

Konsekwencje wyboru typu skali odpowiedzi w badaniach ankietowych

Nadesłany: 27.10.13 | Zaakceptowany do druku: 03.01.14

Jerzy Wierziński*, **Anna O. Kuźmińska****, **Grzegorz Król*****

Podjęwane często bezrefleksyjnie decyzje dotyczące wyboru skali odpowiedzi mają istotne konsekwencje, ponieważ respondent musi transponować swoją „prywatną” skalę do różnicowania w dostarczonej przez badacza. Ta transpozycja wyznacza jego styl odpowiadania na pytania (*response style*). Analizując dane empiryczne z 5 badań, pokazaliśmy, że: (1) respondenci mają zróżnicowane preferencje co do długości skali odpowiedzi (niektórzy preferują skale krótkie {TAK, NIE}, inni wolą wielopunktowe); (2) liczby odpowiedzi skrajnych w różnych technikach pomiarowych korelują wysoko dodatnio między sobą, więc preferencję odpowiedzi skrajnych można uznać za cechę stylu odpowiadania respondenta. W artykule przedstawiliśmy też argumenty przemawiające za 4 тезami: (1) Wskaźniki zbudowane z odpowiedzi na skalach szacunkowych mogą być traktowane jako zmienne ilościowe; (2) Należy unikać odpowiedzi środkowej na skali odpowiedzi, pozwalając jednak na udzielenie odpowiedzi beztreściowej; (3) Odpowiedzi beztreściowe (nie wiem, nie mam zdania) powinny się pojawiać na końcu skali, a nie w środku; (4) Kształt rozkładu wskaźnika zbudowanego z odpowiedzi na wiele pytań zależy od liczby punktów na skali odpowiedzi.

Słowa kluczowe: skala odpowiedzi, styl odpowiadania, odpowiedzi: beztreściowe, skrajne.

Consequences of using different types of rating scales

Submitted: 27.10.13 | Accepted: 03.01.14

The choice of the type of rating scale, made often without careful consideration, has important implications, because respondents are forced to transpose the implicit scale they use to differentiate objects, into a scale provided by the researcher. Analysis of empirical data from five studies showed that: (1) respondents have different preferences as to the length of the scale (some people prefer short {YES, NO} scales, others prefer multipoint scales), (2) the numbers of extreme responses in different measurement techniques correlate highly and positively with each other, so the preference for extreme responses can be regarded as respondent's response style. The article also presented arguments for four theses: (1) an index constructed from responses on Likert-type scales can be treated as quantitative variables, (2) middle response category should be avoided, however, respondents should be allowed to provide "don't know" answers, (3) "Don't know" answer category should appear at the end of the scale - not in the middle, (4) the shape of a distribution of an index constructed from answers to multiple questions depends on the number of points in the Likert-type scale.

Keywords: rating scale, extreme response style, don't know answers.

JEL: C18

* **Jerzy Wierziński** – prof. dr hab., Wydział Zarządzania, Uniwersytet Warszawski.

** **Anna O. Kuźmińska** – mgr, Wydział Zarządzania, Uniwersytet Warszawski.

*** **Grzegorz Król** – dr, Wydział Zarządzania, Uniwersytet Warszawski.

W naukach społecznych nadal głównym źródłem danych są opinie zbierane w badaniach ankietowych. Niestety wybór skali odpowiedzi na pytania często jest dokonywany przez badacza bezrefleksyjnie. Skale odpowiedzi wykorzystywane w badaniach różnią się m.in. długością i obecnością (lub nie) odpowiedzi środkowej i beztreściowej („nie mam zdania”, „trudno powiedzieć”) (np. Wieczorkowska i Wierziński, 2011). Przykłady szacunkowych skal odpowiedzi znajdują się poniżej:

1	2	3	4	5	8
Zdecydowanie TAK	Raczej TAK	Ani TAK, ani NIE	Raczej NIE	Zdecydowanie NIE	NIE WIEM

1	2	3	4	8
Zdecydowanie TAK	Raczej TAK	Raczej NIE	Zdecydowanie NIE	NIE WIEM

1	2	8
TAK	NIE	NIE WIEM

1. Poziom pomiaru szacunkowych skal odpowiedzi

Puryści metodologiczni twierdzą, że opisana poniżej skala szacunkowa nie jest skalą ilościową, ponieważ nie ma **stałej** jednostki pomiaru: różnica, np. między {zdecydowanie negatywna} a {negatywna}, **nie jest** taka sama jak różnica między {negatywna} i {pozytywna}:

1	2	3	4
zdecydowanie negatywna	negatywna	pozytywna	zdecydowanie pozytywna

Jednocześnie powszechnie używa się średniej ocen, np. na studiach, a przecież nikt nie jest w stanie wykazać, że różnica między {2} i {3} równa się różnicy między {4} i {5}.

2	3	4	5	6
Ocena niedostateczna	Ocena dostateczna	Ocena dobra	Ocena bardzo dobra	Ocena celująca

W naukach fizycznych, ekonomicznych, itp., przedmiotem pomiaru są *byty realne* (rzeczywiste wielkości, np. długość, temperatura, ciężar, ciepło,

wzrost produkcji). Jest ustalony wzorzec, np. metra, i możemy obliczyć jak dobrze nasz pomiar odpowiada wzorcowi. W naukach społecznych przedmiotem pomiaru są *byty wirtualne* (konstrukty teoretyczne, np. reprezentacje poznawcze, postawy, cechy, stany afektywne). Istnieje podstawowa różnica między pomiarem wielkości fizycznych a pomiarem konstruktów teoretycznych. Modele pomiarowe wypracowane w naukach fizycznych nie przystają w pełni do pomiaru w naukach społecznych. Nie mamy wzorca zaangażowania w pracę ani inteligencji, z którym moglibyśmy wyniki naszego pomiaru porównać. Ekonomista jest zainteresowany ilością pieniędzy dostępnych na rynku, więc będzie starał się mierzyć obiektywną wielkość zarobków. W badaniach społecznych, nawet jeśli prosimy o podanie wielkości zarobków w złotych, to tak naprawdę interesujemy się reprezentacją poznawczą zarobków w umyśle respondenta.

Zarobki, wyrażone np. w złotych, są czystym przykładem stosunkowej skali pomiarowej (z precyzyjnie zdefiniowanym zerem i jednostką pomiaru). W wielu zastosowaniach metodologowie zalecają jednak stosowanie przekształceń, takich jak logarytm, zmieniających pierwotną stałość jednostki pomiaru. Przekształcenie logarytmiczne redukuje skośność i jest korzystne dla wykonywanych analiz, ale różnica między logarytmami z 4000 [3,6] i 2000 zł [3,3], nie jest równa różnicy między logarytmami 3000 [3,48] z 1000 [3] zł. Badacze społeczni wiedzą, że zachowanie czy poglądy respondenta wyjaśnia nie tyle wielkość obiektywna dochodu, ile jego wartość subiektywna. Obiektywne zmiany pensji (ważne dla ekonomisty) mogą być przez respondenta różnie traktowane, np. w zależności od tego czy podwyższono mu pensję z 800 do 1000 zł, czy też z 12 200 do 12 400 zł. Już dawno stwierdzono, że wartość subiektywna nie jest liniową funkcją wysokości zarobków (por. np. Wierzbński, 2007). Oznacza to, że stała jednostka stosunkowej skali pomiarowej <zarobki> jest reprezentowana w umyśle respondenta w odmienny sposób w zależności od pozycji wartości zmiennej na skali. Wzorcowa dla rygorystów metodologicznych stosunkowa skala pomiarowa po bliższej analizie też zaczyna budzić wątpliwości dotyczące stałości jednostki pomiaru.

W naukach społecznych mierzy się cechy nieobserwowalne, takie jak ekstrawersja, zaangażowanie w pracę itp., zadając zestaw pytań związanych, co najmniej fasadowo, z operacjonalizowaną zmienną. Wyjaśnijmy to na przykładzie pomiaru ekstrawersji, definiowanej w następujący sposób: „Na ekstrawersję składają się: towarzyskość, żywość, aktywność, asertywność, poszukiwanie doznań. Ekstrawertycy są przyjacielscy i serdeczni, towarzyscy i rozmowni, skłonni do zabawy oraz poszukiwania stymulacji. Wykazują tendencje do dominowania w kontaktach społecznych i są życiowo aktywni oraz pełni wigoru. Wykazują optymizm życiowy i pogodny nastrój. Introwertycy są ich przeciwieństwem” (Strelau, 2004).

Możemy więc zakładać, że ekstrawertycy powinni odpowiadać na poniższe pytanie inaczej niż introwertycy:

Czy w opisanej poniżej sytuacji zachowałbyś się:

- a) tak jak osoba A,
- b) raczej jak osoba A,
- c) tak jak osoba B,
- d) raczej jak osoba B,
- e) trudno powiedzieć.

W nowym towarzystwie **osoba A** natychmiast nawiązuje nowe znajomości. Musi minąć trochę czasu, nim **osoba B** poczuje się swobodnie w rozmowie z obcymi.

