

Porównywalność danych sondażowych zebranych w różnych krajach

Grażyna Wieczorkowska-Wierzbńska^{1,2,3}, Jerzy Wierzbński², Anna Olga Kuźmińska²

¹ Instytut Studiów Społecznych, Uniwersytet Warszawski

² Wydział Zarządzania, Uniwersytet Warszawski

³ Wydział Psychologii, Szkoła Wyższa Psychologii Społecznej

Nieporównywalność zebranych danych jest podstawowym problemem w badaniach międzykulturowych. Uniwersalne problemy analiz porównawczych zostały omówione na przykładzie zrealizowanych w 2009 roku internetowych badań własnych włoskiej i japońskiej prób reprezentatywnych i danych zebranych w obu krajach w ramach *International Social Survey Programme* (1998, 2008). Oprócz omówienia problemu równoważności pomiaru i badanych prób testowano różnice w stylach odpowiadania, przejawiające się w liczbie odpowiedzi beztreściowych (*nie mam zdania, trudno powiedzieć*) i skrajnych (*zdecydowanie tak, zdecydowanie nie*). W prawie wszystkich z analizowanych sześciu zestawów pytań Japończycy udzielali mniej odpowiedzi skrajnych i uchylali się od zajęcia stanowiska istotnie częściej niż Włosi. Wyniki porównania są interpretowane jako przejaw różnic w stylach myślenia (Nisbett, 2009), polegających na dominacji akceptacji sprzeczności (uznania, że dobre i złe, mocne i słabe itd. istnieją we wszystkim) w kulturze Wschodu i reguły wyłączonego środka (zdanie jest albo prawdziwe, albo fałszywe) w kulturze Zachodu. Tekst kończy się rekomendacją dotyczącą wykrywania i usuwania z dalszych analiz tzw. fałszywych respondentów, którzy udzielają odpowiedzi losowo. Fałszywi respondenci są większym zagrożeniem dla sondaży analitycznych (nastawionych na testowanie zależności między zmiennymi) niż brak reprezentatywności próby (która jest niezbędna, gdy naszym celem jest, tak jak w sondażach sprawozdawczych, określenie rozkładu zmiennej w populacji).

Słowa kluczowe: *równoważność pomiaru, styl odpowiadania, odpowiedzi beztreściowe, odpowiedzi skrajne, międzykulturowe badania sondażowe, ISSP, fałszywi respondenci*

Globalizacja stawia przed nami nowe wyzwania. Nie tylko powiązania biznesowe, lecz także media i kultura symboliczna dają nam możliwość/konieczność pozostawania w kontakcie z całym światem. Formowanie międzynarodowych zespołów nie jest wyjątkiem, ale regułą¹. Owocna współpraca międzynarodowa wymaga jednak

wiedzy o podobieństwach i różnicach. Jednym ze źródeł tej wiedzy są wyniki porównawczych badań sondażowych prowadzonych równoległe w różnych krajach. Celem takich porównań jest wykrywanie tego, co uniwersalne dla wszystkich społeczeństw (taką zależnością jest na przykład dodatnia korelacja między wykształceniem a wiekiem respondenta), i tego, co specyficzne dla danego kraju (ilustracją jest istotna tylko w 11 na 31 analizowanych krajów interakcja płci i wieku w wyjaśnianiu poziomu wykształcenia, która oznacza szybszy awans edukacyjny kobiet w tych społeczeństwach – por. Wierzbński, 2009). Jednak nawet w bardzo obszernym i świetnie

Grażyna Wieczorkowska-Wierzbńska, Wydział Zarządzania, Uniwersytet Warszawski, ul. Szturmowa 1/3, 02-678 Warszawa, e-mail: gw@uw.edu.pl

Jerzy Wierzbński, Wydział Zarządzania, Uniwersytet Warszawski, ul. Szturmowa 1/3, 02-678 Warszawa, e-mail: wierzbinski@mail.wz.uw.edu.pl

Anna Kuźmińska, Wydział Zarządzania, Uniwersytet Warszawski, ul. Szturmowa 1/3, 02-678 Warszawa, e-mail: akuzminska@mail.wz.uw.edu.pl

¹ Przykładem może tu być załoga statku Norwegian Majesty kursującego między Filadelfią a Bermudami pod banderą norweską,

na którym pływa skrajnie zróżnicowana narodowościowo załoga. Więcej nawet – armator nie chce zatrudniać ludzi z tego samego kraju (wyjątek stanowią stewardzi z Filipin), ponieważ nie chce, by załoga dzieliła się na jednonarodowe kliki.

napisanym podręczniku Pawła Boskiego *Kulturowe ramy zachowań społecznych* (2009) nie ma rozdziału poświęconego problemom metodologicznym, a przecież nierównowalność zebranych danych jest podstawową trudnością w badaniach międzykulturowych².

Informacja o porównywanych krajach i analizowanych danych

O uniwersalnych problemach analizy danych międzykulturowych opowiemy na przykładzie zrealizowanych w 2009 roku porównawczych badań włosko-japońskich poświęconych reakcjom na kryzys finansowy. Analizy zostaną następnie rozszerzone na porównanie obu krajów na podstawie danych zebranych w ISSP (International Social Survey Programme, 1998, 2008). ISSP to stały międzynarodowy program cyklicznych badań porównawczych. Jego celem jest przeprowadzanie raz na dwa lata przez wszystkie kraje członkowskie badań na określony temat, według tej samej metodologii, z zastosowaniem identycznego zestawu pytań. Dystrybucją danych ISSP dla ośrodków naukowych w Polsce zajmuje się, wraz z Archiwum Danych Społecznych (<http://www.ads.org.pl>), Instytut Studiów Społecznych Uniwersytetu Warszawskiego. Wybór dwóch krajów był podyktowany zainteresowaniem Włochami jednego z członków zespołu badawczego. Jako kraj porównawczy została wybrana Japonia – z powodu różnic pod względem indywidualizmu i unikania niepewności (wyższe nasilenie obu kulturowych wymiarów w Japonii według klasyfikacji Hofstede) oraz podobieństw pod względem zamożności (zbliżony PKB), wzorca męskości i dystansu wobec władzy (Johnson, Kulesa, Cho, Shavitt, 2005).

Choć tekst ilustrowany jest wieloma wynikami analiz powyższych danych (tabela 1), nie zamieszczamy szczegółów, bo znacznie zmniejszyłoby to czytelność wywodu.

RÓWNOWAŻNOŚĆ TŁUMACZENIA NARZĘDZIA

Polsko-amerykański zespół badawczy (Eugene Burnstein, Richard Gonzales, Grażyna i Jerzy Wierzbiniński), uzbrojony we wskazówki dotyczące tworzenia łatwego do tłumaczenia narzędzia, przygotował angielską wersję pytań. Zastosowanie ich we Włoszech i w Japonii wymagało ich przetłumaczenia. W takich wypadkach najczęściej stosowana jest metoda „tam i z powrotem” (*back-translation*), polegająca na tłumaczeniu na język docelowy, a następnie przez innego tłumacza ponownie na język oryginału (Su, Parham, 2002). Efektem zastosowania tej metody mogą być sformułowania stanowiące zbyt dokładną reprodukcję tekstu źródłowego i przez to brzmiące nienaturalnie. Ten sposób tłumaczenia może okazać się pomocny w ocenie trafności przekładu. Jest on jednak niezwykle rzadko wykorzystywany przez profesjonalnych tłumaczy, którzy opierają się najczęściej na ocenie zespołu kompetentnych znawców obu języków. Inną metodą jest (Werner, Campbell, 1970; za: van de Vijver, Leung, 1996) jednoczesne dopasowywanie do siebie obu wersji – oryginalnej i przetłumaczonej. W tym przypadku celem nie jest więc dosłowna reprodukcja pierwowzoru, ale osiągnięcie naturalności i zrozumiałości.

Brislin, Lonner i Campbell (1973; za: van de Vijver, Leung, 1996) sformułowali zestaw heurystyk pomocnych przy tworzeniu łatwego do tłumaczenia narzędzia:

1. Stosuj krótkie, proste zdania, co minimalizuje obciążenie poznawcze.
2. Używaj raczej strony czynnej niż biernej.
3. Powtarzaj rzeczowniki zamiast wykorzystywać zaimeki, które mogą być w niektórych językach trudne do przetłumaczenia.
4. Unikaj metafor i zwrotów potocznych, które są zazwyczaj kłopotliwe w przekładzie.

Tabela 1
Informacja o analizowanych danych³

Próba	N	% kobiet	Wiek (M, s)	Wykształcenie (M, s)
2009 Włochy	797	51,1	45,9 (14,13)	13,02 (2,85)
2009 Japonia	833	48,3	44,95 (13,27)	14,47 (2,04)
1998 + 2008 ISSP Włochy	2031	50,6	48,43 (16,54)	10,45 (4,43)
1998 + 2008 ISSP Japonia	2495	51,9	47,63 (16,95)	12,24 (2,63)

² Ze względu na temat numeru *Psychologii Społecznej* używamy terminu *porównanie międzykulturowe*, choć, co będziemy wykazywać w tekście, omawiane zagadnienia mają charakter bardziej uniwersalny.

³ Autorzy dziękują Giulii Ranuzii de Bianchi za zainicjowanie i zebranie funduszy na przeprowadzenie internetowych badań prób reprezentatywnych w 2009 roku we Włoszech i Japonii.

5. Unikaj trybu przypuszczającego (w języku angielskim sformułowania typu *could* czy *would*).

6. Nie unikaj redundancji w zdaniach zawierających kluczowe pojęcia.

7. Unikaj przysłówków i przyimków określających miejsce i czas, np. ponad, poza.

8. Unikaj formy dzierżawczej, jeśli to tylko możliwe.

9. Wykorzystuj słowa określające raczej konkretne obiekty (np. Polacy, Amerykanie) niż sformułowania bardziej ogólne (np. „ludzie różniący się narodowością”).

