

Wpływ konstrukcji skali interwałowych różniących się kalibracją wartości oraz szerokością przedziałów na odpowiedzi badanych

Magdalena Woronecka, Piotr Zakrzewski, Monika Zielińska

Gdański Uniwersytet Medyczny

Abstrakt

Skala interwałowa pozwala przypisać zmiennym określoną wartość i mierzyć różnicę pomiędzy wartościami tej samej zmiennej. W 1985 roku został przeprowadzony eksperyment, którego sens sprowadzał się do tego, aby sprawdzić, czy odpowiedzi badanych mogą się różnić ze względu na to, jakie warianty przedstawia skala, do której mają się odnieść. Poniżej przedstawiamy replikację badania Schwarza, Hipplera, Deutscha i Stracka, przy użyciu opracowanych przez nas ankiet, biorących pod uwagę dwie zmienne – kalibrację wartości i szerokość przedziałów. Kalibrację podzielono na niską i wysoką, czyli odpowiedzi na skali interwałowej przedstawiono za pomocą niskich lub wysokich wartości. Szerokość przedziałów podzielono na wąską i szeroką, czyli rozpiętość między wartościami na skali może być mniejsza lub większa. Każdy ze stu losowo wybranych uczestników (N=100) wypełnił jedną z czterech ankiet internetowych, które różniły się jedynie konstrukcją skali interwałowej. Wyniki badania mogłyby przyczynić się do zrozumienia, czy konkretna budowa skali ma wpływ na odpowiedzi ludzi. Mogłyby też pomóc opracować metodę, jak ten wpływ ograniczyć do minimum, by odpowiedzi były jak najbliższe prawdy, nie obarczone żadnym efektem manipulacji wywieranym przez konstrukcję skali.

Słowa kluczowe: skala interwałowa, wpływ skali, kalibracja wartości, szerokość przedziałów

1 Wprowadzenie

Podjęmowane często bezrefleksyjnie decyzje dotyczące wyboru skali odpowiedzi mają istotne konsekwencje, ponieważ respondent musi transponować swoją „prywatną” skalę do różnicowania w dostarczoną przez badacza. Ta transpozycja wyznacza jego styl odpowiadania na pytania (Wierzbński, Kuźmińska, & Król, 2014). Rozwój technologii oraz łatwa dostępność różnorodnych elektronicznych kanałów komunikacyjnych sprawiły, że badania ankietowe są jedną z chętniej wykorzystywanych metod pozyskiwania informacji w społeczeństwie informacyjnym. Na odpowiedzi udzielane przez respondentów rzutuje zarówno struktura pytań, ich kolejność, długość, forma, zastosowane skale, jak i użyte słownictwo, styl oraz sformułowania (Krok, 2016). Wcześniejsze badania koncentrowały się głównie na wpływie respondenta na przebieg wywiadu bez uwzględnienia procesów poznawczych leżących u podstaw stosowania kategorii skali. (Schwarz, Hippler, Deutsch, & Strack, 1985)

Schwarz, Hippler, Deutsch i Strack (1985) zbadali wpływ zakresu kategorii skali na odpowiedzi badanych w pytaniach zamkniętych. Respondenci mieli odpowiedzieć na pytanie dotyczące dziennego użytku telewizji w skali, która wahała się od „do pół godziny” do „ponad dwie i pół godziny” lub wahała się od „do dwóch i pół godziny” do „więcej niż cztery i pół godziny”. Pierwsza grupa zgłaszała mniejsze wykorzystanie telewizji niż druga i oszacowali, że średnie wykorzystanie telewizji jest niższe. Wyniki te wskazują, że badani wywnioskowali, iż średnia oglądalność telewizji z alternatywnych odpowiedzi została im podana i zastosowali ją jako standard porównania w ocenie ich zachowania. (Schwarz et al., 1985)

W replikacji badania Schwarza i in. (1985) użyliśmy stworzonego przez nas kwestionariusza w czterech wersjach, które różniły się kombinacją zmiennych niezależnych. Oprócz niskiej i wysokiej kalibracji wartości wprowadziliśmy podział na wąski i szeroki przedział odpowiedzi.