Na mocy definicji ekstrawersji zakładamy, że respondent, który stwierdził, że jest bardziej podobny do osoby A (*w nowym towarzystwie natychmiast nawiązuje nowe znajomości*), lokuje się na wymiarze ekstrawersji w innym miejscu niż respondent, który stwierdził, że jest bardziej podobny do osoby B (*potrzebuje czasu, aby poczuć się swobodnie*). Odpowiedź na to pytanie pozwala nam uporządkować respondentów na wymiarze ekstrawersji. Rygorystyci metodologiczni powiedzieliby, że jest to jedynie skala porządkowa. Przypisanie poszczególnym odpowiedziom liczb, np. w następujący sposób:

1 = tak jak osoba A	2 = raczej jak osoba A
4 = tak jak osoba B	3 = raczej jak osoba B

oznacza, że zakładamy równość różnicy m.in. między odpowiedzią <1> i <2> (<tak jak A> i <raczej jak A>) oraz <2> i <3> (<raczej jak A> i <raczej jak B>).

Czy możemy założenie o równości przedziałów (odległości między kolejnymi punktami skali) udowodnić? Nie, ale w naukach społecznych tego rodzaju skale są traktowane jako ilościowe i wykonujemy analizy tego typu danych, np. licząc średnią lub współczynniki korelacji. Trzeba pamiętać, że w naukach społecznych najczęściej budujemy jeden wskaźnik z wielu pytań, czyli sumujemy/uśredniamy wyniki z kilku, czasem nawet kilkudziesięciu skal odpowiedzi. Tak otrzymana skala daje nam więcej informacji niż tylko porządek i dlatego powszechnie jest traktowana jako skala ilościowa.

Podsumowując, założenie, że wyniki odpowiedzi na skalach szacunkowych, takich jak omawiana powyżej, stanowią dane ilościowe, jest akceptowane w najlepszych zagranicznych czasopismach prezentujących wyniki analiz badań społecznych. Podobnie traktowane są odpowiedzi na pytania dotyczące zgadzania się z danym poglądem, np.:

Zdecydowanie TAK	TAK	Raczej TAK	Raczej NIE	NIE	Zdecydowanie NIE
---------------------	-----	---------------	---------------	-----	---------------------

We wszystkich prowadzonych przez nas analizach danych sondażowych okazywało się, że niedoskonałe skale szacunkowe, np. satysfakcji finansowej, są lepszym predyktorem, np. adaptacji do zmiany systemowej czy satysfakcji z różnych dziedzin życia, niż wyrażona na eleganckiej skali stosunkowej wielkość zarobków.

2. Odpowiedzi beztreściowe na skali odpowiedzi

Braki odpowiedzi są zmurą każdego badania. Oprócz odmowy odpowiedzi lub pominięcia kolejnego pytania, dużo problemów dostarczają odpowiedzi: <nie wiem> czy <trudno powiedzieć>, które dalej będą nazywane (Sułek, 2002) *odpowiedziami beztreściowymi*. Odpowiedzi beztreściowe pojawiać się mogą w danych ankietowych w trzech typach pytań:

- *pytanie standardowe* (odpowiedź beztreściowa nie jest odczytywana przez ankietera, natomiast jest zapisywana, jeśli respondent udzieli jej spontanicznie);
- *pytanie z quasi-filtrem* (możliwa odpowiedź beztreściowa jest odczytywana respondentowi);
- *pytanie z pełnym filtrem* (pytanie właściwe zadaje się tylko tym respondentom, którzy odpowiedzieli twierdząco na pytanie filtrujące: „Czy masz zdanie w tej sprawie?”, „Czy zastanawiałeś się nad tym?” itp.).

W samowypełnialnych ankietach pytania standardowe oznaczają brak możliwości udzielenia odpowiedzi beztreściowej. Każdy z powyższych sposobów zadawania pytania daje inne rozkłady odpowiedzi beztreściowych, w zależności od treści pytania, stopnia krystalizacji opinii respondenta i postrzeganych przez niego oczekiwań społecznych (poprawność społeczno-polityczna opinii).

Interpretacja odpowiedzi beztreściowych przysparza wielu problemów. Odpowiedzi te mogą wynikać zarówno z braku wiedzy i/lub zdecydowania, znużenia wywiadem, jak i z chęci ucieczki respondenta przed odpowiedzią na niewygodne dla niego pytanie. Ale też mogą sugerować, że skala odpowiedzi nie uwzględnia odpowiedniej opcji. Liczba odpowiedzi beztreściowych zależy od charakteru pytania – jest ich mniej, gdy pytamy o łatwo obserwowalne zachowania. Na przykład w pytaniach o to, jak często modli się respondent (Wierzbiński, 2009), na skali od „nigdy” do „parę razy dziennie” odpowiedzi beztreściowych było 5,4%, gdy zaś pytano o częstość uczestniczenia w nabożeństwach – tylko 1,4%.

Badania (Krosnick i in., 2002) wykazały, że liczba odpowiedzi beztreściowych koreluje negatywnie z miarą samooceny wysiłku wkładanego w odpowiadanie przez respondenta na pytania, co sugeruje, że uchylanie się od odpowiedzi wynika często z lenistwa. Niebezpieczna jest także motywacja przeciwna, gdy respondent czuje się zobowiązany do udzielenia treściowej odpowiedzi. Wstyd przed przyznaniem się do niewiedzy może powodować wybieranie odpowiedzi treściowych losowo, zgodnie z domniemywanym oczekiwaniem badacza lub uznawaną normą społeczną, o czym przekonują cyto-

wane wcześniej badania pokazujące formułowanie przez respondentów opinii na temat nieistniejących obiektów. Takie odpowiedzi Sufek nazywa „pustymi”.

Autorzy sondaży obawiają się zarówno zbyt dużej, jak i zbyt małej liczby odpowiedzi beztreściowych i dlatego starają się konstruować pytania w taki sposób, aby nie skłaniać respondentów do generowania zarówno odpowiedzi „pustych”, jak i fałszywie beztreściowych.

2.1. Liczba odpowiedzi beztreściowych jako cecha pytania i/lub cecha respondenta

Brakujące dane można analizować jako cechę pytania lub cechę respondenta. Duża liczba odpowiedzi <nie wiem> lub <trudno powiedzieć> oznaczonych w tekście jako <TP> dla któregoś z pytań może być wskaźnikiem błędów w sformułowaniu pytania i dlatego zawsze powinna być przedmiotem dociekań. Czasem warto takie pytanie wyłączyć z dalszych analiz. W badaniach zaufania do 18 instytucji, analizując liczbę odpowiedzi beztreściowych jako cechę pytania (w tym przypadku instytucji), stwierdzimy, że w omawianym przykładzie najwięcej odpowiedzi <trudno powiedzieć> pojawia się w przypadku związków zawodowych (OPZZ i „Solidarność”)¹. Budując wskaźnik zaufania organizacyjnego, można się zastanawiać, czy nie pominąć tych 2 pytań.

Warto także przeanalizować rozkład odpowiedzi beztreściowych traktowanych jako cecha respondenta, czyli zliczać, ile razy każdy respondent wybrał w zestawie pytań (np. 18 instytucji) odpowiedź <TP>. W analizowanym przykładzie tylko 68,2% respondentów ani razu nie uchyliło się odpowiedzi. Gdybyśmy potraktowali odpowiedzi beztreściowe jako braki danych, w wielu analizach (np. analizie czynnikowej) stracilibyśmy ponad 30% próby. Zdecydowanie lepszym rozwiązaniem jest przekodowanie odpowiedzi beztreściowych na środkową wartość skali. Ale zanim to zrobimy, musimy przeanalizować rozkład. W analizowanym przykładzie 67 osób (1,1% badanej próby) udzieliło ponad 50% odpowiedzi beztreściowych (więcej niż 9 na 18 pytań). Warto zastanowić się, czy nie powinniśmy usunąć ich z analiz, ponieważ można sądzić, że tematyka pytań ich nie interesuje lub nie mają opinii/wiedzy na temat większości instytucji. Sprawdzenie rozkładów odpowiedzi beztreściowych, zarówno dla pytań, jak i dla respondentów, powinno być wstępem do kolejnych analiz.

Liczba odpowiedzi beztreściowych zależy od ich dostępności. Pokazuje to przeprowadzone przez zespół PGSS (Cichomski, Jerzyński i Zieliński, 2010) badanie eksperymentalne, w którym zadano następujące pytania:

Zawsze spotyka się jakieś osoby, których poglądy inni ludzie uważają za niewłaściwe lub niebezpieczne. Weźmy na przykład kogoś, kto jest przeciw Kościołom i religii.

A) Jeśli taka osoba chciałaby wygłosić w Pana mieście/gminie przemówienie przeciwko Kościołom i religii, czy powinno się jej na to pozwolić, czy też nie?

B) Czy takiej osobie powinno się pozwolić uczyć w wyższej uczelni, czy też nie?

C) Jeśli jacyś ludzie z Pana otoczenia chcieliby usunąć z publicznej biblioteki książkę wymierzoną przeciwko Kościołom i religii napisaną przez tę osobę, to czy byłby Pan za usunięciem tej książki, czy przeciwko jej usunięciu?

