam się” – „zdecydowanie nie zgadzam”. Odpowiedzi „zgadzam się” wydają się być bardziej adresowane do respondenta i być może z tego powodu dokładniej identyfikują postawy jednostek. Jednak byłoby ryzykowne potraktowanie tego wyniku jako generalnej rekomendacji. Nie można wykluczyć, że większa trafność tej formy pytania jest rezultatem zadziałania całkiem innego efektu, a mianowicie – umieszczenia ich na początku kwestionariusza wywiadu.

Ostatnia kwestia związana jest z posługiwaniem się kartą respondenta. Jest to rutynowy element w badaniach surveyowych prowadzonych metodą bezpośredniego wywiadu. Posługiwanie się kartą wskazane jest zwłaszcza w przypadku pytań z rozbudowaną skalą odpowiedzi, w których karta traktowana jest jako narzędzie większej precyzji pomiaru. W analizowanych przez nas pytaniach respondenci posługiwali się dziesięciopunktową skalą – przedstawioną w postaci podziałki – na której mieli wyrazić swe zaufanie do policji, prawa i rządu. Wbrew oczekiwaniom, zastosowanie karty nie przyniosło wzrostu „siły wyjaśniającej” tych pytań. Tytułem komentarza nie można stwierdzić nic więcej, jak tylko to, że nawet w przypadku sprawdzonych zasad realizacji wywiadu, mogą się zdarzyć odstępstwa od reguł.

Wnioski te nie pretendują do miana uogólnień. Niewątpliwym ich ograniczeniem jest fakt, nie że wyszliśmy poza dane dla Polski. Należy je traktować tylko jako punkt wyjścia do analiz porównawczych z innymi krajami, których celem byłoby potwierdzenie tych prawidłowości i zastosowanie ich w praktyce badawczej.

BIBLIOGRAFIA

- Bollen K. [1989], *Structural Equations with Latent Variables*, New York, John Wiley & Sons.
- Bradburn N. M., Sudman S. [1979], *Improving Interview Method and Questionnaire Design*. San Francisco, Jossey – Bass Publishers.
- Górnjak J. [2000], *My i nasze pieniądze. Studium postaw wobec pieniądza*, Kraków, Aureus.
- Hyman H. [1954], *Interviewing in Social Research*. Chicago, The University of Chicago Press.
- Koniarnek J. [1983], *Reguły poprawnego formułowania pytań kwestionariuszowych*. [w:] K. Lutyńska, A. P. Wejland (red.), *Wywiad kwestionariuszowy*. Wrocław: Ossolineum.

Ragunathan T. i in. [1997]. *A Multivariate Technique for Imputing Missing Values Using a Sequence of Regression Models*, Ann Arbor.

Sullivan J. H., Feldman S. [1979], *Multiple Indicators*, London, Sage Publications.

Sztabiński P. [1997], *Ankieterzy i ich respondenci*. Warszawa: Wydawnictwo IFiS PAN.

Sztabiński P. [2001], *Europejski Sondaż Społeczny. Opis Projektu*, „Ask. Społeczeństwo, Badania, Metody” Nr 10, s. 91–114.

Henryk Domański

Institute of Philosophy and Sociology
The Polish Academy of Sciences

EVALUATION OF ALTERNATIVE QUESTIONS IN A QUESTIONNAIRE

Summary

The analysis deals with the construction of questions in survey research. The author uses quantitative techniques to compare question alternatives taking into account quality criteria such as validity and reliability (with a special focus on missing data and ‘do not knows’). Question alternatives differ (i) in their wording, (ii) as regards the number of response categories in a scale, (iii) as to whether (or not) show cards are used. An attempt has been made to determine which version is better. Data, for Poland, come from the European Social Survey carried out in 2000.

(HF30)	Ludzie przede wszystkim dbają o własny interes	1
	Ludzie przede wszystkim starają się służyć pomocą innym	2
	(Trudno powiedzieć)	8

(B7) Na ile osobiście ma P. zaufanie do polskiego parlamentu, tzn. sejm i senatu?

Całkowity brak zaufania	00	01	02	03	04	05	06	07	08	09	Całkowite zaufanie	10	(Trudno po- wiedzieć)	88
----------------------------	----	----	----	----	----	----	----	----	----	----	-----------------------	----	--------------------------	----

(HF31)

Ocena:

(Trudno powiedzieć) 88

(B8) Na ile osobiście ma P. zaufanie do systemu prawnego?