Staraliśmy się ustalić, czy to, w jaki sposób została skonstruowana skala interwałowa – za pomocą mniejszych lub większych wartości oraz wąskich i szerokich przedziałów – ma wpływ na odpowiedzi badanych. Postawiliśmy hipotezę, że odpowiedzi na skalach przedziałowych o niskiej kalibracji będą mniejsze w porównaniu do odpowiedzi na te same pytania o wysokiej kalibracji. Analogicznie odpowiedzi na skali z wąskim przedziałem będą mniejsze od odpowiedzi z szerokim przedziałem.

2 Metody

2.1. Uczestnicy

W badaniu wzięło udział $N=100$ osób w przedziale wiekowym 14-59 lat, w tym 76% stanowiły kobiety. Średnia wieku wynosiła $M=25,9$ lat a odchylenie standardowe $SD=10,11$. 56% osób badanych posiadała wykształcenie średnie, 33% wyższe, 6% gimnazjalne, 3% zawodowe, a 2% podstawowe. Osoby były dobierane w sposób przypadkowy.

2.2 *Procedura*

Eksperyment przeprowadzono przy pomocy ankiety internetowej. Badani odpowiadali na pięć pytań:

1. Ile mniej więcej książek czytasz w ciągu roku?
2. Średnio ile czasu dziennie spędzasz na portalach społecznościowych typu Facebook, Instagram, YouTube itp.?
3. Jak często korzystasz z restauracji typu Fast-food?
4. Mniej więcej ile pieniędzy miesięcznie przeznaczasz na własne przyjemności (np. słodycze, kino, wyjścia ze znajomymi, zakupy)?
5. Ile czasu tygodniowo przeznaczasz na sport/ćwiczenia/aktywność fizyczną?

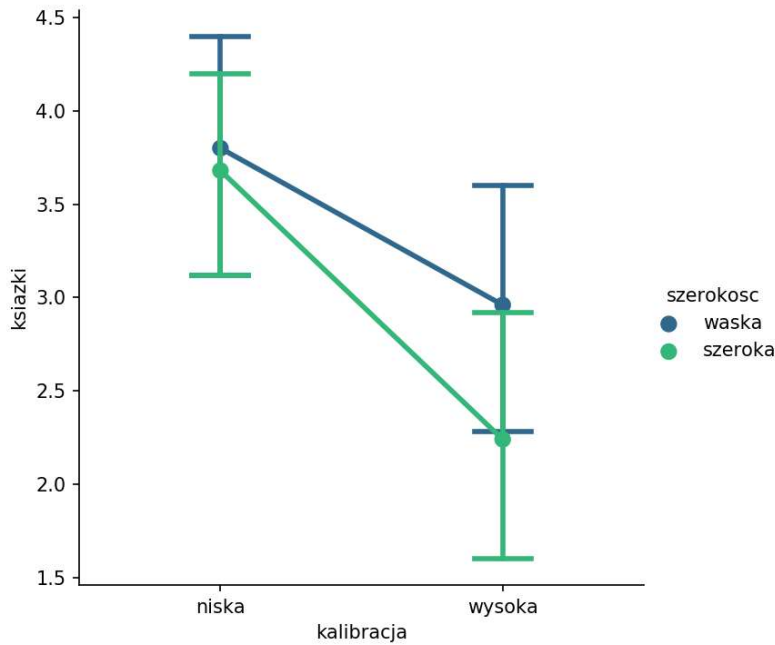
Odpowiedzi udzielali na skali interwałowej w dwóch wariantach szerokości przedziału i w dwóch wariantach kalibracji. Badani losowali jedną z czterech grup i zostali przeniesieni do odpowiedniej ankiety o skali z niską kalibracją i wąskim przedziałem, wysoką kalibracją i wąskim przedziałem, niską kalibracją i szerokim przedziałem lub wysoką kalibracją i szerokim przedziałem.

2.3 *Analiza statystyczna*

3 Wyniki

3.1 *Książki*

Średnia w grupie o niskiej kalibracji i szerokim przedziale wyniosła $M = 3,68$ ($SD = 1,44$). Średnia w grupie o niskiej kalibracji i wąskim przedziale wyniosła $M = 3,80$ ($SD = 1,80$). Średnia w grupie o wysokiej kalibracji i szerokim przedziale wyniosła $M = 2,24$ ($SD = 1,64$). Średnia w grupie o wysokiej kalibracji i wąskim przedziale wyniosła $M = 2,96$ ($SD = 1,72$).