10. Unikaj słów wskazujących na brak precyzji, typu „prawdopodobnie” czy „często”.

11. Gdzie jest to tylko możliwe, używaj sformułowań znanych tłumaczom.

12. Unikaj zdań z dwoma czasownikami, które dotyczą różnych czynności (np. „myślę i mówię”).

W literaturze wymieniane są trzy sposoby sprawdzania trafności tłumaczenia: (1) osoby dwujęzyczne sprawdzają wersję oryginalną i docelową, (2) osoby znające jedynie język źródła sprawdzają oryginał i wersję powtórnie przetłumaczoną, (3) osoby kompetentne jedynie w języku docelowym sprawdzają tłumaczenie, natomiast znające tylko język oryginału sprawdzają powtórne tłumaczenie – jest to metoda najczęściej stosowana (van de Vijver, Leung, 1996). W naszym badaniu angielska wersja była tłumaczona na odpowiednie języki przez agencje badawcze, które realizowały badania w Japonii i we Włoszech. Przetłumaczone wersje były sprawdzane pod kątem zgodności przez dwujęzycznych psychologów: japońska przez prof. Shinobu Kitayamę, a włoska przez Giulię Ranuzii de Bianchi – czyli wybrany został wariant pierwszy. W dokumentacji sondażu ISSP możemy przeczytać, że jakość japońskiego tłumaczenia pytań sprawdzona została przez osobę anglojęzyczną od lat mieszkającą w Japonii i pozostającą w związku małżeńskim z Japończykiem (Harkness, Langfeldt, Scholz, 2000). Jakość tłumaczenia włoskiego oceniano na podstawie porównania dwóch niezależnych tłumaczeń kwestionariusza. Trafność tłumaczenia może być szacowana przez ekspertów, ale zdarza się, że konieczne jest wykazanie jej za pomocą analizy zebranych danych. Równoważność tłumaczenia możemy oceniać na trzech poziomach (Poortinga, 1989), poprzez porównanie: (1) treści pytania (*item wording*), (2) rozkładów odpowiedzi w obu kulturach, (3) korelacji odpowiedzi na pytania. Czytając pytania zadane Amerykanom w sondażu ISSP (1998, 2008), zauważymy, że zaufanie jest określane jako *trust*, gdy dotyczy ludzi, i jako *confidence*, gdy dotyczy instytucji. W polskiej wersji sondażu oba terminy przetłumaczono jako „zaufanie”, choć słowo *confidence* w innych kontekstach jest tłumaczone jako „ufność” (np. *confidence interval* jako „przedział ufności”). Czy w związku

z tą różnicą w sformułowaniu pytań możemy porównywać wyniki? Pierwszy poziom porównania równoważności pomiaru to poziom leksykalny, drugi dotyczy porównania rozkładów odpowiedzi, trzeci porównania korelacji. Wierziński (2009, 2012) porównał korelacje w obu krajach, wykonując serie analiz regresji, w których zmienną zależną było zaufanie do instytucji (Y), a na liście predyktorów znalazły się odpowiedzi na pytanie o zaufanie do ludzi (X), wiek, płeć, wykształcenie (w latach nauki), wskaźnik sytuacji finansowej, kraj oraz interakcja kraju z X. Brak istotnej interakcji X z krajem w wyjaśnianiu Y upoważnia do wniosku, że mimo użycia w badaniach amerykańskich dwóch terminów: *trust* (w odniesieniu do ludzi) i *confidence* (w odniesieniu do organizacji), przetłumaczonych w wersji polskiej jednakowo jako „zaufanie” (różnice na poziomie leksykalnym), korelacje zaufania instytucjonalnego z zaufaniem interpersonalnym nie różnią się w obu próbach. Można więc stwierdzić, że chociaż w języku angielskim terminy *confidence* i *trust* różnią się na poziomie leksykalnym, to jednak charakteryzują się równoważnością funkcjonalną, co pozwala na dokonywanie porównań międzynarodowych z krajami, w których takich różnic w określaniu zaufania nie ma.

RÓWNOWAŻNOŚĆ POMIARU (MEASUREMENT EQUIVALENCE)

Gdy porównujemy poziomy zmiennych psychologicznych, np. szczęścia w różnych kulturach, i mamy wątpliwości co do trafności odpowiedzi na pytania, to warto przypomnieć argumenty Gilberta (2007). Jego wywód można streścić następująco: jeśli nawet określimy (w przybliżeniu), co rozumiemy pod pojęciem szczęścia, nie możemy być pewni, czy dwie osoby, które zaznaczyły tę samą odpowiedź, odczuwają to samo. Szczęście – i wiele innych terminów psychologicznych – należy do tych słów, które dla zbyt wielu ludzi mają zbyt wiele różnych znaczeń. Jeśli jednak dany obiekt jest niemierzalny na żadnej skali pomiarowej, nie może być badany naukowo. Może być przedmiotem badań, ale nauka wymaga pomiarów, więc jeśli czegoś nie można zmierzyć – określić jego właściwości za pomocą zegara, linijki lub jakiegokolwiek narzędzia – nie może to być przedmiotem badań naukowych. Bardzo trudno jest zmierzyć szczęście i jednocześnie być przekonanym o wiarygodności i rzetelności dokonanego pomiaru. Badacz nigdy nie będzie pewny, na ile wiernie te doznania są odwzorowane w postaci liczb, ponieważ respondenci mogą się mylić co do swoich uczuć, mogą nie pamiętać, co odczuwali, nawet jeśli sądzą, że wiedzą i pamiętają to dokładnie.

Niedoskonałe narzędzia pomiaru (np. odpowiedzi na pytanie) są jednak lepsze niż brak narzędzi (użycie

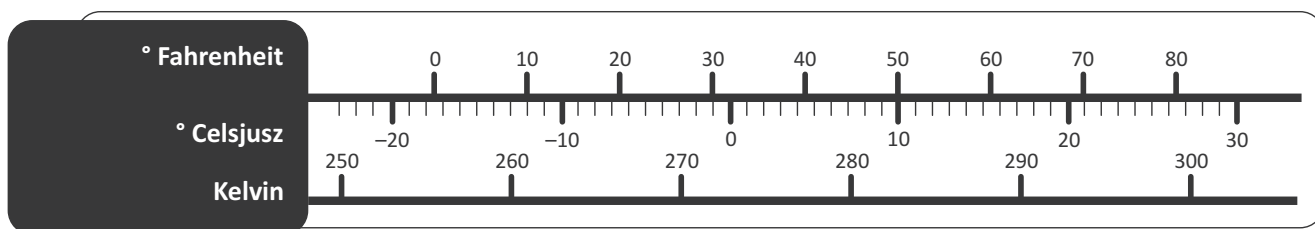
kiepskiej maszyny do szycia jest lepsze niż szycie garnituru ręcznie). Natura doznań subiektywnych każe nam z pokorą sądzić, że powstanie niezawodnego przyrządu pozwalającego zmierzyć z całkowitą precyzją szczęście odczuwane przez daną osobę jest bardzo mało prawdopodobne. Nawet interpretacja wyników obrazowania mózgu zależy od współpracy analizowanej osoby. Ale doskonałych narzędzi nie mają także inne dyscypliny. Pomiar dokonywane we wszystkich dyscyplinach (z definicji empirycznych), np. w fizyce, też zawsze zawierają błąd pomiaru. Niedoskonałości pomiaru to problem, który dyskwalifikuje nasze wyniki tylko wówczas, kiedy ich nie dostrzegamy. Jeśli zdajemy sobie sprawę z nieuniknionych błędów relacji na temat subiektywnych doznań, możemy starać się o dokonanie odpowiednich korekt.

Skoro każdy człowiek ma trochę inaczej skalibrowane skale szczęścia, zaufania i ich opisu, to nie jesteśmy w stanie porównać twierdzeń dwóch osób. Jeśli ocenę ogólnie nieufnej osoby A, określającej swoje zaufania do X na poziomie 3, porównywalibyśmy z identyczną oceną osoby B, obdarzającej innych hojnie zaufaniem, to moglibyśmy mylnie sądzić, że oceny te są identyczne, nie zdając sobie sprawy z faktu, że wartość 3 to najwyższa nota, jaką przyznała osoba A, a najniższa dla osoby B. Jednakże gdy mamy dane pochodzące z dużo większej próby, błędy spowodowane przez odmienne skalibrowanie skal różnych osób zaczynają się nawzajem neutralizować. Gilbert (2007) ilustruje to obrazowo za pomocą opowieści o pracownikach fabryki, która produkuje wszystkie linijki (wyskalowane pręty miernicze) świata i wytwarza miliony przyrządów pomiarowych nieco różniących się od siebie skalą. Porównanie pojedynczych pomiarów obiektów przy pomocy nietrzymających standardów przyrządów nie pozwala być pewnym, że X jest większy od Y. Ale jeśli setki osób z setkami linijek w dłoniach dokonałyby pomiaru każdego z tych obiektów, to moglibyśmy te pomiary uśrednić i czuć się pewniejsi, że X jest jednak większy od Y. Bo prawdopodobieństwo, że wszyscy ludzie, którzy zmierzili X, posługiwali się linijkami o rozciągniętej skali, a ci, którzy zmierzili Y, tymi o skali ściśniętej, jest minimalne.

W przypadku badań międzykulturowych zagrożeniem jest sytuacja, w której linijki używane w danym kraju różnią się w sposób systematyczny od linijek używanych do pomiaru w drugim kraju. Dlatego tak ważne jest sprawdzenie *równoważności pomiaru*, które polega na porównaniu jednostki i punktu zakotwiczenia skali odpowiedzi (*origin*; van de Vijver, Leung, 1996). Choć to mało psychologiczny przykład, najłatwiej wyjaśnić to w odniesieniu do pomiaru masy i temperatury. W Polsce wagę (masę) określamy w kilogramach, w USA w funtach. Choć obie te skale pomiarowe mają to samo zero, różnią się jednostką. Informacja, że Cezary przytył 10 kg jest równoważna stwierdzeniu, że jego waga zmieniła się aż o 22,04 funta. Podobnie w Polsce do pomiaru temperatury używa się skali Celsjusza, w USA skali Fahrenheita, fizycy zaś posługują się skalą Kelvina. Jeżeli zapytamy znajdujących się w tym samym miejscu trzech mężczyzn: Cezarego (Polaka), Franka (Amerykanina) i Krzyśka (fizyka) o temperaturę, to bezpośrednie porównywanie ich odpowiedzi nie będzie miało sensu. Choć obiekt pomiaru jest ten sam, liczbowe różnice w odpowiedziach będą bardzo duże.