Połowa respondentów (balot B) miała odpowiedź „nie wiem” na karcie odpowiedzi (pytanie z quasi-filtrem), druga połowa (balot A) otrzymywała kartę tylko z odpowiedziami treściowymi (np. „tak, powinno się pozwolić” i „nie, nie powinno się pozwolić”). Odpowiedź „nie wiem” była zapisywana przez ankietera, gdy respondent udzielał takiej odpowiedzi spontanicznie. Dostępność odpowiedzi beztreściowej spowodowała we wszystkich pytaniach wzrost liczby osób uchylających się od zajęcia stanowiska (dla pytania A: 10,3% → 17,6%; dla pytania B: 13,8% → 22,5%; dla pytania C: 16,5% → 26,9%). Nie ma powodu, by sądzić, że respondenci w obu grupach różnili się pod względem opinii, ponieważ podział grup na baloty był losowy.

Czasem zdarza się, że badacz traktuje odpowiedź beztreściową jako odpowiedź środkową i umieszcza ją w środku skali odpowiedzi. W celu sprawdzenia tego efektu, przeprowadzono badanie eksperymentalne, w którym dwie losowo wyodrębnione grupy respondentów odpowiadały na pytania różniące się skalą odpowiedzi. Jak widać w tabeli 1, umieszczenie opcji beztreściowej w środku skali odpowiedzi drastycznie zwiększa liczbę osób ją wybierających.

	Wersja A					Wersja B				
	1	2	3	4	5	1	2	3	4	5
Mam wokół siebie wiele bliskich mi osób	4,8	9,6	10,8	43,4	31,3	6,8	18,2	47,7	22,7	4,5
Zwykle pozostaję sam ze swoimi problemami	33,7	33,7	13,3	14,5	4,8	31,8	38,6	15,9	9,1	4,5
W ważnych sprawach mogę liczyć na pomoc przyjaciół	4,8	3,6	9,6	32,5	49,4		18,2	22,7	52,3	6,8
Niewiele wiem o ludziach, którzy mnie otaczają	51,8	38,6	7,2	2,4		52,3	25,0	9,1	4,5	9,1
Dużo czasu spędzam ze znajomymi i przyjaciółmi	7,2	21,7	20,5	38,6	12,0	4,5	31,8	36,4	18,2	9,1
Często czuję się samotny	27,7	39,8	8,4	18,1	6,0	31,8	40,9	13,6	9,1	4,5

Kafeteria odpowiedzi:

wersja A: [1] – zdecydowanie tak, [2] – tak, [3] – **trudno powiedzieć**, [4] – nie, [5] – zdecydowanie nie,

wersja B: [1] – zdecydowanie tak, [2] – tak, [3] – nie, [4] – zdecydowanie nie, [5] – **trudno powiedzieć**.

Tab. 1. Liczba odpowiedzi beztreściowych w zależności od miejsca opcji beztreściowej na skali odpowiedzi (badania eksperymentalne). Źródło: G. Wieczorkowska, J. Wierzbński i M. Siarkiewicz. (2009). *Wybrane problemy metodologiczne analitycznych badań sondażowych*. W: M. Zahorska i E. Nasalska (red.). *Wartości – polityka – społeczeństwo*. Warszawa: Wydawnictwo Naukowe Scholar.

3. Wpływ punktu środkowego skali

Wybierając skalę odpowiedzi, stajemy przed dylematem, czy ma ona mieć *punkt środkowy* (nazywany dalej *odpowiedzią środkową*), np. <ani zgadzam się, ani nie zgadzam się> oraz czy powinna się na niej pojawić opcja odpowiedzi beztreściowej. Warto zwrócić uwagę, że zarówno odpowiedzi środkowe, jak i beztreściowe są niemożliwe do zinterpretowania. Z tego powodu niektórzy badacze obstają przy tym, aby w ogóle nie pojawiały się one na skali – respondenci będą musieli dokonać wyboru jednej „konkretnej” opcji, dzięki czemu otrzymamy bardziej rzetelne dane. Zrównoważona skala odpowiedzi wymaga ustalenia opcji środkowej i wyznaczenia jednakowej liczby odpowiedzi pozytywnych i negatywnych. Zapewnienie opcji środkowej jest ważne dla respondentów, którzy mogą nie mieć wyrobionej opinii (Kaden, 2008). Dla tych, których postawy są skryształizowane, lepsze są skale bez opcji środkowej (Brzezińska i Brzeziński, 2004).

Środkowe odpowiedzi wybierane są przez dwie kategorie respondentów: *niedoinformowanych* – tych, którym brak wiedzy potrzebnej do sformułowania opinii, i *niezdecydowanych* – tych, którzy tę wiedzę posiadają, ale, rozważywszy

	Wersja A						Wersja B				
	1	2	3	4	5	TP	1	2	3	4	TP
Wolę być obywatelem Polski niż jakiegokolwiek innego kraju na świecie	31,0	31,5	19,0	11,1	5,1	2,0	34,0	29,3	15,7	8,4	12,6
Są pewne sprawy w dzisiejszej Polsce, których jako Polacy możemy się wstydzić	45,4	42,6	6,9	2,3	1,4	1,0	44,5	41,4	6,3	3,7	4,2
Ogólnie rzecz biorąc, Polska jest lepszym krajem niż większość innych krajów	9,7	20,4	32,4	18,5	13,9	5,0	9,4	21,5	30,9	14,7	23,6
Trzeba popierać swój kraj, nawet gdy postępuje niewłaściwie	7,4	9,3	30,1	40,7	9,3	3,0	8,4	12,6	29,8	38,2	11,0
Chciałbym być dumny z Polski częściej niż jestem	53,2	31,9	6,5	3,7	4,2	1,0	54,5	29,3	5,2	1,6	9,4

Kafeteria odpowiedzi:

wersja A: [1] – zdecydowanie się zgadzam, [2] – zgadzam się, [3] – **ani się zgadzam ani nie zgadzam**, [4] – nie zgadzam się, [5] – zdecydowanie się nie zgadzam, [TP] – **trudno powiedzieć**,

wersja B: [1] – zdecydowanie się zgadzam, [2] – zgadzam się, [3] – nie zgadzam się, [4] – zdecydowanie się nie zgadzam, [?] – trudno powiedzieć.

Tab. 2. Rozkłady odpowiedzi uzyskanych za pomocą skali z punktem środkowym i bez. Źródło: G. Wieczorkowska, J. Wierziński i M. Siarkiewicz. (2009). *Wybrane problemy metodologiczne analitycznych badań sondażowych*. W: M. Zahorska i E. Nasalska (red.). *Wartości – polityka – społeczeństwo*. Warszawa: Wydawnictwo Naukowe Scholar.

argumenty za i przeciw, nie potrafią zadeklarować, która z dostępnych opcji odpowiada im najbardziej (Sułek, 2001). Dodanie wówczas wyraźnej odpowiedzi beztreściowej przeznaczonej dla niedoinformowanych (np. „nie znam się na tym na tyle, aby mieć pogląd”) nie zmienia odpowiedzi tych respondentów, którzy wybrali którąś z opinii zdecydowanych, ale redukuje liczbę wskazań środkowego punktu skali.

Aby porównać konsekwencje umieszczenia na skali punktu środkowego dla liczby odpowiedzi beztreściowych, przeprowadzono badanie eksperymentalne. W badaniu dwóch równoważnych (podzielonych losowo) grup studentów wykorzystano dwie skale odpowiedzi (zob. tabela 2). Tylko w wersji A dostępna była opcja środkowa: <ani zgadzam się, ani nie zgadzam się>. Jeżeli sądzilibyśmy, że odpowiedź beztreściową wybierają osoby niezdecydowane, to procent odpowiedzi TP w wersji B powinien być większy niż procent osób wybierających opcję środkową w wersji A. Tak było tylko w jednym z pięciu analizowanych pytań.

Dostępność na skali odpowiedzi opcji zarówno środkowej, jak i beztreściowej powoduje wzrost liczby odpowiedzi nieinterpretowalnych. Dlatego należy w badaniach unikać odpowiedzi środkowej, ale udostępniać odpowiedź <trudno powiedzieć>, którą można następnie przekodować w środek skali odpowiedzi.

4. Wpływ długości skali odpowiedzi

Jak wspomnieliśmy wcześniej, bardzo często, układając kwestionariusz, decyząc o rodzaju skali odpowiedzi podejmujemy bezrefleksyjnie. Jedni preferują *skale długie* (np. 101-punktowa Michigan Feeling Thermometer czy 10-punktowa Gallupa), inni proszą wyłącznie o odpowiedź „tak” lub „nie”. Przykład: osoba może oceniać stopień kontyngencji (związku między działaniami a wynikami) w danej sytuacji poprzez zakreślenie punktu na skali: <-4, ..., +4> (Wasserman i Shaklee, 1984). Inni badacze (Shank, 1985) proszą o ocenę stopnia kontyngencji na skali: <-100, ..., +100>. Wpływ długości skali był przedmiotem wielu badań, na podstawie których sformułowano opisane poniżej wnioski:

- Gdy badano (Green i Rao, 1970) wpływ zmiany liczby kategorii na stopień odtworzenia początkowej konfiguracji przestrzennej za pomocą skalowania wielowymiarowego, okazało się, że najlepsze wyniki dają skale co najmniej 6-punktowe (dla co najmniej 8 pytań związanych z jednym wymiarem).
- W innych badaniach (Benson, 1971) pokazano, że skale krótkie (2-, 3-punktowe) są użyteczne i nie powinny być eliminowane.
- W badaniach (Peterson, 1985) Generalnego Sondażu Społecznego, porównanie pytań o zaufanie do organizacji ze skalą 3-punktową (*a great deal of confidence; only some confidence; hardly some confidence*) vs 7-punktową (od *complete confidence* do *no confidence at all*) nie wykazały różnic.
- Porównywano również (Wieczorkowska, Wierzbński i Siarkiewicz, 2009) wpływ długości skali (5 vs 9 punktów) na odpowiedzi na czterech skalach skróconej wersji kwestionariusza kolektywizmu/indywidualizmu.