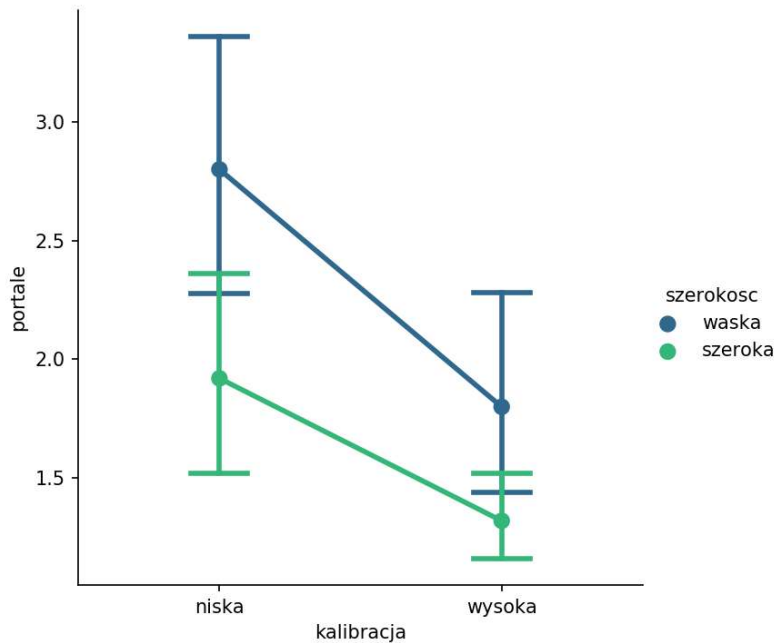


Rys. 1. Odpowiedzi badanych dotyczące książek

Efekt główny kalibracji dla oceny książek okazał się istotny statystycznie; $F(1, 96) = 11,862$, $p < 0,001$. Efekt główny szerokości dla oceny książek okazał się nieistotny statystycznie; $F(1, 96) = 1,610$, $p > 0,05$. Efekt interakcji kalibracji i szerokości okazał się nieistotny statystycznie; $F(1, 96) = 0,821$, $p > 0,05$.

3.2 Portale społecznościowe

Średnia w grupie o niskiej kalibracji i szerokim przedziale wyniosła $M = 1,92$ ($SD = 1,08$). Średnia w grupie o niskiej kalibracji i wąskim przedziale wyniosła $M = 2,80$ ($SD = 1,44$). Średnia w grupie o wysokiej kalibracji i szerokim przedziale wyniosła $M = 1,32$ ($SD = 0,48$). Średnia w grupie o wysokiej kalibracji i wąskim przedziale wyniosła $M = 1,80$ ($SD = 1,12$).