Gdybyśmy mieli porównać ich odpowiedzi na pytanie, o ile wzrosła temperatura w czasie ostatniej doby, odpowiedź Krzyśka, używającego skali Kelvina, byłaby tożsama z odpowiedzią Cezarego, bo jednostka na używanych przez nich skalach jest taka sama, choć różny punkt zakotwiczenia (zero). Odpowiedź Franka byłaby istotnie odmienna, ponieważ skala Fahrenheita różni się od pozostałych zarówno jednostką, jak i punktem zakotwiczenia, co można odczytać z równań pozwalających na przeliczanie wartości (zgodnie ze wzorami: $K = 273,15 + C$ oraz $F = 32 + [9/5] \times C$) i zobaczyć na rysunku 1.

W przypadku równoważności jednostki pomiaru (tak jak w skalach Celsjusza i Kelvina) różnice pomiędzy dwoma wynikami (np. wynikami dwóch uczniów czy jednej osoby badanej w dwóch punktach czasowych) mogą być porównywane zarówno wewnątrz, jak i pomiędzy kulturami, podczas gdy pojedyncze wyniki mogą być jedynie porównywane wewnątrz danej kultury. Z tego powodu niektóre testy inteligencji, charakteryzujące się różnymi



Rysunek 1. Porównanie pomiaru temperatury za pomocą trzech różnie skalibrowanych skal (20°C odpowiada 68°F i 293,15 K).

Tabela 2

Zestawienie możliwości porównań przy różnie skalibrowanych skalach

Skale używane w obu kulturach mają:	Bezpośrednie porównanie między wynikami w obu kulturach	Bezpośrednie porównania różnic (np. wielkości przyrostu, spadku) między kulturami
Ten sam punkt zakotwiczenia, tę samą jednostkę, np. wiek respondenta, liczba dzieci	Możliwe	Możliwe
Różny punkt zakotwiczenia, tę samą jednostkę, np. temperatura na skali Celsjusza i Kelvina	Niemożliwe	Możliwe
Ten sam punkt zakotwiczenia, różną jednostkę, np. waga w kg i funtach, odległość w km i milach	Niemożliwe	Niemożliwe
Różny punkt zakotwiczenia i różną jednostkę, np. temperatura na skali Celsjusza i Fahrenheita	Niemożliwe	Niemożliwe

punktami zakotwiczenia skali inteligencji w różnych kulturach, mogą być trafnie zastosowane jedynie wewnątrz grupy. Jeśli możemy stwierdzić, że wyniki posiadają nie tylko identyczną jednostkę pomiaru, ale i wspólny punkt zakotwiczenia, możemy porównywać otrzymane wyniki zarówno wewnątrz, jak i pomiędzy grupami kulturowymi. Przykładami takich zmiennych są np. wiek respondenta czy liczba dzieci.

W przypadku zmiennych psychologicznych również czujemy, że ludzie używają różnie skalibrowanych skal, np. do oceny własnej „temperatury emocjonalnej”, czyli stanu emocjonalnego. Choć używanie skal numerycznych stało się ostatnio popularne przez zadawanie pytań typu: „Jak oceniłbyś swoje zaangażowanie na skali od 1 do 10?”, to trudno byłoby napisać równania – jak dla przeliczania funtów na kilogramy czy stopni Celsjusza na Fahrenheita – do porównania odpowiedzi różnych osób, choć ewidentnie różnią się one sposobem używania danej skali. Problem oceny tożsamości jednostki i punktu zakotwiczenia skali nie jest tak abstrakcyjny, jak mogłoby się Czytelnikowi wydawać. Coraz częściej do oceny projektów badawczych powołuje się międzynarodowe zespoły recenzentów. Opinie Amerykanów są zazwyczaj bardziej entuzjastyczne niż opinie Europejczyków (Wieczorkowska, 1995), co może świadczyć o przesunięciu punktu zakotwiczenia skali. Jeżeli jednak odległości między opiniowanymi projektami są takie same w obu kulturach, to możemy dokonać porównania opinii różnicy w ocenie dwóch kandydatów. W przypadku pomiarów psychologicznych wykazanie równoważności punktu zakotwiczenia i jednostki skali jest niezwykle trudne. Łatwiej jest podważyć niż udowodnić tego typu równoważność. Możliwości dokonywania porównań wyników

w zależności od tożsamości punktu zakotwiczenia i jednostki skali zestawiono w tabeli 2.

Warto podkreślić, że brak równoważności pomiaru nie uniemożliwia wykazania równoważności struktury (*structural equivalence*), na którą jest ukierunkowana część badań międzykulturowych. Badania zorientowane na porównywanie struktury (*structure-oriented design*) skupiają się na zależnościach pomiędzy zmiennymi i starają się zidentyfikować międzykulturowe podobieństwa i różnice w tych zależnościach, sprawdzając na przykład, czy struktura inteligencji jest uniwersalna. Warto uświadomić sobie, że liczenie współczynnika korelacji oznacza standaryzację zmiennych, więc dla omawianego przykładu współczynniki korelacji między pomiarami temperatury wyrażonymi w skalach Celsjusza, Fahrenheita i Kelvina będą wynosić 1. Oznacza to, że przy braku równoważności pomiaru możemy wciąż uzyskać równoważność strukturalną, którą ocenia się przez interkorelacje pozycji dotyczących danego konstruktów teoretycznego. Równoważność strukturalna jest zazwyczaj oceniana przy pomocy skalowania wielowymiarowego, analizy czynnikowej lub modelowania strukturalnego. Otrzymanie zbliżonych struktur czynnikowych pozwala na stwierdzenie identyczności konstruktów psychologicznych leżących u podstaw danego narzędzia.

PORÓWNYWALNOŚĆ PRÓB

Kolejnym ważnym problemem badań porównawczych może być niekompatybilność porównywanych danych. Omawiane w tekście sondaże zostały przeprowadzone przez komercyjne firmy na próbach określanych przez nie jako reprezentatywne, możemy więc zakładać, że nie zostały popełnione błędy podczas doboru respondentów (błędny operat, błędna lub nieodpowiednia do wymagań

procedura losowania). Reprezentatywność próby użytkownik może sprawdzić jedynie przez porównanie rozkładów zmiennych socjodemograficznych z informacjami dostępnymi z innych źródeł. Niestety choć takie próby losowane są z populacji bardzo starannie, systematycznie maleje stopień ich realizacji. W roku 1995 Polska mogła szczerzyć się jednym z największych, bo 81-procentowym poziomem realizacji badań ISSP (1995). W 2008 roku było to już tylko 45% (1263 wśród wylosowanych 2800; dla poszczególnych krajów por. <http://www.gesis.org/en/home>). Jerzyński (2008) pokazał, że wnioskowanie z 30-procentowej próby (a tyle np. wyniósł stopień realizacji próby ISSP w USA w 1995 roku) może łatwo prowadzić do fałszywych wniosków. Jak pisze, w każdej populacji występuje grupa jednostek, które nie wezmą udziału w badaniu, ponieważ nie mają ochoty, możliwości czasowych lub poznawczych, aby odpowiadać na pytania. Określane są one mianem jednostek niedostępnych. Podczas losowania pewna proporcja z tej grupy trafi do próby, co oznacza, że zawsze pewna część próby losowej nie zostanie zrealizowana. Może to spowodować przekroczenie prognozy minimalnej liczebności próby, który jest konieczny do spełnienia wymagań warunkujących dokładność wnioskowania. Niestety nie można tego problemu rozwiązać przez uzupełnienie próby – dolosowanie

respondentów – ponieważ podział populacji na część dostępną i niedostępną nie jest losowy. Wyniki wielu badań (Hilgard i in.; za: Jerzyński, 2008) wskazują, że niedostępność jednostki jest często silnie związana z jej innymi cechami ważnymi dla celu badania. Chcąc dolosować próbę, losujemy z całej populacji, a nie tylko z części niedostępnych respondentów. Na podstawie informacji pochodzącej tylko z grupy dostępnych respondentów nie możemy generalizować obrazu struktury cech na całą populację. Siła zniekształceń wyników analiz rośnie wraz ze spadkiem poziomu realizacji próby.

Niestety tendencja spadkowa dotyczy większości krajów, także Włoch i Japonii (International Social Survey Programme, 1998, 2008), ponieważ respondenci zdają sobie sprawę, że badania sondażowe stały się biznesem dla wielu firm i nie chcą poświęcać im nieodpłatnie swojego czasu. W związku z tym, że tak wiele wylosowanych osób odmawia udziału w badaniach, jak wynika z przykładów zawartych w tabeli 3, w niektórych krajach losuje się od razu więcej prób reprezentatywnych lub dokonuje się dolosowywania, co – jak pisaliśmy wcześniej – nie jest dobrym rozwiązaniem, choć pozwala na osiągnięcie zakładanej liczebności.

W tabeli 3 pokazano ponadto, jak na stopień realizacji próby wpływa także sposób zbierania danych. Wywiady

Tabela 3
Stopień realizacji próby w wybranych krajach

Kraj	Rok	Metoda	Liczebność wylosowanej próby	Liczebność zrealizowanej próby	Poziom realizacji (%)
Japonia	1998	Ankieta wypełniana samodzielnie przez respondenta	1 800	1368	76
Japonia	2008	Ankieta wypełniana samodzielnie przez respondenta	1 800	1200	67
Włochy	1998	Wywiad ankietarski	1 369	1009	74
Włochy	2008	CAPi (<i>Computer assisted personal interview</i>)	6 900*	1078	16
Polska	1998	Ankieta wypełniana samodzielnie przez respondenta	1 708	1147	67
Polska	2008	CAPi (<i>Computer assisted personal interview</i>)	2 800*	1263	45
Kanada	1998	Ankieta pocztowa	4 500	974	22
Kanada	2008	Ankieta pocztowa	160	421	10
Czechy	1998	Wywiad ankietarski	3 120	1223	39
Czechy	2008	Wywiad ankietarski	3 460	1512	44
Francja	1998	Ankieta pocztowa	11 000	1133	10
Francja	2008	Ankieta pocztowa	10 000	2564	26

* Podana jest łączna liczba zawierająca substytucje i dolosowywanie respondentów.