Wskaźniki w obu porównywanych grupach nie różniły się własnościami psychometrycznymi (wariancje, interkorelacje).

Rozwiązaniem tych sprzecznych rekomendacji jest propozycja (Lehmann i Hulbert, 1972), aby liczba punktów na skali zależała od celu badania. Jeżeli interesujemy się średnimi dla grup respondentów, skale dwu-, trzy-punktowe są wystarczające. Jeżeli interesują nas wyniki pojedynczych osób, skale powinny być pięcio-, siedmiopunktowe. Należy pamiętać, że:

- zwiększanie liczby punktów na skali redukuje błąd zaokrąglenia (*rounding error*), ale zwiększa koszty przetwarzania danych i zmęczenie respondenta, co może owocować brakami odpowiedzi;
- jak wspomnieliśmy wcześniej, zazwyczaj rekomenduje się magiczną w psychologii pamięci cyfrę 7 (+/-2) – większa liczba punktów (np. często badacze preferują skalę procentową) zmniejsza rzetelność; w pytaniach o procenty ludzie wybierają znaczące punkty będące wielokrotnościami liczby 5, pomijając resztę, czyli efektywna liczba punktów skali jest mniejsza niż ta, którą przedstawia się respondentowi.

5. Styl odpowiadania (*response style*)

To, że respondent otrzymuje 5-punktową skalę odpowiedzi, np. przy ocenie ważności różnych celów życiowych, nie oznacza, że w taki właśnie sposób dokonuje ich „prywatnego” różnicowania. Jego wartościowanie może być dużo bardziej zgrubne (np. „ważne” vs „nieważne”) lub dużo bardziej subtelne. Ktoś może uważać, że „nic nie jest warte zachodu” i odpowiadać na pytanie, używając tylko lewego krańca skali (1 = „zupełnie nieważne”, 2 = „nieważne”), ktoś inny używa tylko opcji 3 i 4 („ważne” i „bardzo ważne”), inna osoba może wykorzystywać wszystkie opcje (Wieczorkowska, 1992). Oznacza to, że każdy respondent musi transponować swoją „prywatną” skalę do różnicowania w dostarczonej przez badacza skali odpowiedzi. Ta transpozycja wyznacza jego *styl odpowiadania na pytania* (*response style*), który może być uniwersalny, tzn. w niewielkim stopniu zależny od tego, jaki obiekt jest aktualnie oceniany.

5.1. Przykład 1. Transformacja skali odpowiedzi zachodząca w umysłach respondentów

W badaniu internetowym studenci kursu metodologicznego mieli za zadanie odpowiadać na to samo pytanie zaczerpnięte z PGSS („Sumiennie pracuję, aby wypełnić moje codzienne obowiązki, nawet jeśli jestem trochę chory lub jest jakiś inny uzasadniony powód, abym zrobił sobie przerwę”) za pomocą różnych skal odpowiedzi. Odpowiedzi te były przedzielone zadaniami arytmetycznymi, aby pozwolić na wyczyszczenie pamięci roboczej z udzielonej wcześniej odpowiedzi.

W tabeli 3 pokazane jest, jak uczestnicy (N = 81) odpowiadali na to pytanie, gdy skala odpowiedzi była 8-punktowa (zawierała 8 opcji do wyboru, od <zdecydowanie TAK> do <zdecydowanie NIE>, z opcją środkową i odpowiedzią <TP>, czyli trudno powiedzieć).

Sumiennie pracuję, aby wypełnić moje codzienne obowiązki, nawet jeśli jestem trochę chory lub jest jakiś inny uzasadniony powód, abym zrobił sobie przerwę	Częstość	Procent
zdecydowanie NIE	2	2,5
NIE	0	0,0
raczej NIE	13	16,0
ani TAK, ani NIE	10	12,3
raczej TAK	23	28,4
TAK	24	29,6
zdecydowanie TAK	7	8,6
trudno powiedzieć	2	2,5

Tab. 3. Rozkłady odpowiedzi na ośmiopunktowej skali odpowiedzi. Źródło: opracowanie własne.

Gdy skala odpowiedzi została zredukowana do 3 punktów: <TAK> <NIE> <TP>, osoby zaznaczające na skali ośmiopunktowej odpowiedzi <zdecydowanie TAK> [7 osób] i <TAK> [24 osoby] wybrały <TAK>, natomiast dwie osoby, które wybrały wcześniej <zdecydowanie NIE>, wybrały <NIE>. Interesujące zmiany widoczne są w przypadku pierwotnych odpowiedzi <raczej NIE> i <raczej TAK>. Wśród 13 osób, które na skali ośmiopunktowej wybrały odpowiedź <raczej NIE>, przy skali trzypunktowej aż 6 zmieniło ją na <TP>, natomiast 7 na <NIE>. Przy skali dwupunktowej (bez możliwości udzielenia odpowiedzi beztreściowej TP) 13 osób (czyli wszystkie) opowiedziało <NIE>. Odpowiedź <raczej TAK> wybrały przy ośmiopunktowej skali 23 osoby. Przy skali trzypunktowej 3 z nich zazaczyły <TP>, przy dwupunktowej 22 osoby wybrały odpowiedź <TAK>, a jedna – paradoksalnie – <NIE>.

Odpowiedź <ani TAK, ani NIE> na skali ośmiopunktowej zaznaczyło 10 osób. Kiedy ograniczono skalę do trzypunktowej, 8 osób wybrało <TP>, 2 zdecydowały się na odpowiedź <NIE>. Przy skali dwupunktowej 8 z początkowych 10 osób opowiedziało <NIE>, natomiast 2 osoby <TAK>. Podobnie osoby wybierające – gdy było to możliwe – odpowiedź <TP> przy skali dwupunktowej wybrały <NIE>. Ograniczenie liczby punktów skali odpowiedzi wymusiło więc na nich zajęcie stanowiska.

Na odpowiedzi respondenta może wpływać potrzeba aprobaty społecznej (szczególnie silna w czasie wywiadów), przejawiająca się w *tendencji do potakiwania* (*Acquiescence Response Style*) lub, wręcz przeciwnie, potrzeba prezentowania siebie jako osoby nonkonformistycznej, przejawiająca się w *tendencji do zaprzeczania* (*Disacquiescence Response Style*) (Harzing, 2006). W literaturze jest także omawiany *styl potakiwania netto* (*Net Acquiescence Response Style*), obliczany jako różnica w tendencji do potakiwania i zaprzeczania (Baumgartner i Steenkamp, 2001).

Innym stylem odpowiadania jest transformacja wewnętrznej skali w skrajne wartości skali odpowiedzi. Tendencja do udzielania skrajnych odpowiedzi bywa interpretowana (Baumgartner i Steenkamp, 2001) jako brak tolerancji dla wieloznaczności. Można przypuszczać, że jest przejawem słabiej wyartykułowanych struktur poznawczych i jest silniejsza dla ważnych, angażujących emocjonalnie bodźców. Strach przed nietrafnością może prowadzić do udzielania większej liczby odpowiedzi środkowych. Opiswane w literaturze porównania międzykulturowe preferencji dla udzielania skrajnych odpowiedzi pokazały, że więcej odpowiedzi skrajnych udzielają:

- w USA zarówno Latynosi, jak i Afroamerykanie niż Amerykanie pochodzenia europejskiego (Clarke III, 2000);
- Amerykanie i Kanadyjczycy niż Japończycy (Shiomi i Loo, 1999; Takahashi, Ohara, Antonucci i Akiyama, 2002);
- Amerykanie niż Koreańczycy (Chun, Campbell i Yoo, 1974; Lee i Green, 1991);
- Grecy niż Włosi i Hiszpanie, którzy z kolei udzielają ich więcej niż Brytyjczycy, Niemcy i Francuzi (Van Herk, Poortinga i Verhallen, 2004);
- Włosi niż Japończycy (Wieczorkowska-Wierzińska, Wierziński i Kuźmińska, 2014).

Skoro stwierdzono istotne różnice międzykulturowe, to zadaliśmy sobie pytanie, czy w ramach tej samej kultury warto analizować różnice indywidualne.

5.2. Przykład 2. Preferowanie odpowiedzi skrajnych

Analizie poddano zbiory danych pochodzące z trzech badań. W badaniach pracowników różnych firm, **A** [$N = 111$ (78,4 % kobiet) w wieku od 21 do 71 lat ($M = 37,83$; $s = 10,42$)] i **B** [$N = 129$ (25,6% kobiet) w wieku od 25 do 65 lat ($M = 37,93$; $s = 6,96$)], analizowano odpowiedzi na pytania 4 kwestionariuszy wykorzystywanych w Sondażu Stylów Aktywności (Jeśka, 2014): „Style Działania”, „Typy Osobowości”, „Wykonywana Praca”, „Wymarzona Praca”. W badaniu **C** (Mieczkowska, 2013) analizowano odpowiedzi uczniów ($N = 186$) drugiej i trzeciej klasy gimnazjum w dwóch kwestionariuszach: „Emocje Nastolatków”, „Postawy Nastolatków” – zbiór C_1 ($N = 186$) i powtórzony pomiar postaw C_2 ($N = 99$).