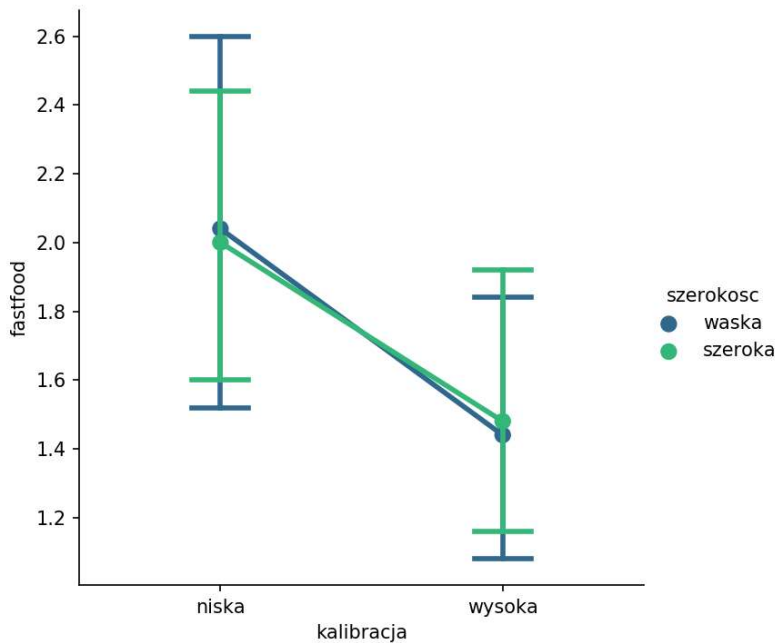


Rys. 2. Odpowiedzi badanych dotyczące portali społecznościowych

Efekt główny kalibracji dla oceny portali społecznościowych okazał się istotny statystycznie; $F(1, 96) = 13,560$, $p < 0,001$. Efekt główny szerokości dla oceny portali społecznościowych okazał się istotny statystycznie; $F(1, 96) = 9,800$, $p < 0,05$. Efekt interakcji kalibracji i szerokości okazał się nieistotny statystycznie; $F(1, 96) = 0,850$, $p > 0,05$.

3.3 *Fast food*

Średnia w grupie o niskiej kalibracji i szerokim przedziale wyniosła $M = 2,00$ ($SD = 1,15$). Średnia w grupie o niskiej kalibracji i wąskim przedziale wyniosła $M = 2,04$ ($SD = 1,43$). Średnia w grupie o wysokiej kalibracji i szerokim przedziale wyniosła $M = 1,48$ ($SD = 1,00$). Średnia w grupie o wysokiej kalibracji i wąskim przedziale wyniosła $M = 1,44$ ($SD = 1,04$).

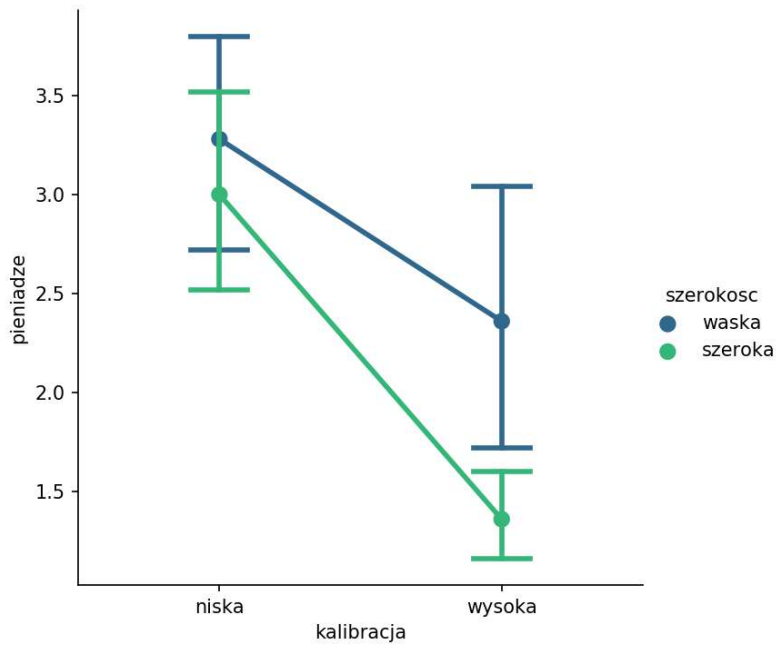


Rys. 3. Odpowiedzi badanych dotyczące restauracji typu fast food

Efekt główny kalibracji dla oceny restauracji typu fast food okazał się istotny statystycznie; $F(1, 96) = 5,730, p < 0,05$. Efekt główny szerokości dla oceny restauracji typu fast food okazał się nieistotny statystycznie; $F(1, 96) = 1,034, p > 0,05$. Efekt interakcji kalibracji i szerokości okazał się nieistotny statystycznie; $F(1, 96) = 2,923, p > 0,05$.

3.4 Pieniądze

Średnia w grupie o niskiej kalibracji i szerokim przedziale wyniosła $M = 3,00$ ($SD = 1,26$). Średnia w grupie o niskiej kalibracji i wąskim przedziale wyniosła $M = 3,28$ ($SD = 1,37$). Średnia w grupie o wysokiej kalibracji i szerokim przedziale wyniosła $M = 1,36$ ($SD = 0,57$). Średnia w grupie o wysokiej kalibracji i wąskim przedziale wyniosła $M = 2,36$ ($SD = 1,66$).