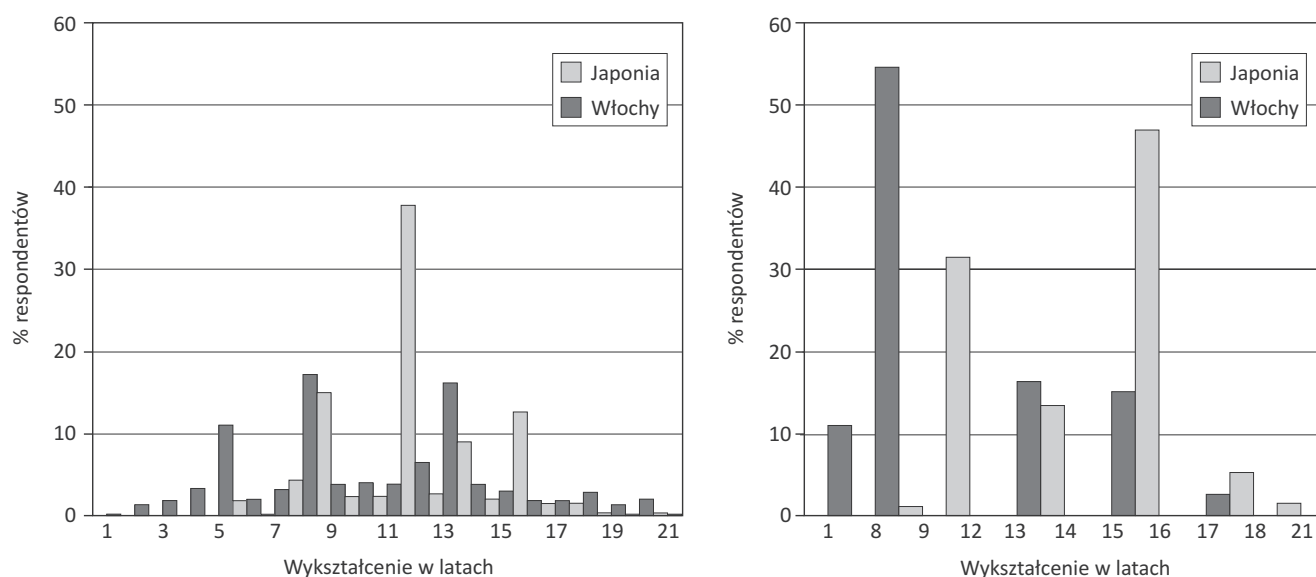
ankieterskie z kwestionariuszem papierowym lub programem komputerowym, w którym badacz zadaje pytania i wprowadza od razu odpowiedzi do komputera (*computer assisted personal interview*), dają o wiele mniej odmów niż wręczenie lub wysłanie pocztą respondentowi ankiety do samodzielnego wypełnienia. Stopień realizacji próby w przypadku badań pocztowych lub na niezobligowanych do udziału internautach jest bardzo niski. W naszych badaniach internetowych przeprowadzające je agencje miały podpisane umowy z wylosowanymi wcześniej próbami reprezentatywnymi (udział w wielu badaniach w zamian za korzyści materialne – notebooki).

Poziom realizacji próby niesie dużo większe zagrożenie dla badań sprawozdawczych, których celem jest porównywanie procentów odpowiedzi, niż dla sondazy analitycznych (Wieczorkowska, Wierzbiński, 2011). Dla potrzeb badań porównawczych ważniejsza od reprezentatywności próby dla populacji jest porównywalność prób. Zarówno w naszym badaniu, jak i w sondażu ISSP (1998, 2008) próba włoska nie różniła się od japońskiej pod względem wieku i proporcji kobiet. Porównanie poziomu wykształcenia wykazało natomiast istotną przewagę Japończyków. Rozkłady wykształcenia w obu próbach znajdują się na rysunku 2. Pominięto na nim procent respondentów deklarujących status ucznia lub studenta (w próbie ISSP odpowiedź „nadal w szkole” wybrało 3,7% Japończyków i 1,2% Włochów, natomiast „nadal na uniwersytecie” – 2,6% Japończyków i 3,2% Włochów).

Ponadto część próby włoskiej przyznała się do braku formalnej edukacji (ISSP – 2,6%; badania własne – 0,5%). W próbie japońskiej takie odpowiedzi nie wystąpiły. Obserwowane różnice w wykształceniu wynikać mogą z odmienności systemów edukacji. Chociaż w obu krajach edukacja jest powszechna i obowiązkowa (w Japonii do piętnastego, we Włoszech od 2006 r. do szesnastego roku życia), istnieją rozbieżności dotyczące długości każdego z etapów nauki (J Education, 2012; Italian Education, 2012). We Włoszech szkoła podstawowa trwa pięć lat, szkoła średnia pierwszego stopnia trzy lata, szkoła średnia drugiego stopnia (lub szkoła zawodowa) pięć lat, licencjat cztery, magisterium zaś dwa lata (w sumie 19 lat nauki). W Japonii szkoła podstawowa jest dłuższa (sześć lat), ale szkoły średnie obydwu stopni trwają po trzy lata, licencjat otrzymać można po czterech latach nauki, a tytuł magistra po dodatkowych dwóch (w sumie 18 lat). Wynikające z tych różnic przesunięcia można łatwo zaobserwować na rysunku 2.

RÓŻNICE W STYLU ODPOWIADANIA NA PYTANIA (*RESPONSE STYLE*)

Międzykulturowe badania postaw polegają zazwyczaj na porównywaniu średnich z odpowiedzi na zestaw pytań. Takie podejście zakłada, że odpowiedzi respondenta zależą wyłącznie od merytorycznego znaczenia pytania. Najczęściej prosimy o wyrażenie opinii za pomocą narzuconej przez badacza skali odpowiedzi. Niestety wybór



Rysunek 2. Porównanie wykształcenia (przez liczbę lat nauki) w dwóch reprezentatywnych próbach, włoskiej i japońskiej, w badaniach ISSP (po lewej) i internetowych (po prawej).

skali odpowiedzi na pytania często jest dokonywany przez badacza bezrefleksyjnie. Skale wykorzystywane w badaniach różnią się m.in. długością i obecnością (lub nie) odpowiedzi środkowej i beztreściowej (*nie mam zdania, trudno powiedzieć*; Wieczorkowska-Wierzbińska, Wierzbiński, 2011). To, że respondent otrzymuje pięciopunktową skalę odpowiedzi, np. przy ocenie ważności różnych celów życiowych, nie oznacza, że w taki właśnie sposób prywatnie dokonuje ich różnicowania. Jego wartościowanie może być dużo bardziej zgrubne (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. Oznacza to, że każdy respondent musi transponować swoją prywatną skalę różnicowania w dostarczoną przez badacza skalę 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.

Braki odpowiedzi są zmorą każdego badania. Oprócz odmowy odpowiedzi lub pominięcia kolejnego pytania dużo problemów dostarczają odpowiedzi *nie wiem, trudno powiedzieć*, które nazywane są odpowiedziami beztreściowymi (Sułek, 2002) i oznaczane skrótem LOB. Odpowiedzi beztreściowe pojawiać się mogą w danych sondażowych w trzech typach pytań: (1) pytanie standardowe (odpowiedź LOB nie jest odczytywana przez ankietę, natomiast jest zapisywana, jeśli respondent udzieli jej spontanicznie), (2) pytanie z quasi-filtrem (możliwa odpowiedź LOB jest odczytywana respondentowi), (3) 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.).

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, ze znudzenia wywiadem, jak i z chęci ucieczki respondenta przed odpowiedzią na niewygodne dla niego pytanie. Mogą też sugerować, że skala odpowiedzi nie uwzględnia adekwatnej opcji. Liczba odpowiedzi LOB zależy od charakteru pytania – jest ich mniej, gdy pytamy o łatwo obserwowalne zachowania. Na przykład w pytaniach o to, jak często modli się

respondent (por. Wierzbiński, 2009), na skali od *nigdy* do *parę razy dziennie* było 5,4% odpowiedzi LOB, gdy pytano zaś o częstość uczestniczenia w nabożeństwach – tylko 1,4%. Badania Krosnicka ze współpracownikami (2002) wykazały, że liczba wybranych LOB koreluje negatywnie z miarą samooceny wysiłku wkładanego w odpowiadanie przez respondenta na pytania, co sugeruje, że udzielanie odpowiedzi beztreściowych 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 badania pokazujące formułowanie przez respondentów opinii na temat nieistniejących obiektów. Takie odpowiedzi Sułek (2002) 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.

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 tendencjach do potakiwania i zaprzeczania (Baumgartner, Steenkamp, 2001).

Innym stylem odpowiadania jest transformacja wewnętrznej skali oceniania tylko w skrajne wartości skali odpowiedzi. Tendencja do udzielania skrajnych odpowiedzi bywa interpretowana (Baumgartner, Steenkamp, 2001) jako przejaw sztywności, braku tolerancji dla wieloznaczności lub dogmatyzmu i może być związana z wyższym poziomem lęku, choć strach przed nietrafnością powinien prowadzić do udzielania odpowiedzi środkowych. Można przypuszczać, że jest przejawem słabiej wyartykułowanych struktur poznawczych. Jest silniejsza dla ważnych, angażujących emocjonalnie bodźców.

W tekście skoncentrujemy się na analizie różnic w stosowaniu dwóch strategii odpowiadania na pytania, przejawiających się w: (1) częstym unikaniu wyrażania opinii za pomocą wyboru odpowiedzi beztreściowej (*nie mam zdania, trudno powiedzieć*), (2) preferowaniu odpowiedzi skrajnych (*extreme response style*) vs. środkowych (*mid-point response style*).

Będziemy więc starali się odpowiedzieć na pytanie, czy Włosi różnią się od Japończyków pod względem częstości udzielania odpowiedzi beztreściowych i skrajnych. Nie znaleźliśmy w literaturze informacji na temat międzykulturowych różnic w liczbie odpowiedzi beztreściowych.