Całkowity brak zaufania	00	01	02	03	04	05	06	07	08	09	Całkowite zaufanie	10	(Trudno po- wiedzieć)	88
----------------------------	----	----	----	----	----	----	----	----	----	----	-----------------------	----	--------------------------	----

(HF32)

Ocena:

(Trudno powiedzieć) 88

(B9) Na ile osobiście ma P. zaufanie do policji?

Całkowity brak zaufania	00	01	02	03	04	05	06	07	08	09	Całkowite zaufanie	10	(Trudno po- wiedzieć)	88
----------------------------	----	----	----	----	----	----	----	----	----	----	-----------------------	----	--------------------------	----

(HF33)

Ocena:

(Trudno powiedzieć) 88

(B30) Na ile jest P. zadowolony/a z obecnego stanu polskiej gospodarki?

Zupełnie niezadowolony/a	00	01	02	03	04	05	06	07	08	09	Całkowicie zadowolony/a	10	(Trudno powiedzieć)	88
-----------------------------	----	----	----	----	----	----	----	----	----	----	----------------------------	----	------------------------	----

(HF25)

Zupełnie niezadowolony/a										Całkowicie zadowolony/a	(Trudno powiedzieć)
0	1	2	3	4						5	8

(B31) Na ile jest P. zadowolony/a ze sposobu działania obecnego polskiego rządu?

Zupełnie niezadowolony/a												Całkowicie zadowolony/a	(Trudno powiedzieć)
00	01	02	03	04	05	06	07	08	09	10		88	

(HF26)

Zupełnie niezadowolony/a										Całkowicie zadowolony/a	(Trudno powiedzieć)
0	1	2	3	4						5	8

(B32) W jakim stopniu jest P. zadowolony/a z funkcjonowania demokracji w Polsce?

Zupełnie niezadowolony/a												Całkowicie zadowolony/a	(Trudno powiedzieć)
00	01	02	03	04	05	06	07	08	09	10		88	

(HF27)

Zupełnie niezadowolony/a										Całkowicie zadowolony/a	(Trudno powiedzieć)
0	1	2	3	4						5	8

(B43) W jakim stopniu zgadza się P. lub nie zgadza ze stwierdzeniem: im mniej rząd angażuje się w sferę gospodarczą, tym lepiej dla Polski?

Zdecydowanie się zgadzam	Zgadzam się	Ani się zgadzam, ani nie zgadzam	Nie zgadzam się	Zdecydowanie się nie zgadzam	(Trudno powiedzieć)
1	2	3	4	5	8

(HF34)	Na pewno tak	1
	Prawdopodobnie tak	2
	Nie jestem pewny/a czy tak, czy nie	3
	Prawdopodobnie nie	4
	Na pewno nie	5
	(Trudno powiedzieć)	8

(B44) W jakim stopniu zgadza się P. lub nie zgadza ze stwierdzeniem: rząd powinien podjąć działania zmierzające do zmniejszenia różnic w dochodach?

Zdecydowanie się zgadzam	Zgadzam się	Ani się zgadzam, ani nie zgadzam	Nie zgadzam się	Zdecydowanie się nie zgadzam	(Trudno powiedzieć)
1	2	3	4	5	8

(HF35)	Na pewno tak	1
	Prawdopodobnie tak	2
	Nie jestem pewny/a czy tak, czy nie	3
	Prawdopodobnie nie	4
	Na pewno nie	5
	(Trudno powiedzieć)	8

(B45) W jakim stopniu zgadza się P. lub nie zgadza ze stwierdzeniem: pracownicy potrzebują silnych związków zawodowych dla ochrony warunków pracy i poziomu płac?

Zdecydowanie się zgadzam	Zgadzam się	Ani się zgadzam, ani nie zgadzam	Nie zgadzam się	Zdecydowanie się nie zgadzam	(Trudno powiedzieć)
1	2	3	4	5	8

(HF36)	Na pewno tak	1
	Prawdopodobnie tak	2
	Nie jestem pewny/a czy tak, czy nie	3
	Prawdopodobnie nie	4
	Na pewno nie	5
	(Trudno powiedzieć)	8