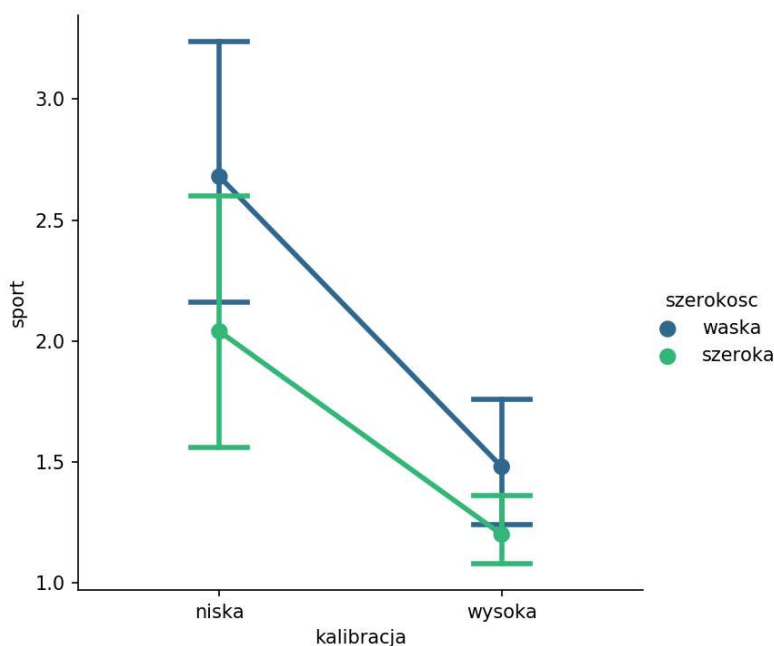


Rys. 4. Odpowiedzi badanych dotyczące pieniędzy

Efekt główny kalibracji dla oceny pieniędzy okazał się istotny statystycznie; $F(1, 96) = 25,120$, $p < 0,001$. Efekt główny szerokości dla oceny pieniędzy okazał się istotny statystycznie; $F(1, 96) = 6,280$, $p < 0,05$. Efekt interakcji kalibracji i szerokości okazał się nieistotny statystycznie; $F(1, 96) = 1,990$, $p > 0,05$.

3.5 Sport

Średnia w grupie o niskiej kalibracji i szerokim przedziale wyniosła $M = 2,04$ ($SD = 1,31$). Średnia w grupie o niskiej kalibracji i wąskim przedziale wyniosła $M = 2,68$ ($SD = 1,46$). Średnia w grupie o wysokiej kalibracji i szerokim przedziale wyniosła $M = 1,20$ ($SD = 0,41$). Średnia w grupie o wysokiej kalibracji i wąskim przedziale wyniosła $M = 1,48$ ($SD = 0,65$).



Rys. 5. Odpowiedzi badanych dotyczące sportu

Efekt główny kalibracji dla oceny sportu okazał się istotny statystycznie; $F(1, 96) = 23,414, p < 0,001$. Efekt główny szerokości dla oceny sportu okazał się istotny statystycznie; $F(1, 96) = 4,762, p < 0,05$. Efekt interakcji kalibracji i szerokości okazał się nieistotny statystycznie; $F(1, 96) = 0,730, p > 0,05$.

4 Dyskusja

Konstrukcja skali to ważne zagadnienie dla badań i testów z wielu różnych dziedzin psychologii. Może ona mieć istotny wpływ na to, co jest dla nas najważniejsze – wyniki. Inaczej zbudowana skala znacząco wpłynęła na odpowiedzi badanych, jednak nie wszystkie nasze założenia były słuszne.