Opisywane porównania międzykulturowe w preferencji dla udzielania skrajnych odpowiedzi pokazały, że więcej odpowiedzi skrajnych udzielają:

(1) w USA zarówno Latynosi, jak i Afroamerykanie niż Amerykanie pochodzenia europejskiego (Clarke III, 2000),

(2) Amerykanie i Kanadyjczycy niż Japończycy (Shiomi, Loo, 1999; Takahashi, Ohara, Antonucci, Akiyama, 2002),

(3) Amerykanie niż Koreańczycy (Chun, Campbell, Yoo, 1974; Lee, Green, 1991),

(4) Grecy niż Włosi i Hiszpanie, którzy z kolei udzielają ich więcej niż Brytyjczycy, Niemcy i Francuzi (van Herk, Poortinga, Verhallen, 2004).

Z kolei badanie przeprowadzane na dużej reprezentatywnej próbie konsumentów z 11 krajów Unii Europejskiej

(Baumgartner, Steenkamp, 2001) pokazało, że w porównaniu z wariancją pomiędzy różnymi zestawami pytań (skalami), wariancja pomiędzy krajami była niewielka. Ta rozbieżność doniesień zachęciła nas do sprawdzenia różnic w stylach odpowiedzi operacjonalizowanych przez liczbę odpowiedzi beztreściowych i skrajnych w analizowanych zbiorach.

Analiza formalnej charakterystyki stylu odpowiadania na pytania zakłada abstrahowanie od treści pytań. Do przeprowadzenia analizy wyodrębniono więc z badania internetowego 20 pytań z jedną skalą odpowiedzi i 18 pytań z inną. Podobnie z badań ISSP (z lat 1998 i 2008) wyodrębniono cztery zestawy pytań z różnymi skalami odpowiedzi.

Liczba odpowiedzi beztreściowych

W analizowanych badaniach opcja odpowiedzi beztreściowej (*nie wiem* lub *trudno powiedzieć*) była dostępna na skali odpowiedzi. Możemy więc dla każdego badanego policzyć liczbę odpowiedzi beztreściowych (LOB) w osobnych zestawach pytań – czyli to, ile razy wykorzystał możliwość uchylecia się od zajęcia stanowiska.

Tabela 4

Wyniki przeprowadzonych porównań liczby odpowiedzi beztreściowych (LOB)

A	B	C	D	E	F	G
Badanie	Liczba pytań	Wszystkie beztreściowe (%)	Włochy <i>M</i> (s)	Japonia <i>M</i> (s)	Porównanie LOB	Istotne kowarianty LOB
1	Zestaw I.1 k = 18	0–0	1,25 (2,07)	1,39 (3,24)	Brak różnic	kobiety ↑ edu+↓
1	Zestaw I.2 k = 20	0,5–3,2	0,93 (2,17)	1,84 (4,17)	Japonia ↑	Brak
2	Skala II.1 k = 4	4,3–17,4	0,44 (1,04)	1,14 (1,58)	Japonia ↑	kobiety ↑
2	Skala II.2 k = 5	0,4–3,8	0,09 (0,45)	0,69 (1,23)	Japonia ↑	kobiety ↑ edu+↓
2	Skala II.3 k = 4	0,5–3,4	0,08 (0,39)	0,36 (0,91)	Japonia ↑	kobiety ↑ edu+↓
2	Skala II.4 k = 5	0,1–1,4	0,25 (0,63)	0,63 (1,09)	Japonia ↑	starsi ↓ kobiety ↑ edu+↓

Kolumna A – (1) internetowe badanie włoskiej i japońskiej prób reprezentatywnych (2009), (2) ISSP w latach 1998 i 2008, powtarzano te same pytania; kolumna B – zestaw pytań z tą samą skalą odpowiedzi, k – liczba pytań; kolumna C – procent respondentów, którzy udzielili wszystkich odpowiedzi beztreściowych w danym zestawie – pierwsza liczba dotyczy próby włoskiej, druga japońskiej; kolumna D – średnie i odchylenia standardowe LOB we Włoszech; kolumna E – średnie i odchylenia standardowe LOB w Japonii; kolumna F – kierunek istotnej różnicy w LOB między krajami (przy kontroli kowariantów – zmiennych socjodemograficznych), symbol „Japonia↑” oznacza, że było więcej LOB w Japonii niż we Włoszech; kolumna G – kierunek istotnej zależności LOB od zmiennych socjodemograficznych, symbol „edu+↓” oznacza, że liczba LOB zmniejszała się wraz ze wzrostem wykształcenia (liczba lat nauki).

Okazało się, że w prawie wszystkich zestawach były osoby, które uchyliły się od odpowiedzi. W tabeli 4 w kolumnie C podano procent osób, które udzieliły wszystkich odpowiedzi beztreściowych na pytania z danego zestawu – pierwsza liczba dotyczy próby włoskiej, druga japońskiej. W kolumnach D i E tabeli 4 podano średnie i odchylenia standardowe LOB w obu krajach.

W następnym kroku sprawdzono związek między liczbą odpowiedzi beztreściowych – po zlogarytmizowaniu⁴ – a krajem pochodzenia respondenta, przy kontroli płci, wieku i wykształcenia. Kierunek stwierdzonych istotnych różnic został przedstawiony w kolumnie F tabeli 4. Symbol „Japonia↑” oznacza, że więcej odpowiedzi beztreściowych było w Japonii niż we Włoszech. Kierunek zależności liczby odpowiedzi beztreściowych ze zmiennymi socjodemograficznymi przedstawiony jest

w kolumnie G tabeli 4. Symbol „edu+↓” oznacza, że liczba LOB zmniejszała się wraz ze wzrostem wykształcenia (liczba lat nauki), symbol „kobiety↑” oznacza, że mężczyźni udzielili takich odpowiedzi mniej.

Przeprowadzone analizy wykazały, że – z wyjątkiem zestawu pierwszego – Japończycy uchyłali się od zajęcia stanowiska istotnie częściej niż Włosi.

Liczba odpowiedzi skrajnych w obu kulturach

Dla każdego z analizowanych sześciu zestawów pytań policzono – dla każdego respondenta – liczbę odpowiedzi skrajnych, np. *bardzo, zdecydowanie tak* lub *zdecydowanie nie, bardzo prawdopodobne, bardzo nieprawdopodobne*. Przed analizą wykluczono respondentów, którzy udzielili zbyt dużej liczby odpowiedzi beztreściowych w ramach jednego zestawu pytań. Określenie „zbyt duża”

Tabela 5
Wyniki przeprowadzonych porównań liczby odpowiedzi skrajnych (LOS)

A	B	C	D	E	F	G
Badanie	Liczba pytań	Wykluczeni (%)	Włochy M (s)	Japonia M (s)	Porównanie LOS	Istotne kowarianty LOS
1	Zestaw I.1 k = 18	0,7–3	7,22 (3,48)	7,02 (3,54)	Japonia ↓	starsi ↑
1	Zestaw I.2 k = 20	0,7–5	9,61 (5,63)	5,77 (4,88)	Japonia ↓	starsi ↑ kobiety ↑
2	Skala II.1 k = 4	12,4–30,6	1,92 (1,69)	1,02 (1,49)	Japonia ↓	starsi ↑
2	Skala II.2 k = 5	0,9–9,6	1,47 (1,42)	2,57 (1,54)	Japonia ↑	edu+↑
2	Skala II.3 k = 4	1,6–10	0,60 (0,99)	0,47 (0,83)	Brak różnic	starsi ↓ edu+↓
2	Skala II.4 k = 5	1,4–7,3	3,00 (1,39)	1,66 (1,33)	Japonia ↓	starsi ↑ edu+↓

Kolumna A – jak w tabeli 4; kolumna B – jak w tabeli 4; kolumna C – procent respondentów, którzy zostali wykluczeni z dalszych analiz ze względu na zbyt dużą liczbę LOB, pierwsza liczba dotyczy próby włoskiej, druga japońskiej; kolumna D – średnie i odchylenia standardowe LOS we Włoszech; kolumna E – średnie i odchylenia standardowe LOS w Japonii; kolumna F – kierunek istotnej różnicy w LOS między krajami (przy kontroli kowariantów – zmiennych socjodemograficznych), symbol „Japonia↑” oznacza, że więcej odpowiedzi skrajnych było w Japonii niż we Włoszech; kolumna G – kierunek istotnej zależności liczby odpowiedzi skrajnych od kowariantów (zmiennych socjodemograficznych), symbol „edu+↓” oznacza, że liczba odpowiedzi skrajnych zmniejszała się wraz ze wzrostem wykształcenia (liczba lat nauki).

⁴ Rozkłady liczby odpowiedzi beztreściowych są zazwyczaj prawoskośne, co utrudnia analizy danych. Modele liniowe dla takich danych (nazywane modelami liczebności) dopasowuje się na podstawie rozkładów z rodziny Poissona. Jerzyński (2008) w swoich analizach, ze względu na występującą nadmierną dyspersję, zastosował modele oparte o rozkład ujemny dwumianowy (*generalized linear model with negative-binomial link function*) i pokazał, że prowadzą one do tych samych konkluzji, co metody wymagające spełnienia założeń normalności rozkładu badanych zmiennych. Brak spełnienia założenia o normalności rozkładu naraża nas na ryzyko, że uzyskane estymatory mogą być obciążone, ale celem prezentowanych tutaj analiz nie jest estymacja parametrów populacji, lecz porównanie wartości statystyk w różnych zbiorach danych. Dlatego zniekształcenia związane z niespełnieniem założeń w podobnym stopniu wpływają na porównywane wyniki. Dla redukcji stopnia prawoskośności rozkładów zmiennych zależnych w testach statystycznych używano logarytmów liczby odpowiedzi.

wymaga doprecyzowania: przy długich zestawach pytań przyjęliśmy, że jest to ponad 50% odpowiedzi, przy skalach pięciopytaniowych więcej niż dwie, przy skali czteropytaniowej więcej niż jedna. Procenty osób wykluczonych według tej reguły znajdują się w kolumnie C tabeli 5.