Dla każdego respondenta policzono liczbę wybranych przez niego odpowiedzi skrajnych w wypełnianych przez niego kwestionariuszach, składających się z dużej liczby pytań. Skale odpowiedzi różniły się liczbą punktów oraz opisem możliwych do wybrania odpowiedzi. Jeśli skala odpowiedzi miała tylko opisane końce: (1) całkowicie nieprawdziwe – (6) całkowicie prawdziwe, to liczba odpowiedzi skrajnych oznaczała, ile razy respondent zaznaczył 1 lub 6. Jeśli pytania dotyczyły podobieństwa do zachowania osoby A vs osoby B na skali odpowiedzi opisanej <zazwyczaj jak A>, <częściej jak A>, <częściej jak B>, <zazwyczaj jak B>, <trudno powiedzieć>, to liczba odpowiedzi skrajnych oznaczała, ile razy respondent zaznaczył <zazwyczaj jak A> lub <zazwyczaj jak B>.

1	2	3	4		
				Bad. A	Bad. B
Liczba pytań	Przykładowe pytanie	Skala odpowiedzi			
Style działania k = 48 w Badaniu A k = 50 w Badaniu B	Czy Twoje zachowanie w sytuacji opisanej w pytaniu byłoby bardziej zbliżone do reakcji osoby A czy też osoby B. • Osoba A stara się realizować wiele celów równocześnie („piec wiele pieczeni przy jednym ogniu”). Osoba B woli je realizować pokolei.	• zazwyczaj tak jak A • częściej tak jak A • częściej tak jak B • zazwyczaj tak jak B • trudno powiedzieć	M	14,21	18,31
			Me	14	18
			s	14,21	12,66
			min	0	0
			max	38	50
			0,25	7	8
0,75	20,75	25			
Typy osobowości k = 20 w badaniu A k = 37 w badaniu B	A: Prosimy, abyś ocenił w jakim stopniu charakterystyka opisuje Ciebie. • Niektóre osoby lubią podejmować się funkcji kierowniczych – nawet jeśli nie jest to związane ze wzrostem wynagrodzenia. B: Określ, na ile poniższe stwierdzenia są prawdziwe w odniesieniu do Ciebie. • Wywieranie wpływu na innych sprawia mi satysfakcję.	A: z opisaniem tylko końcami: 1 – w ogóle 7 – w bardzo dużym stopniu B: z opisaniem tylko końcami: 1 – całkowicie nieprawdzie 6 – całkowicie prawdziwe	M	2,89	13,5
			Me	2	11
			s	2,58	10,5
			min	0	0
			max	16	37
			0,25	1	6
0,75	4	18			
Wykonywana praca k = 37 w badaniu A k = 30 w badaniu B	Praca, którą wykonuję, wymaga: • realizacji szczegółowych procedur	• nigdy • bardzo rzadko • rzadko • często • bardzo często • zawsze • trudno powiedzieć	M	6,03	6,34
			Me	5	5
			s	5,08	4,89
			min	0	0
			max	18	26
			0,25	1,25	2,5
0,75	10	9,5			

cd. tab. 4

1	2	3	4		
			M	s	max
Wymarzona praca k = 37 w badaniu A k = 30 w badaniu B	Czy chciałbyś wykonywać pracę, która wymaga: <ul style="list-style-type: none"> realizacji szczegółowych procedur 	<ul style="list-style-type: none"> w żadnym wypadku bardzo niechętnie raczej niechętnie ani chętnie, ani niechętnie raczej chętnie bardzo chętnie z największą przyjemnością 	M	9,53	4,1
			Me	9	2
			s	5,36	5,6
			min	0	0
			max	25	30
			0,25	5	0
0,75	12,75	5,5			
				Bad. C₁	Bad. C₂
Emocje nastolatków k = 24	Określ, jak często w ciągu ostatniego roku przeżywałeś: <ul style="list-style-type: none"> poczucie bezsilności 	<ul style="list-style-type: none"> bardzo często lub zawsze często czasami rzadko bardzo rzadko lub nigdy ? – nie myślę w tych kategoriach 	M	6,37	
			Me	6	
			s	3,81	
			min	0	
			max	19	
			0,25	4	
0,75	9				
Postawy nastolatków k = 28 C ₁ : N = 185 C ₂ : powtórzony pomiar po manipulacji eksperymentalnej po dwóch tygodniach (N=99)	Często robię to, czego oczekują ode mnie rodzice, a nie to, czego sam chcę	<ul style="list-style-type: none"> całkowicie nieprawdziwe lub całkowicie się nie zgadzam nieprawdziwe lub nie zgadzam się prawdziwe lub zgadzam się całkowicie prawdziwe lub całkowicie się zgadzam ? – trudno powiedzieć (nie mogę określić) 	M	9,76	8,38
			Me	9	8
			s	5,0	4,92
			min	0	0
			max	25	22
			0,25	6	5
0,75	12	12			

Legenda: kolumna 1. – nazwa kwestionariusza wraz z liczbą pytań (k) w danym badaniu; kolumna 2. – przykładowe pytanie pochodzące z danego kwestionariusza; kolumna 3. – skala odpowiedzi; kolumna 4. – charakterystyki dotyczące liczby odpowiedzi skrajnych w danym kwestionariuszu (średnia, mediana, odchylenie standardowe, wartości minimalne i maksymalne oraz pierwszy i trzeci kwartył) dla poszczególnych grup badawczych: A, B oraz C₁ i C₂ (powtórzony pomiar).

Tab. 4. Wyniki przeprowadzonych porównań liczby odpowiedzi skrajnych dla trzech badanych prób. Źródło: opracowanie własne.

W tabeli 4 podano nazwę kwestionariusza wraz z liczbą pytań w danym kwestionariuszu (kolumna 1.), przykładowe pytanie (kolumna 2.) oraz opis skali odpowiedzi (kolumna 3.). Kolumna 4. prezentuje charakterystyki dotyczące rozkładu liczby odpowiedzi skrajnych wybieranych przez respondenta (średnia, mediana, odchylenie standardowe, wartości minimalne i maksymalne oraz pierwszy i trzeci kwartyl).

Jak widać w tabeli 4, w każdej z wykorzystywanych technik były osoby, które **ani razu** nie wybrały skrajnej wartości na skali odpowiedzi, ale także takie, które zaznaczały wartości skrajne bardzo często – w badaniu B przy trzech kwestionariuszach (Style działania, Typy osobowości, Wymarzona praca) znalazły się osoby, które wykorzystały maksymalną możliwą liczbę odpowiedzi skrajnych (odpowiednio: 50, 37 i 30).

W kolejnym kroku policzono współczynniki korelacji dla liczby zaznaczonych odpowiedzi skrajnych w różnych kwestionariuszach przez tych samych badanych. Jak zaprezentowano w tabelach 5 i 6, wszystkie korelacje są dodatnie – jedynie trzy z 15 współczynników korelacji nie osiągnęły progu istotności statystycznej ($r = 0,186$, $p = 0,066$; $r = 0,148$, $p = 0,09$; $r = 0,115$, $p = 0,19$).

	Wykonywana praca	Wymarzona praca	Style aktywności
Typy osobowości	A $r = 0,504^{***}$	A $r = 0,307^{**}$	A $r = 0,369^{***}$
	B $r = 0,148$	B $r = 0,299^{**}$	B $r = 0,207^*$
Wykonywana praca	A	A $r = 0,476^{***}$	A $r = 0,310^{**}$
	B	B $r = 0,633^{***}$	B $r = 0,115$
Wymarzona praca	A	A	A $r = 0,227^*$
	B	B	B $r = 0,219^*$

* $p = 0,05$; ** $p = 0,01$; *** $p = 0,001$.

Tab. 5. Współczynniki korelacji pomiędzy liczbą odpowiedzi skrajnych w różnych kwestionariuszach w badaniach pracowników: A ($N = 108$) i B ($N = 129$). Źródło: opracowanie własne.

W badaniu C, mimo iż pod wpływem manipulacji eksperymentalnej (gimnazjaliści oglądali film wpływający na postawy) zmniejszyła się istotnie ($t = 3,23$, $df = 97$, $p = 0,002$) liczba odpowiedzi skrajnych przy powtórzonym pomiarze postaw ($M_1 = 9,83$, $s_1 = 5,06$ vs $M_2 = 8,37$, $s_2 = 4,94$), liczba wybranych odpowiedzi skrajnych w obu pomiarach była bardzo wysoko skorelowana ($r = 0,61$, zob. tabela 6).

Istotne dodatnie współczynniki korelacji między liczbą odpowiedzi skrajnych wybieranych przez respondenta w różnych wypełnianych przez niego kwestionariuszach z odmiennie opisaną skalą odpowiedzi można pokazać

	Postawy nastolatków (pomiar 1 – badanie C1, N = 186)	Postawy nastolatków (pomiar 2 – badanie C2, N = 99)
Emocje nastolatków	$r = 0,366^{***}$	$r = 0,186+$
Postawy nastolatków (pomiar 1)		$r = 0,608^{***}$

*** $p = 0,001$; + tendencja statystyczna.

Tab. 6. Współczynniki korelacji pomiędzy liczbą odpowiedzi skrajnych w różnych kwestionariuszach w badaniach C1 i C2. Źródło: opracowanie własne.

także w innych nie przedstawianych tutaj badaniach. Oznacza to, że preferencję do wybierania skrajnych wartości na skali odpowiedzi można traktować jako *cechę respondenta*.