Zgodnie z pierwszą hipotezą, na skali z wartościami o niskiej kalibracji odpowiedzi były niższe od tych przedstawionych za pomocą większych wartości, mimo iż pytania były te same. Efekt główny kalibracji skali okazał się istotny w każdym z pięciu przypadków. Hipoteza odnosząca się do szerokości przedziałów nie została w pełni potwierdzona. W trzech pytaniach efekt główny szerokości przedziałów okazał się istotny statycznie (pytania nr 2, 4 i 5), jednak w dwóch pozostałych nie stwierdzono istotnych różnic. Należy pamiętać, że wartości przypisane do konkretnych pozycji w grupach o niskiej kalibracji są mniejsze od odpowiedzi, które znajdują się na tej samej pozycji w grupach o wysokiej kalibracji wartości. Choć średnia odpowiedzi przedstawionych na powyższych wykresach może wydawać się wyższa dla grup o niskiej kalibracji, to wartości przypisane tym średnim są niższe od tych dla grup o wysokiej kalibracji.

Wyciągając wnioski z uzyskanych przez nas wyników jesteśmy skłonni stwierdzić, że potwierdziła się hipoteza replikowanego przez nas badania. Wyniki są zależne od tego, jak skonstruowaną skalę badani otrzymali. Wyjaśnieniem tego może być fakt, że badani uważają zaprezentowaną im skalę za pewne odzwierciedlenie rzeczywistości, która zawiera informację o normach występujących w populacji. Skala ta jednak jest uprzednim założeniem badacza na temat świata, a nie jego odzwierciedleniem. Badani jednak, nie chcąc odstawać od normy, udzielają odpowiedzi w oparciu o zaprezentowaną im skalę i są znacznie mniej skłonni do podawania prawdziwych wyników, które mogłyby wydawać się nadzwyczajne w kontekście zaprezentowanej im skali.

Podczas kolejnego badania należałoby skupić się efekcie szerokości przedziałów, gdyż obecne wyniki nie pozwalają jednoznacznie potwierdzić lub wykluczyć ich wpływu na odpowiedzi badanych. Przyczyną może być między innymi małe zróżnicowanie względem płci i wykształcenia. Znaczną większość stanowią kobiety, a ponad połowa respondentów posiada wykształcenie średnie. Dalsze badania powinny uwzględnić rozłożenie tych cech w próbce i zrekrutować więcej osób badanych w celu uzyskania bardziej wiarygodnej grupy badawczej.

Warto by było również przyłożyć większą wagę do samej treści pytań. W ankiecie badani dostali pytania takie, w których wyższa odpowiedź jest odbierana pozytywniej niż niższa, jak i na odwrót. Przykładowo osoby, które więcej czasu poświęcają na sport, będą lepiej odbierane niż te, które poświęcają go mniej lub wcale, za to w przypadku częstości chodzenia do restauracji typu fast food sytuacja się odwraca – lepiej chodzić rzadziej niż częściej. Choć ankieta była w pełni anonimowa, respondenci mogli odczuwać pewien wstyd lub dyskomfort psychiczny przed zaznaczeniem prawdziwych odpowiedzi, które mogły być wyższe lub niższe od tych przychylniej odbieranych przez społeczeństwo.

Wyniki zarówno naszego, jak i oryginalnego badania Schwarza i in. (1985) wskazują, że badacze powinni przykładać szczególną uwagę do tego, w jaki sposób konstruują skale wykorzystywane w ich badaniach. Źle dobrane wartości, za pomocą których przedstawiane są odpowiedzi, mogą całkowicie zmienić wyniki tych badań, co skutkuje chociażby fałszywym obrazem populacji. Być może nie da się w pełni zapobiec temu efektowi, jaki konstrukcja skali wywiera na człowieka, ale warto zdawać sobie z tego sprawę i mieć to na uwadze podczas przeprowadzania badań jak i interpretacji ich wyników.

5 Bibliografia

Krok, E. (2016). Survey questionnaire structure and its impact on the research results. *Zeszyty Naukowe Uniwersytetu Szczecińskiego. Studia Informatica*, 37(874), 55–73.
<https://doi.org/10.18276/si.2015.37-05>

- Schwarz, N., Hippler, H.-J., Deutsch, B., & Strack, F. (1985). Response Scales: Effects of Category Range on Reported Behavior and Comparative Judgments. *Public Opinion Quarterly*, 49(3), 388. <https://doi.org/10.1086/268936>
- Wierzbński, J., Kuźmińska, A. O., & Król, G. (2014). Consequences of using different types of rating scales. *Problemy Zarządzania*, 12(1), 113–136. <https://doi.org/10.7172/1644-9584.45.8>