W następnym kroku sprawdzaliśmy związek liczby odpowiedzi skrajnych (LOS) – po zlogarytmizowaniu – z krajem pochodzenia respondenta, przy kontroli płci, wieku i wykształcenia. W kolumnach D i E tabeli 5 podano średnie i odchylenia standardowe LOS w obu krajach. Kierunek stwierdzonych istotnych różnic między krajami został przedstawiony w kolumnie F tabeli 5. Kierunek zależności liczby odpowiedzi skrajnych ze zmiennymi socjodemograficznymi przedstawiony jest w kolumnie G tabeli 5. Poza jednym zestawem Japończycy udzielili znacząco mniejszej liczby odpowiedzi skrajnych niż Włosi.

INTERPRETACJA STWIERDZONYCH RÓŻNIC

I tu właściwie moglibyśmy zakończyć analizę, podsumowując, że Japończycy w porównaniu z Włochami istotnie częściej uchylają się od udzielania odpowiedzi treściowej, a także rzadziej korzystają z odpowiedzi skrajnej. Wyniki dobrze zaplanowanych badań eksperymentalnych, dzięki manipulacji wartościami zmiennych niezależnych, pozwalają na łatwą interpretację obserwowanych różnic w wartościach zmiennych zależnych. W porównaniach międzykulturowych, gdy nie ma możliwości losowego przydziału do grup, kluczowe jest wyeliminowanie konkurencyjnych hipotez. Tłumaczenie różnic kulturą jest ryzykowne, jeśli nie potrafimy powiedzieć, jakie konkretnie aspekty danej kultury wyjaśniają stwierdzone różnice. Przydatne może być wprowadzanie do analiz kowariantów. W przeprowadzonych analizach kontrolowaliśmy zmienne socjodemograficzne, ponieważ wiemy, że od nich zależy liczba odpowiedzi beztreściowych. Przykładowo we wcześniejszych analizach liczby LOB w ocenach zaufania do instytucji (Wierziński, 2009) pokazano, że opcję LOB częściej wybierają: (1) osoby starsze niż młodsze, (2) kobiety niż mężczyźni, (3) mieszkańcy wsi niż mieszkańcy miast, (4) słabiej niż lepiej wykształceni, (5) gorzej niż lepiej sytuowani, (6) niezainteresowani polityką niż nią zainteresowani. Uzyskano także istotne efekty interakcji zmiennych socjodemograficznych. W analizach naszych danych zależności typu odpowiedzi od zmiennych socjodemograficznych nie są już takie oczywiste. Choć siła związku LOB i LOS z krajem przewyższała o rząd wielkości siłę relacji tych zmiennych z danymi socjodemograficznymi, to trzeba zauważyć, że (opisujemy tylko te zależności, dla których współczynnik regresji był istotnie różny od zera):

- (1) odpowiedzi beztreściowe były wybierane przez:
 - (a) osoby starsze rzadziej w jednym zestawie,
 - (b) kobiety częściej w pięciu zestawach,
 - (c) osoby lepiej wykształcone rzadziej w czterech zestawach,
- (2) odpowiedzi skrajne były wybierane przez:
 - (a) osoby starsze częściej w czterech zestawach, w jednym rzadziej,
 - (b) kobiety częściej w jednym zestawie,
 - (c) osoby lepiej wykształcone rzadziej w dwóch zestawach, częściej w jednym.

Choć przedmiotem analizy są formalne charakterystyki stylu odpowiadania, to ewidentne jest, że należy przyrzeć się zestawom pytań (por. tabela 6).

Pierwszy zestaw pytań (I.1) w badaniach internetowych dotyczył różnych reakcji na kryzys finansowy – sprawy aktualnej w czasie przeprowadzania tego badania w roku 2009, ponieważ aż 55,8% Japończyków i 49,2% Włochów zadeklarowało, że ich sytuacja finansowa pogorszyła się w ciągu wcześniejszych 12 miesięcy.

Zestaw drugi (I.2) dotyczył różnych strategii aktywności, stąd większe zdecydowanie w przypadku kobiet, które są zazwyczaj bardziej zainteresowane psychologicznym samoopisem niż mężczyźni.

Skala pierwsza z sondażu ISSP (II.1) dotyczyła wiary w aksjomaty związane bardziej z religią katolicką niż z buddyzmem. Stąd aż 17,4% Japończyków uchylilo się od odpowiedzi.

Skala druga (II.2) dotyczyła możliwości kierowania własnym życiem, w tym roli Boga, stąd procent osób uchylających się od odpowiedzi w Japonii zmalał.

Skala trzecia (II.3) dotyczyła zaufania do czterech instytucji i tu nie widać powodu, dla którego Japończycy powinni się częściej uchylać od odpowiedzi – a jednak to robili.

Skala czwarta (II.4) dotyczyła postaw moralnych – permissywności seksualnej i postawy wobec aborcji.

Można więc sformułować dość oczywisty wniosek, że liczba odpowiedzi beztreściowych zależy od tego, jak istotna jest dla danej osoby tematyka pytań. Jednak w prawie wszystkich z analizowanych sześciu zestawów pytań Japończycy udzielali mniej odpowiedzi skrajnych i uchylali się od zajęcia stanowiska istotnie częściej niż Włosi. Stosując dychotomię przyjętą przez Nisbetta (2009), możemy myśleć o Włochach jako o przedstawicielach kultury Zachodu, czerpiących swój rodowód od starożytnych Greków, którzy skupiali się w swoich rozważaniach na obiektach traktowanych jako odrębne, niezależne od kontekstu bytu i przez to możliwe do zdefiniowania przez ich abstrakcyjne, często dychotomiczne właściwości. Japończycy natomiast to przedstawiciele

Tabela 6
Zestawy pytań wykorzystane w porównaniach

Numer	Liczba pytań	Skala odpowiedzi	Pytania w polskiej wersji
I.1	18	1 – bardzo nieprawdopodobne 2 – raczej nieprawdopodobne 3 – raczej prawdopodobne 4 – bardzo prawdopodobne ? – trudno powiedzieć	Pytania dotyczyły przewidywanych reakcji na kłopoty finansowe Adama (9 pytań) i Ewy (9 pytań), których sytuacja została opisana we wstępie. Przykładowa pozycja: „Adam będzie obwinił system bankowy”.
I.2	20	1 – tak jak X 2 – raczej jak X 3 – raczej jak Y 4 – tak jak Y ? – trudno powiedzieć	Każde pytanie było opisem zachowania osoby X i osoby Y, a zadaniem respondenta była ocena, czy jego zachowanie/reakcje są bardziej podobne do X czy do Y. Przykładowa pozycja: „X uważa, że sukces zależy od szczęścia. Y uważa, że jest to przejaw zdolności i wysiłku”.
II.1	4	1 – zdecydowanie tak 2 – raczej tak 3 – raczej nie 4 – zdecydowanie nie 8 – trudno powiedzieć	Czy wierzy Pan(i) ... (1) w życie po śmierci? (2) w Niebo? (3) w Piekło? (4) w cuda religijne?
II.2	5	1 – zdecydowanie się zgadzam 2 – zgadzam się 3 – ani się zgadzam, ani nie zgadzam 4 – nie zgadzam się 5 – zdecydowanie się nie zgadzam 8 – trudno powiedzieć	Czy zgadza się Pan(i), czy też nie zgadza z poniższymi stwierdzeniami? 1. Istnieje Bóg, który Osobiście zajmuje się każdą istotą ludzką. 2. Ludzie niewiele mogą zrobić, aby zmienić bieg swojego życia. 3. Życie ma dla mnie sens tylko dlatego, że istnieje Bóg. 4. Moim zdaniem życie nie służy żadnemu celowi. 5. Życie ma tylko wtedy znaczenie, gdy ty sam nadasz mu sens.
II.3	4	1 – całkowite zaufanie 2 – duże zaufanie 3 – umiarkowane zaufanie 4 – bardzo małe zaufanie 5 – w ogóle nie mam zaufania 8 – trudno powiedzieć	Jak duże zaufanie posiada Pan(i) w stosunku ... (1) do Sejmu? (2) do organizacji przemysłowych i handlowych? (3) do sądów i systemu prawnego? (4) do szkół i systemu edukacji?
II.4	5	1 – zawsze niewłaściwe 2 – prawie zawsze niewłaściwe 3 – tylko czasem niewłaściwe 4 – nie ma w tym nic złego 8 – trudno powiedzieć	1. Jeśli kobieta i mężczyzna rozpoczynają stosunki seksualne przed ślubem, to czy uważa Pan(i), że jest to niewłaściwe, czy też nie ma w tym nic złego? 2. A jakie jest Pana(-i) zdanie na temat stosunków seksualnych w stanie małżeńskim z kimś innym niż jego żona lub jej mąż. Czy jest to...? 3. A czy sądzi Pan(i), że stosunki seksualne między dwiema dorosłymi osobami tej samej płci są...? 4. Czy osobiście sądzi Pan(i), że jest to niewłaściwe, czy też nie ma w tym nic złego, gdy kobieta przerywa ciążę, jeśli... (a) istnieje wielkie prawdopodobieństwo, że dziecko urodzi się z poważnymi wadami? (b) rodzina ma bardzo niskie dochody i nie może sobie pozwolić na więcej dzieci?

kultury Wschodu wywodzącej się z taoizmu, konfucjanizmu i buddyzmu. *Jin* i *jang* nie są przeciwieństwami, ale uzupełniają się i przenikają. Konfucjanizm promuje zasadę złotego środka. Również buddyzm zakłada oddziaływanie każdej rzeczy na niemal wszystkie inne. Wytworem kultury Wschodu jest dialektycyzm, który służy integracji pozornych przeciwieństw i akceptacji odmiennych punktów widzenia. Jedną z zasad wschodniego myślenia jest zasada sprzeczności: stare i nowe,

dobre i złe, mocne i słabe istnieją we wszystkim. Także wschodnie przysłowia są pełne sprzeczności: „A man is stronger than iron and weaker than a fly” („Człowiek jest silniejszy niż żelazo i słabszy od muchy”; Nisbett, 2009, s. 173). Ludzie Zachodu kierują się w myśleniu zasadą identyczności – istota rzeczy nie zależy od relacji, i zasadą wyłączonego środka – zdanie *p* jest albo prawdziwe, albo fałszywe. Wynika więc z tego, że odpowiadanie na pytania ankiety wymaga refleksji zgodnej z zachodnim

stylem myślenia. Stwierdzona w naszych analizach większa liczba odpowiedzi beztreściowych i mniejsza skrajnych u przedstawicieli kultury Wschodu (Japończycy) niż Zachodu (Włosi) jest zgodna z tą interpretacją.