5.3. Przykład 3. Preferencje badanych w odniesieniu do typu skali odpowiedzi

Przykład 1. zilustrował dość oczywisty wniosek – badani muszą (i to robią) dostosować swoje opinie do dostarczonej przez badacza skali odpowiedzi. Co by się stało, gdyby mogli wybierać?

Sytuacja dość abstrakcyjna (niezdarzająca się w badaniach), ale takie pytanie mogliśmy zadać studentom rozumiejącym problem, bo uczestniczącym w kursie metodologicznym. Ich zadaniem było porangowanie 6 skal (zob. tabela 7) od najbardziej (ranga = 1) do najmniej preferowanej (ranga = 6).

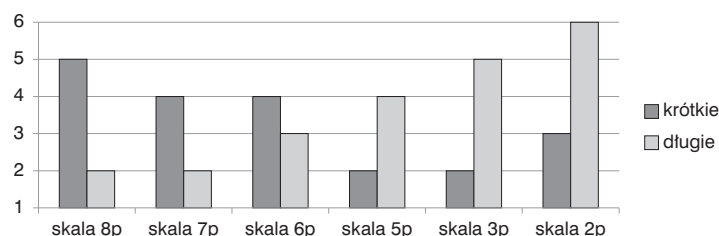
8p [8-punktowa]	7p [7-punktowa]	6p [6-punktowa]	5p [5-punktowa]	3p [3-punktowa]	2p [2-punktowa]
zdecydowanie NIE					
NIE					
raczej NIE					
ani TAK, ani NIE					
raczej TAK					
TAK					
zdecydowanie TAK					
? trudno powiedzieć					

Tab. 7. Opcje dostępne na skalach odpowiedzi (szare pola oznaczają opcje niedostępne). Źródło: opracowanie własne.

Skalę 8-punktową, z odpowiedzią środkową i dostępną odpowiedzią TP, jako najbardziej preferowaną wybrało 29 osób (34,9%), jako najmniej preferowaną wybrały ją 22 osoby (26,5%). Skalę 2-punktową (TAK, NIE) jako najbardziej preferowaną wybrało 13 osób (15,7%), za najmniej preferowaną

uznało ją 40 osób (48,2%). Widać, że nawet w tak homogenicznej próbie, jaką są studenci UW uczestniczący w kursie internetowym, występują różne preferencje.

Za pomocą analizy skupień porządków rangowych 6 typów skal odpowiedzi, metodą k-średnich wyodrębniono – z 59 osób, które wykonały to zadanie – dwie grupy różniące się preferencjami. Centroid dla grupy 1. (określony na rysunku 1 jako „krótkie”, $N = 26$) pokazuje, że najbardziej preferowane są skale 3- i 5-punktowe (ranga = 2), potem 2-punktowa, najmniej preferowana jest skala 8-punktowa (ranga = 5). Centroid dla grupy 2. (określony na rysunku 1 jako „długie”, $N = 33$) wskazuje na preferowanie skal długich (8- i 7-punktowych, ranga = 2) i awersję do skal krótkich (ranga = 6 dla skali 2-punktowej).



Rys. 1. Wyniki analizy skupień porządków rangowych 6 typów skal. Źródło: opracowanie własne.

Ci sami badani parę tygodni później wypełniali skalę samooceny Rosenberga (Laguna, Lachowicz-Tabaczek i Dzwonkowska, 2007) w wersji pojedynczych itemów i w wersji wyboru między dwoma opisami (zob. tabela 8).

Dla każdej osoby została policzona liczba odpowiedzi skrajnych dla obu technik pomiaru samooceny. W przypadku samooceny skrajne odpowiedzi są wybierane o wiele częściej w przypadku „zdań pozytywnych” („Mam do siebie pozytywny stosunek”) niż negatywnych („Czuję, że mam mało powodów do dumy z siebie”), więc jest to przypadek techniki, w której liczba odpowiedzi skrajnych koreluje ze średnią, która z kolei koreluje z płcią respondenta. W wielu badaniach stwierdza się bowiem podwyższoną w stosunku do mężczyzn depresyjność kobiet (np. Kessler, 2003). Dlatego hipotezę, mówiącą, że osoby preferujące skale dłuższe wybierają mniej skrajnych odpowiedzi niż osoby preferujące skale krótsze testowaliśmy, kontrolując płeć.

W przypadku obu technik efekt preferencji okazał się istotny ($F(1,55) = 6,0; p < 0,05; \eta^2 = 0,098$ dla techniki 10-itemowej; $F(1,55) = 7,2; p < 0,05; \eta^2 = 0,116$ dla techniki 5-itemowej). Preferujący krótsze skale wybrali istotnie mniej odpowiedzi skrajnych na skali 5- i 10-itemowej ($M_1 = 2,54, M_2 = 3,11$) niż preferujący skale dłuższe ($M_1 = 1,51; M_2 = 2,12$).

10 itemów	5 itemów
Skala odpowiedzi: zdecydowanie nie zgadzam się; nie zgadzam się; zgadzam się; zdecydowanie zgadzam się; ? trudno powiedzieć, nie mam zdania	Skala odpowiedzi: [A] Tak jak osoba A; [rA] Raczej jak osoba A; [rB] Raczej jak osoba B; [B] Tak jak osoba B; [?] trudno powiedzieć
1. W sumie skłonny jestem sądzić, że marnie mi w życiu idzie	1. Osoba A skłonna jest sądzić, że marnie jej w życiu idzie. Osoba B uważa, że potrafi radzić sobie równie dobrze, jak większość innych ludzi
2. Potrafię radzić sobie równie dobrze, jak większość innych ludzi	
3. Mam do siebie pozytywny stosunek	2. Osoba A ma do siebie pozytywny stosunek. Osoba B czuje, że ma mało powodów do dumy z siebie
4. Czuję, że mam mało powodów do dumy z siebie	
5. Czuję, że jestem osobą wartościową, przynajmniej na równi z innymi	3. Osoba A uważa, że jest osobą wartościową, przynajmniej na równi z innymi. Osoba B czasami czuję się kompletnie bezużyteczna
6. Czasami czuję się kompletnie bezużyteczny	
7. Niekiedy myślę, że jestem do niczego	4. Osoba A niekiedy myśli, że jest do niczego. Osoba B uważa, że ma wiele dobrych cech
8. Mam wiele dobrych cech	
9. Chciałbym mieć więcej szacunku dla samego siebie	5. Osoba A chciałaby mieć więcej szacunku dla samego siebie. Osoba B, ogólnie rzecz biorąc, jest z siebie zadowolona
10. Ogólnie rzecz biorąc, jestem z siebie zadowolony	

Tab. 8. Pytania ze skali samooceny 10- i 5-itemowej. Źródło: opracowanie własne.

Wniosek – istnienie różnic indywidualnych w preferencjach co do długości skali (subtelności różnicowania) powoduje, że trudno opowiedzieć na pytanie, jaka długość skali odpowiedzi jest *optymalna*, ponieważ zawsze dłuższe skale mogą być „skracane” przez transformację dokonywaną się w umysłach badanych. Skoro długie skale mogą być „skracane”, a krótkie nie mogą zostać wydłużane, następne badanie miało za zadanie porównanie wyników pomiaru uzyskanych za pomocą skal 3- i 5- punktowych.

5.4. Przykład 4. Czy lepsze są skale pięciopunktowe czy trzypunktowe?

Jak cytowaliśmy we wprowadzeniu, w literaturze pojawia się rekomendacja, że jeżeli interesujemy się średnimi dla grup respondentów, skale dwu-, trzypunktowe są wystarczające. Jeżeli interesują nas natomiast wyniki pojedynczych osób, skale powinny być pięcio-, siedmiopunktowe. W celu sprawdzenia skutków zastosowania różnych typów skali odpowiedzi, przeprowadzono analizy dotyczące konstrukcji wskaźnika jednego z wymiarów Stylu Aktywności – preferencji dla *symultanicznego* vs *sekwencyjnego* wykonywania zadań (Wieczorkowska, 1998). Do pomiaru wykorzystuje się odpowiedzi na

cztery pytania (zob. tabela 9). Osoba preferująca wykonywanie zadań po kolei jest opisana w dwóch pytaniach jako A, w dwóch jako B.

Porównywano wyniki uzyskane w badaniach 108 studentów psychologii (SW), którzy w ciągu miesiąca odpowiadali na pytania dwukrotnie, za pomocą dwóch różnych skal odpowiedzi:

- raz, korzystając ze skali pięciopunktowej (<tak jak A>, <raczej tak jak A>, <raczej tak jak B>, <tak jak B>, <trudno powiedzieć>) – grupa SW₅,
- drugi raz, korzystając ze skali trzypunktowej (<tak jak A>, <tak jak B>, <trudno powiedzieć>) – grupa SW₃.