Trzeba jednak zauważyć, że większą liczbę odpowiedzi beztreściowych w Japonii może tłumaczyć inny niż we Włoszech sposób przeprowadzania badań ISSP. Jak wynika z dokumentacji, odpowiedzi na pytania ISSP we Włoszech zostały zebrane za pomocą wywiadu prowadzonego przez ankietera w kontakcie bezpośrednim. Kolejność odpowiadania jest dla wszystkich respondentów stała, a czas namysłu nad odpowiedzią jest ograniczony przez przyjęte przez ankietera normy. Mamy zatem do czynienia z efektem interakcji między respondentem i ankieterem (por. Zieliński, 2009), ponieważ odpowiadając na pytania, respondent bierze pod uwagę – świadomie lub nie – jakie wrażenie wywierają jego odpowiedzi na ankieterze. W Japonii ankieter zostawiał respondentowi ankietę do samodzielnego wypełnienia i przychodził po nią później – nie kontrolował więc ani kolejności, ani czasu odpowiadania na pytania. Można więc oczekiwać, że mniej odpowiedzi beztreściowych będzie pojawiać się w czasie wywiadu ankietarskiego niż w czasie samodzielnego wypełniania kwestionariusza. To tłumaczyłoby różnice między Japonią i Włochami w badaniach ISSP, ale nie tłumaczy tego samego kierunku różnic w badaniach internetowych. Co prawda w pytaniach o kryzys finansowy nie stwierdzono istotnych różnic, ale to z kolei wynika ze specyfiki owych pytań (dotyczyły sprawy istotnej oraz reakcji innych ludzi, nie były więc zagrażające).

W sondażu internetowym tempo i kolejność odpowiadania na pytania (jeśli możliwy jest powrót do poprzedniej strony i zmiana odpowiedzi) są całkowicie uzależnione od odpowiadającego. Choć w internecie można narzucić badanemu zarówno czas, jak i kolejność udzielania odpowiedzi, to najczęściej się tego nie robi.

Kolejna alternatywna interpretacja stwierdzonych różnic wskazuje na niedopasowanie kulturowe zadanych pytań w skalach II.1 i II.2 badania ISSP. Mimo wielkiego doświadczenia i wielkiej staranności międzynarodowego zespołu ISSP widzimy w przygotowaniu pytań wyraźną dominację badaczy z kultury judeochrześcijańskiej. W odpowiedzi na pytanie o religię, w której respondent został wychowany, 96,8% Włochów wskazało katolicyzm, 2,4% ateizm (brak religii), a 0,5% kategorię *inne wyznanie*. W Japonii ateizm zadeklarowało aż 52,4% osób, buddyzm i inne religie wschodnie 31% osób, 2,9% zaś inne wyznanie. Udzielenia odpowiedzi na to pytanie odmówiło 25 Japończyków, opcję *nie wiem* wybrało zaś tylko troje Włochów i aż 317 Japończyków. Z tych różnic w religijności wynika, że pytania o wiarę

w cuda religijne, rolę Boga, Biblię⁵ itd. mogły być przez Japończyków odebrane jako niegrzeczne, niezrozumiałe lub nieprzemyślane i tym samym wpłynąć nie tylko na tę konkretną odpowiedź, ale na wszystkie pozostałe odpowiedzi. Trzeba bowiem pamiętać, że każde pytanie ma do spełnienia funkcję ukrytą: powinno motywować do współpracy. Pytania dotyczące religii takiej funkcji w Japonii nie spełniały. Ta interpretacja nie może jednak tłumaczyć różnic w odpowiedziach beztreściowych w badaniu internetowym i różnic w odpowiedziach skrajnych. Warto też dodać, że gdy w 2008 roku dołączono do zestawu pytań ISSP pytania o wiarę w reinkarnację i nirwanę, które uzyskały więcej odpowiedzi pozytywnych (*zdecydowanie tak i raczej tak*) w Japonii (8% + 34,1% dla reinkarnacji; 5,9% + 30,4% dla nirwany) niż we Włoszech (gdzie odpowiedzi *zdecydowanie nie i raczej nie* było odpowiednio: 34,9% + 27,5% i 34,1% + 19,6%), to nadal liczba LOB była większa w Japonii (24,2% i 36,3%) niż we Włoszech (13,8% i 31,7%).

Odrzucenie alternatywnych wyjaśnień skłania do uznania stwierdzonych różnic kulturowych w stylach odpowiadania za prawdziwe, z podkreśleniem modyfikującego wpływu treści pytań. Warto jednak zaznaczyć, że zarówno liczby odpowiedzi beztreściowych, jak i liczby odpowiedzi skrajnych korelowały między zestawami, co pokazuje, że zarówno liczba odpowiedzi skrajnych, jak i liczba odpowiedzi beztreściowych powinna być analizowana także jako cecha stylu odpowiadania respondenta, a nie tylko pytania. Nieuwzględnienie różnic w stylach odpowiadania może prowadzić do zniekształconych wniosków, co zostało pokazane zarówno w badaniach porównawczych polskich i niemieckich nastolatków (Wieczorkowska, 1993), jak i w badaniach marketingowych (Baumgartner, Steenkamp, 2001).

FAŁSZYWI RESPONDENCI

Na zakończenie, choć jest to zawsze pierwsza analiza, od której rozpoczynamy pracę, warto powiedzieć o jednym z większych zagrożeń dla trafności analiz. Pisaliśmy wcześniej, że rosnąca we wszystkich krajach liczba odmów udziału w badaniach sprawia duże problemy dla reprezentatywności badanych prób, które jednak można starać się korygować, np. przez ważenie poststratyfikacyjne (por. Jerzyński, 2008). Większym problemem jest jednak *pozorna zgoda*, polegająca na udzielaniu losowych odpowiedzi. Jest to szczególnie prawdopodobne w sytuacji, gdy respondent wypełnia ankietę samodzielnie (np. tylko w obecności ankietera lub przez internet) i sprawdza się

⁵ Główne wyznawane w Japonii religie (buddyzm i shinto) nie mają nawet tekstów pełniących rolę Biblii w chrześcijaństwie.

jedynie, czy udzielono odpowiedzi na wszystkie pytania. Dystrybutorzy danych próbują wyłapywać takie niezgodności logiczne jak treściowa odpowiedź na pytanie o satysfakcję z dzieci, gdy wcześniej respondent powiedział, że dzieci nie ma, lub gdy respondent twierdzi, że w pracy kieruje innymi, choć wcześniej odpowiedział, że nie pracuje i nigdy nie pracował. Najczęściej jednak analiza zgodności polega na sprawdzaniu rozkładów poszczególnych zmiennych, a nie na porównywaniu odpowiedzi uzyskanych od tego samego respondenta. Najnowsza metoda przeprowadzania wywiadów (CAPI) w sporej części eliminuje powstawanie nielogiczności w odpowiedziach, ponieważ program komputerowy, który wyznacza kolejne pytanie, ominie pytanie o satysfakcję z dzieci, jeśli odpowiedzią na wcześniejsze pytanie o liczbę dzieci było 0.

W analizach próbuje się także szacować efekt braku zgodności (*non contingent response style*), porównując odpowiedzi na skorelowane pary pytań. Operacyjnie oznacza to liczenie sumy bezwzględnych różnic pomiędzy odpowiedziami respondentów na pary najwyżej dodatnio skorelowanych pozycji o zbliżonych średnich w badanej próbie. Wyjaśnijmy to na przykładzie, zakładając dla uproszczenia wywodu, że skala odpowiedzi jest dwustopniowa (*prawda vs. fałsz*). Wskaźnik niezgodności dla pary pozycji: (A) „Lubię układać plany i decydować o tym, co powinni robić inni ludzie” i (B) „Lubię wydawać innym polecenia i organizować ich pracę”, wyniesie 0, jeśli osoba odpowie w ten sam sposób na oba pytania. W innym wypadku wskaźnik niezgodności będzie równy 1. Przy skali wielostopniowej będzie to wartość bezwzględna odległości obu odpowiedzi na skali. W przypadku pytań skorelowanych negatywnie, takich jak: (C) „Nie lubię znajdować się w centrum uwagi”, (D) „Zazwyczaj odgrywam w swojej grupie rolę przywódcy”, wskaźnik niezgodności wyniesie 1, gdy respondent dwa razy odpowie *prawda* lub dwa razy *fałsz*. Choć takie odpowiedzi są sprzeczne, możemy mieć jednak problem z różnicowaniem efektu niezgodności od efektu potakiwania lub zaprzeczania.

Najprostszym – choć nie jedynym – sposobem wykrywania fałszywych respondentów (pozorujących odpowiadanie na pytania, a w rzeczywistości odpowiadających losowo) jest sprawdzenie częstości wybierania tej samej odpowiedzi na wszystkie pytania z tą samą skalą odpowiedzi. W przypadku skal o niskim stopniu skorelowania pytań (np. oba analizowane zestawy w próbie internetowej) takie zachowanie może świadczyć o braku zainteresowania badaniem. Choć w naszych badaniach internetowych średnia korelacja w dziewięciu pytaniach o reakcję na kryzys finansowy wyniosła 0,059, to jednak znalazły się osoby, które udzieliły identycznych odpowiedzi na

wszystkie pytania. W dwóch zestawach po dziewięć pytań 18 identycznych odpowiedzi „treściowych” udzieliło aż 21 osób. W zestawie 20 pytań o strategię aktywności dla każdej z czterech dostępnych treściowych odpowiedzi znalazły się osoby, które udzieliły identycznych 20 odpowiedzi, choć stopień skorelowania pytań w badanej próbie wyniósł 0,17. W sumie takich osób było 67.

Sprawdzanie częstości wykorzystania poszczególnych punktów skali odpowiedzi powinno służyć eliminowaniu fałszywych respondentów z dalszych analiz. Można tego dokonać jedynie w przypadku różnokierunkowych zestawów pytań (kiedy o nasileniu danej postawy/cechy świadczy „zgoda” z częścią itemów i brak zgody z resztą). Nie można tego zrobić np. w jednokierunkowych zestawach pytań – udzielenie takiej samej odpowiedzi na pytania dotyczące aborcji i permissywności seksualnej może świadczyć o silnych skryzalizowanych postawach, a nie o lekceważeniu badania przez respondenta. Jeżeli jednak respondent kilkanaście razy odpowiada raczej jak wzorcowa osoba X przy słabo skorelowanych pozycjach, powinniśmy być podejrzliwi.

W innych analizach merytorycznych naszych badań internetowych przyjęliśmy zasadę, że jeśli respondent odpowiadający na dwukierunkowe⁶ zestawy pytań używa tego samego klawisza (oznaczającego daną odpowiedź) częściej niż w dwóch trzecich wszystkich odpowiedzi (np. 12 pytań na 18, 13 pytań na 20), to eliminujemy go z dalszych analiz. W 18-elementowym zestawie pytań do badania reakcji na kryzys finansowy spowodowało to wykluczenie z dalszych analiz 10,4% włoskich i 16,3% japońskich respondentów. Wykluczeni nie różnili się od pozostałych wiekiem, płcią ani wykształceniem. Zyskiem z wykluczenia było istotne zwiększenie współczynników korelacji między podobnymi znaczeniowo pozycjami.

W zestawie pytań do oceny strategii aktywności przyjęcie zasady dwóch trzecich spowodowało wykluczenie z dalszych analiz 16,6% włoskich i 20,4% japońskich respondentów. Wśród wykluczonych było więcej mężczyzn i więcej osób starszych niż młodszych, co pewnie można tłumaczyć psychologicznym charakterem pytań.

Fałszywi respondenci są większym zagrożeniem dla sondaży analitycznych (nastawionych na testowanie zależności między zmiennymi) niż brak reprezentatywności próby (która jest niezbędna, gdy naszym celem jest, tak jak w sondażach sprawozdawczych, określenie rozkładu zmiennej w populacji).

⁶ Zestawy pytań dotyczących jednego konstruktów teoretycznego mogą być dwukierunkowe (kiedy o nasileniu danej postawy/cechy świadczy zgoda z częścią itemów i brak zgody z resztą) lub jednokierunkowe (gdy wszystkie pozycje są sformułowane pozytywnie).

Chcielibyśmy, aby przedstawione w tekście rozważania i przykłady skłoniły badaczy do starannej wstępnej analizy danych. Zbyt często można bowiem natknąć się w literaturze fachowej na stosowanie do porównań kulturowych modelowania strukturalnego, które jest dla badaczy bardzo wdzięczną metodą analizy, ponieważ prawie zawsze do danych udaje się dopasować jakiś model. Niestety z drugiej strony skutki przeprowadzania zaawansowanych analiz na danych nieoczyszczonych z fałszywych respondentów dają przykrą ilustrację reguły *garbage in, garbage out* („zanieczyszczenia na wejściu, zanieczyszczenia na wyjściu”).

LITERATURA CYTOWANA

- Baumgartner, H., Steenkamp, J. M. (2001). Response styles in marketing, research: A cross-national investigation. *Journal of Marketing Research*, 38 (2), 143–156.
- Boski, P. (2009). *Kulturowe ramy zachowań społecznych*. Warszawa: Wydawnictwo Naukowe PWN, Wydawnictwo Szkoły Wyższej Psychologii Społecznej Academica.
- Chun, K. T., Campbell, J. B., 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.
- Gilbert, D. (2007). *Na tropie szczęścia* (przeł. E. Rajewska). Poznań: Wydawnictwo Media Rodzina.
- Harkness, J., Langfeldt, B., Sholz, E. (2000). ISSP study monitoring 1996–1998. *Reports to the ISSP General Assembly on monitoring work undertaken for the ISSP by ZUMA, Germany*. Pobrano 12.07.2012 z: <http://info1.gesis.org/dbksearch19/Docs.asp?no=3190>.
- Harzing, A. (2006). Response styles in cross-national survey research. *International Journal of Cross Cultural Management*, 6 (2), 243–266.
- International Social Survey Programme (1995). *National Identity I, 1995*. Pobrano 12.07.2012 z: <http://info1.gesis.org/dbksearch19/Docs.asp?no=2880>.
- International Social Survey Programme (1998). *Religion II, 1998*. Pobrano 12.07.2012 z: <http://info1.gesis.org/dbksearch19/Docs.asp?no=3190>.
- International Social Survey Programme (2008). *Religion III, 2008*. Pobrano 12.07.2012 z: <http://info1.gesis.org/dbksearch19/Docs.asp?no=4950>.
- Italian Education & Schooling in Italy (2012). *Angloinfo. The global expat network. Rome*. Pobrano 12.07.2012 z: <http://rome.angloinfo.com/countries/italy/schooling.asp>.
- J Education (2012). *Multilingual Living Information*. Pobrano 12.07.2012 z: <http://www.clair.or.jp/tagengorev/en/j/index.html>.
- Jerzyński, T. (2008). *Wybrane korelaty liczby odpowiedzi beztreściowych w badaniach sondażowych*. Niepublikowana praca doktorska, Uniwersytet Warszawski, Instytut Socjologii, Warszawa.
- Johnson, T., Kulesa, P., Cho, Y., Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology*, 36 (2), 264–277.
- 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., 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.
- Lee, C., Green, R. T. (1991). Cross-cultural examination of the Fishbein Behavioral Intentions Model. *Journal of International Business Studies*, 22 (2), 289–305.
- Nisbett, R. E. (2009). *Geografia myślenia. Dlaczego ludzie Wschodu i Zachodu myślą inaczej?* (przeł. E. Wojtych). Sopot: Smak Słowa.
- Poortinga, Y. H. (1989). Equivalence of cross-cultural data: An overview of basic issues. *International Journal of Psychology*, 24 (2–6), 737–756.
- Shiomi, K., Loo, R. (1999). Cross-cultural response styles on the Kirton Adaptation-Innovation Inventory. *Social Behavior and Personality*, 27 (4), 413–420.
- Su, C., Parham, D. (2002). Generating a valid questionnaire translation for cross-cultural use. *American Journal of Occupational Therapy*, 56 (5), 581–585.
- Sulek, A. (2002). *Ogród metodologii socjologicznej*. Warszawa: Wydawnictwo Naukowe Scholar.
- Takahashi, K., Ohara, N., Antonucci, T. C., Akiyama, H. (2002). Commonalities and differences in close relationships among the Americans and Japanese: A comparison by the individualism/collectivism concept. *International Journal of Behavioral Development*, 26 (5), 453–465.
- van de Vijver, F. J. R., Leung, K. (1996). Methods and data analysis of comparative research. W: J. W. Berry, Y. H. Poortinga, J. Pandley (red.), *Handbook of cross-cultural psychology* (t. 1, s. 257–300). Boston: Allyn & Bacon.
- van Herk, H., Poortinga, Y. H., 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.
- Wieczorkowska, G. (1993). Pułapki statystyczne. W: Z. Smoleńska (red.), *Badania nad rozwojem w okresie dorastania* (s. 211–234). Warszawa: Instytut Psychologii Polskiej Akademii Nauk.
- Wieczorkowska, G. (1995). Nieporozumienia kulturowe (na przykładzie różnic w skryptach społecznych Polaków i Amerykanów). *Nowiny Psychologiczne*, 3, 21–28.
- Wieczorkowska-Wierzbńska, G., Wierzbński, J. (2011). *Statystyka: od teorii do praktyki*. Warszawa: Wydawnictwo Naukowe Scholar.
- Wierzbński, J. (2009). *Badanie zaufania do organizacji: problemy metodologiczne*. Warszawa: Wydawnictwo Naukowe Wydziału Zarządzania Uniwersytetu Warszawskiego.
- Wierzbński, J. (2012). Comparison of data collected in different countries: Methodological problems. W: A. Z. Nowak (red.), *Understanding global economy. National and regional stu-*

dies (s. 110–118). Warszawa: Wydawnictwo Naukowe Wydziału Zarządzania Uniwersytetu Warszawskiego.
 Zieliński, M. (2009). *Efekt ankierski w procesie konstruowania odpowiedzi na pytania w wywiadzie kwestionariuszowym*.

wym. Studium na podstawie Polskiego Generalnego Sondażu Społecznego. Nieopublikowana praca doktorska, Uniwersytet Warszawski, Instytut Socjologii, Warszawa.

Comparability of survey data collected in different countries

Grażyna Wieczorkowska-Wierzińska^{1,2,3}, Jerzy Wierziński², Anna Olga Kuźmińska²

¹ *Institute for Social Studies, University of Warsaw*

² *Faculty of Management, University of Warsaw*

³ *Faculty of Psychology, University of Social Sciences and Humanities*

ABSTRACT

The incomparability of data collected in different countries is a major hindrance to cross-cultural research. We have focused our paper on the issues of measurement and sampling equivalence, as well as response styles manifested in the differences in the numbers of respondents choosing (i) „don't know” and (ii) extreme responses, e.g. “definitely yes” or “definitely no”. To test cross-cultural differences we used an Italian and a Japanese representative internet samples, as well as data collected in both countries as a part of the International Social Survey Programme (1998, 2008). In almost all of the six question sets analysed, Japanese gave fewer extreme responses, as well as more “don't know” responses than Italians. Results of this comparison are interpreted as an indication of the well known differences in the thinking styles (Nisbett, 2009), which in the Eastern cultures are manifested in the acceptance of contradictions (i.e., good and bad, strong and weak etc., exist in everything), whereas thinking in the Western cultures is guided by the law of excluded middle (i.e., a sentence is either true or false). Our analysis also highlights the necessity of detecting “false respondents,” who answer in a random fashion, and eliminating them from further analyses. Such false respondents are a bigger threat to analytical surveys directed at testing of relationships between variables than is a lack of representativeness of the sample, which is crucial when assessing the distribution of a variable in the population.

Keywords: *measurement equivalence, response style, don't know response, extreme response style, cross-cultural survey research, ISSP, false respondents*

Złożono: 2.04.2012

Złożono poprawiony tekst: 17.01.2013

Zaakceptowano do druku: 21.01.2013