W tabeli 9 przedstawiono procenty osób wybierających poszczególne odpowiedzi na dane pytanie, po przekodowaniu odpowiedzi <trudno powie-

		SW ₅	SW _{5→3}	SW ₃	
1. Osoba A denerwuje się, gdy musi jednocześnie myśleć o kilku różnych sprawach. Osoba B stara się mieć kilka rzeczy rozpoczętych równocześnie, aby „prze-rzucać się” z jednej na drugą	A	25,0	45,4	48,1	A
	rA	20,4			
	TP	5,6	5,6	4,6	TP
	rB	32,4	49,1	47,2	B
	B	16,7			
2. Osoba A stara się realizować wiele celów równocześnie („piec wiele pieczeni przy jednym ogniu”). Osoba B woli je realizować po kolei	A	18,5	56,5	54,6	A
	rA	38,0			
	TP	0,9	0,9	3,7	TP
	rB	23,1	42,5	41,7	B
	B	19,4			
3. Osoba A nie lubi mieć rozpoczętych kilku zadań. Osoba B często przerywa pracę, zajmując się w prze-rwach innym zadaniem	A	12,0	34,2	39,8	A
	rA	22,2			
	TP	3,7	3,7	6,5	TP
	rB	38,9	62,0	53,7	B
	B	23,1			
4. Gdy różne zadania konkurują ze sobą co do ważno-ści, osoba A stara się w jakiś sposób realizować je równocześnie. Osoba B lubi koncentrować się tylko na jednym zadaniu jednocześnie	A	18,5	59,2	56,5	A
	rA	40,7			
	TP	1,9	1,9	5,6	TP
	rB	25,0	38,9	38,0	B
	B	13,9			

Legenda:

A – <tak jak A>, rA – <raczej A>, rB – <raczej B>, B – <tak jak B>, TP – <Trudno powiedzieć>.

SW_{5→3} – połączone odpowiedzi A i rA oraz B i rB,

SW₅ – pierwszy pomiar z wykorzystaniem skali 5-punktowej,

SW₃ – drugi pomiar z wykorzystaniem skali 3-punktowej.

Tab. 9. Procenty osób wybierających poszczególne punkty na skali pięcio- oraz trzypunktowej. Źródło: opracowanie własne.

dzieć> na środek skali. W kolumnie oznaczonej jako (i) SW₅ pokazano wyniki dla 5-punktowej skali; (ii) SW₃ pokazano wyniki dla 3-punktowej skali; (iii) SW_{5→3} wyniki po połączeniu procentów odpowiedzi skrajnych z odpowiedziami „raczej” (<A + raczej A> oraz <B + raczej B>).

Porównanie rozkładów odpowiedzi ze skali 5- i 3-punktowej pokazuje, że ograniczenie liczby odpowiedzi do trzech spowodowało wzrost liczby odpowiedzi <trudno powiedzieć> w trzech ostatnich pytaniach.

6. Budowanie wskaźnika symultaniczności

Tak jak w innych badaniach Sondażu Stylów Aktywności (np. Wierzińska-Wierzińska, 2014), odpowiedzi na przytoczone cztery pytania tworzą jeden czynnik. Największy procent wariancji wyjaśnianej przez czynnik oraz najwyższą alfę Cronbacha (zob. tabela 10) uzyskano w przypadku, gdy badani posługiwali się 3-punktową skalą odpowiedzi. W tabeli 10 przedstawione są również inne charakterystyki poszczególnych skal: średnia, mediana, odchylenie standardowe, skośność oraz kurtoza.

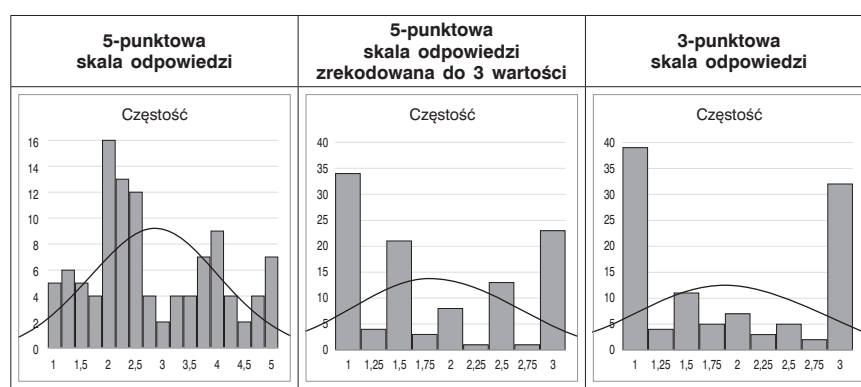
	% wariancji	alfa Cronbacha	M	s	Me	skośność	kurtoza
SW ₅ (5 punktów)	66,29	0,83	2,82	1,16	2,5	0,390	-0,960
SW _{5→3}	65,07	0,82	1,84	0,78	1,5	0,387	-1,435
SW ₃ (3 punkty)	77,00	0,90	1,89	0,85	1,63	0,270	-1,660

Tab. 10. Charakterystyki opisujące poszczególne typy skal. Źródło: opracowanie własne.

Celem zadawania pytań jest zbudowanie wskaźnika zmiennej (w tym przypadku symultaniczności), należy więc zastanowić się, jak liczba punktów na skali odpowiedzi wpływa na kształt jego rozkładu. Jak widać na wykresach (rysunek 2), w rozkładzie wskaźnika sekwencyjności zbudowanym z odpowiedzi na skali 5-punktowej najczęściej występują wartości środkowe (wynik 2 lub mniej niż 2,75 otrzymuje 37,9% osób), a nie skrajne. Gdy wskaźnik jest zbudowany z odpowiedzi na skali 3-punktowej, najczęściej występujące wartości to 1 (36,1% osób) i 3 (29,6% osób). Rozkład sekwencyjności, który wtedy otrzymujemy, daleki jest od rozkładu normalnego, a zbliżony do bimodalnego. Kiedy zrekodujemy wartości skali odpowiedzi z 5 do 3, podobnie jak w przypadku skali 3-punktowej, najwięcej osób badanych znajduje się na krańcach rozkładu (31,5 % i 21,3%), ale także częste są wyniki 1,5 (19,4% osób) i 2,5 (12% osób). Choć przedstawiliśmy analizy, ograniczając się jedynie do jednego wymiaru stylów działania, to analogiczne zależności (*dwumodalność rozkładu wskaźników zbudowanych z wartości na 3-punktowej skali odpowiedzi vs „normalność” rozkładu wskaźników zbudowanych*

z wartości na 5-punktowej skali odpowiedzi) stwierdzono dla pozostałych 8 wymiarów.

W naukach społecznych od dawna trwa dyskusja nad charakterem różnych cech. Czy pracownicy dzielą się na ekstrawertyków i introwertyków, działających sekwencyjnie vs symultaniczne itp.? Jeśli tak, to rozkład ekstrawersji czy sekwencyjności powinien być zbliżony do bimodalnego. A może rozkład ekstrawersji czy sekwencyjności jest zbliżony do normalnego, gdzie większość populacji lokuje się w środku wymiaru? Przedstawiony wyżej przykład pokazuje, że odpowiedź na to pytanie zależy od pomiaru, a konkretnie od użytej w badaniach skali odpowiedzi. Im większa liczba punktów na skali odpowiedzi, tym większa szansa, że otrzymany rozkład wskaźnika zbudowanego z wielu pytań będzie zbliżał się do rozkładu normalnego.



Rys. 2. Rozkłady wskaźnika sekwencyjności zbudowanego z odpowiedzi na skalach o różnej liczbie punktów. Źródło: opracowanie własne.

Słabością opisanego porównania jest to, że w badaniu nie rotowano kolejności typów skal odpowiedzi. Wszyscy uczestnicy najpierw odpowiadali za pomocą skali 5-punktowej, a po miesiącu 3-punktowej. W następnych badaniach należałoby rotować kolejność, ponieważ nie można wykluczyć, że zmieniłaby się liczba odpowiedzi <raczej A>, <raczej B>, ale nie ma to wpływu na konkluzję, że pozwalając na większe różnicowanie odpowiedzi na poszczególne pytania wpływamy na kształt rozkładu wskaźnika z nich zbudowanego, który ma znaczenie dla dalszych analiz. Liczenie współczynników korelacji Pearsona wymaga założenia, że rozkłady zmiennych są normalne, więc jeśli naszym celem jest testowanie zależności między zmiennymi (operacjonalizowanymi przez zbudowane wskaźniki) lepszą skalą odpowiedzi jest skala 5-punktowa. Jeśli natomiast chcemy wyróżnić grupy skrajne, w przypadku analizowanych pytań łatwiej jest to zrobić przy wskaźnikach zbudowanych z odpowiedzi na skali 3-punktowej.

7. Podsumowanie

Decyzje dotyczące wyboru skali odpowiedzi często podejmowane przez badaczy bezrefleksyjnie mają istotne konsekwencje, ponieważ respondent musi transponować swoją „prywatną” skalę do różnicowania w dostarczonej przez badacza. Ta transpozycja wyznacza jego styl odpowiadania na pytania (*response style*), który może być uniwersalny, tzn. w niewielkim stopniu zależny od tego, jaki obiekt jest aktualnie oceniany. Pokazaliśmy, że respondenci mają zróżnicowane preferencje do długości skali odpowiedzi (niektórzy preferują skale krótkie {TAK, NIE}, inni wolą wielopunktowe). Bez względu na to, jak długa jest skala odpowiedzi, respondent może wykorzystywać w swoich odpowiedziach – jeśli taka jego wola – tylko odpowiedzi skrajne. Pokazaliśmy w trzech badaniach, że liczby odpowiedzi skrajnych użytych w różnych technikach pomiarowych korelują wysoko dodatnio między sobą, więc preferencję odpowiedzi skrajnych można uznać za cechę stylu odpowiadania respondenta.

W artykule przedstawiliśmy też argumenty przemawiające za następującymi tezami:

- Wskaźniki zbudowane z odpowiedzi na skalach szacunkowych mogą być traktowane jako zmienne ilościowe.
- Należy unikać odpowiedzi środkowej na skali odpowiedzi, pozwalając jednak na udzielenie odpowiedzi beztreściowej.
- Odpowiedzi beztreściowe (nie wiem, nie mam zdania) powinny się pojawiać na końcu skali, a nie w środku.
- Kształt rozkładu wskaźnika zbudowanego z odpowiedzi na wiele pytań zależy od liczby punktów na skali odpowiedzi.

Nie mamy wątpliwości, że konsekwencje wyboru typu skali odpowiedzi powinny być przedmiotem dalszych badań.

Przypisy

- ¹ Można byłoby oczekiwać, że jest to istotnie związane z wiekiem (dotyczy młodszych badanych), ale okazało się (Wierziński, 2009), że grupa udzielająca odpowiedzi beztreściowej na te pytania była istotnie starsza od reszty. W pytaniu o „Solidarność” średni wiek grupy uchylającej się od odpowiedzi wyniósł $M = 49,12$ ($s = 19,72$, $N = 624$) i był istotnie wyższy od wieku pozostałych, dla których średnia wynosiła $M = 45,83$ ($s = 16,14$, $N = 5362$). Grupa uchylająca się od odpowiedzi charakteryzowała się także istotnie większą wariancją wieku. W pytaniu o OPZZ średni wiek grupy uchylającej się od odpowiedzi wyniósł $M = 48,1$ ($s = 18,98$, $N = 998$) i był istotnie wyższy od wieku pozostałych, dla których średnia wynosiła $M = 45,79$ ($s = 16,0$, $N = 4984$). Grupa uchylająca się od odpowiedzi charakteryzowała się także istotnie większą wariancją wieku.

Bibliografia

- Baumgartner, H. i Steenkamp, J.M. (2001). Response Styles in Marketing Research: A Cross-National Investigation. *Journal Of Marketing Research*, 38 (2), 143–156.
- Benson, P. (1971). How Many Scales and How Many Categories Shall We Use in Consumer Research? A Comment. *The Journal of Marketing*, 35 (4), 59–61.
- Brzezińska, A. i Brzeziński, J. (2004). Skale szacunkowe w badaniach diagnostycznych. W: J. Brzeziński (red.), *Metodologia badań psychologicznych. Wybór tekstów*. Warszawa: Wydawnictwo Naukowe PWN.
- Cichomski, B., Jerzyński, T. i Zieliński, M. (2010). Polskie Generalne Sondáže Społeczne: struktura skumulowanych wyników badań 1992–2010. Warszawa: Instytut Studiów Społecznych, Uniwersytet Warszawski. [Baza danych i dokumentacja metodologiczna dostępne także na stronach: <http://www.iss.uw.edu.pl>; <http://pgss.iss.uw.edu.pl> oraz na stronie Archiwum Danych Społecznych (ADS): <http://ads.org.pl>].
- Chun, K.-T., Campbell, J.B. i Yoo, J.H. (1974). Extreme Response Style in Crosscultural Research. *Journal of Cross-Cultural Psychology*, 5 (4), 465–480.
- Clarke III, I. (2000). Extreme Response Style in Cross-Cultural Research: An Empirical Investigation. *Journal of Social Behavior & Personality*, 15 (1), 137–152.
- Green, P.E. i Rao, V.R. (1970). Rating Scales and Information Recovery. How Many Scales and Response Categories to Use. *Journal of Marketing*, 34 (July), 33–39.
- Harzing, A. (2006). Response Styles in Cross-national Survey Research. *International Journal of Cross Cultural Management*, 6 (2), 243–266.
- Kaden, R.J. (2008). *Badania marketingowe*. Warszawa: Polskie Wydawnictwo Ekonomiczne.
- Kessler, R.C. (2003). Epidemiology of Women and Depression. *Journal of Affective Disorders*, 74 (1), 5–13.
- Krosnick, J.A., Holbrook, A.L., Berent, M.K., Carson, R.T., Hanemann, W.M., Kopp, R.J., Mitchell, R.C., Presser, S., Ruud, P.A., Smith, V.K., Moody, W.R., Green, M.C. i Conaway, M. (2002). The Impact of “No Opinion” Response Options on Data Quality: Non-Attitude Reduction or an Invitation to Satisfice? *The Public Opinion Quarterly*, 66 (3), 371–403.
- Laguna, M., Lachowicz-Tabaczek, K. i Dzwonkowska, I. (2007). Skala samooceny SES Morrisa Rosenberga – polska adaptacja metody. *Psychologia Społeczna*, 2 (4), 164–176.
- Lee, C. i Green, R.T. (1991). Cross-Cultural Examination of the Fishbein Behavioral Intentions Model. *Journal of International Business Studies*, 22 (2), 289–305.
- Lehmann, R. i Hulbert, J. (1972). Are Three-point Scales Always Good Enough? *Journal of Marketing Research*, 9 (November), 444–446.
- Mieczkowska, E. (2013). *Wpływ perspektywy przyjmowanej podczas oglądania filmu na przeżywane emocje i postawy rodzinne nastolatków*. Niepublikowana praca magisterska, Szkoła Wyższa Psychologii Społecznej, Warszawa.
- Peterson, B.L. (1985). Confidence: Categories and Confusion. GSS Methodological Reports #31.ISR.
- Shank, D.R. (1985). Continuous Monitoring of Human Contingency Judgment across Trials. *Memory & Cognition*, 13, 158–176.
- Shiomi, K. i Loo, R. (1999). Cross-cultural Response Styles on the Kirton Adaptation-Innovation Inventory. *Social Behavior and Personality*, 27 (4), 413–420.
- Strelau, J. (2004). Osobowość jako zespół cech. W: J. Strelau (red.), *Psychologia. Podręcznik akademicki. Psychologia ogólna* (t. 2, s. 525–560). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Sułek, A. (2001). *Sondaż polski*. Warszawa: IFIS PAN.
- Sułek, A. (2002). *Ogród metodologii socjologicznej*. Warszawa: Wydawnictwo Naukowe Scholar.
- Takahashi, K., Ohara, N., Antonucci, T.C. i Akiyama, H. (2002). Commonalities and Differences in Close Relationships among the Americans and Japanese: A Compari-

- son by the Individualism/Collectivism Concept. *International Journal of Behavioral Development*, 26 (5), 453–465.
- Van Herk, H., Poortinga, Y.H. i Verhallen, T.M. (2004). Response Styles in Rating Scales: Evidence of Method Bias in Data From Six EU Countries. *Journal of Cross-Cultural Psychology*, 35 (3), 346–360.
- Wasserman, E.A. i Shaklee, H. (1984). Judging Response-Outcome Relation: The Role of Response-Outcome Contingency, Outcome Probability, and Method of Information Presentation. *Memory & Cognition*, 3, 270–286.
- Wieczorkowska-Siarkiewicz, G. (1992). *Punktowe i przedziałowe reprezentacje celu*. Warszawa: OWWP.
- Wieczorkowska-Nejtardt, G. (1998). *Inteligencja Motywacyjna. Mądre strategie wyboru celu i sposobu działania*. Warszawa: Wydawnictwa Instytutu Studiów Społecznych.
- Wieczorkowska-Wierzińska, G. (2014). Diagnoza psychologiczna predyspozycji pracowników. *Problemy Zarządzania*, 12 (1).
- Wieczorkowska, G. i Wierziński, J. (2011). *Statystyka: od teorii do praktyki*. Warszawa: Wydawnictwo Naukowe Scholar.
- Wieczorkowska-Wierzińska, G., Wierziński, J. i Kuźmińska, A. (2014). Porównywalność danych sondażowych zebranych w różnych krajach. *Psychologia Społeczna*, (w druku).
- Wieczorkowska, G., Wierziński, J. i Siarkiewicz, M. (2009). Wybrane problemy metodologiczne analitycznych badań sondażowych. W: M. Zahorska i E. Nasalska (red.). *Wartości – polityka – społeczeństwo*. Warszawa: Wydawnictwo Naukowe Scholar.
- Wierziński J. (2007). Wykorzystanie własności funkcji wartości subiektywnej do przewidywania dynamiki zachowań celowych. W: K. Winkowska-Nowak, A. Nowak i A. Rychwalska (red), *Modelowanie matematyczne i symulacje komputerowe w naukach społecznych*. Warszawa: Wydawnictwo Academica Szkoły Wyższej Psychologii Społecznej.
- Wierziński, J. (2009). *Badanie zaufania do organizacji: problemy metodologiczne*. Warszawa: Wydawnictwo Naukowe Wydziału Zarządzania UW.