Open Peer Review on Qeios

RESEARCH ARTICLE

Detection and Correction of Likert Scale Multiplicative Response-Style Bias

Reinhard Oldenburg[1]

1 Universität Augsburg, Germany

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

Abstract

Individuals may differ in their tendency to fully use the range of a Likert scale. Such individual response styles may have an effect on estimates of correlations. It is shown with simulated data that a relatively simple nonlinear latent model can be used to detect and correct (to some extent) for the bias introduced by differences in multiplicative response styles. A real-world dataset is examined to show that the effect may occur in reality and may lead to biased conclusions.

Corresponding author: Reinhard Oldenburg, reinhard.oldenburg@math.uni-augsburg.de

Introduction

The use of Likert scales[1] to measure latent constructs in the social sciences is extremely widespread[2]. Some debates have investigated which uses of Likert scales are statistically valid given that they are ordinal, and thus parametric statistical methods may not be adequate (see, e.g., Carifio and Perla[3] and Norman[4] for a discussion). Another strand of critical discussion is concerned with the bias that Likert scales may introduce. James et al.[5] described the bias of central tendency, i.e., the preference for the neutral position. Chung and Monroe[6] discuss a bias that comes from social desirability. Kreitchmann et al.[7] compare graded-scale items and forced-choice format items to assess bias. Pimentel[8] explored bias related to the question of odd or even Likert scales. It has been shown that response styles[9] depend on culture[10] and that they affect applied areas such as marketing research[11]. Similarly, Liu et al.[12] investigate extreme response styles and conclude that this can have significant implications, but they do not investigate ways to correct this bias. Douven[13] gives an explanation for the central tendency in Likert scale studies by cognitive modeling of the decision process. All these studies (and many more) describe the origin of the bias and give recommendations on how to avoid or reduce it, e.g., by choosing prompts with care and designing the questionnaire appropriately. Detection and correction of bias with Likert scales are studied as well. Van de Vries[14] discusses a wide spectrum of methods that are mainly targeted at the detection of bias, less on the correction. Among the studies that discuss methods to correct bias introduced from response styles is Moors[15]. It uses a latent class analysis to check and correct response styles when they are of discrete nature. Somewhat similarly, but simpler in its approach, is Greenleaf[16]. He uses individuals' variances over many responses to weight responses in a linear regression model. In contrast to Moors, the model I will investigate in this paper has no discrete classes of response styles but measures individuals' response types approximately on a continuum. And in contrast to Greenleaf, the model considered here uses the flexibility of latent variables to estimate the individual bias. In Rossi et al.[17], the general idea is very similar to our approach: An additive and a multiplicative bias are treated as a latent variable for each case. However, the method estimates a covariance matrix between Likert scale items, not between constructs built from these scales. Böckenholt[18] uses IRT methods based on the assumption that the underlying "true" construct to be measured is normally distributed. He models the individuals' answering process by binary trees. This approach may give further insight, but it is conceptually more difficult than the model applied in the present paper. Furthermore, several advanced IRT methods are described in the literature that model the probability of selecting a grade at a Likert scale[19][20][21]. Such methods are very flexible and can account for different forms of response styles. However, their complexity makes it difficult for the user to understand and adapt the response style model to the situation at hand. The present study is directed at detecting and correcting a multiplicative form of response style bias by using a simple nonlinear confirmatory factor analysis model. The kind of response style considered is similar to the extreme response style in the classification of van Vaerenbergh and Thomas[9].

The research questions that will be answered are:

- Can individual multiplicative response style have a substantial impact on estimations of correlations between constructs?
- Can nonlinear latent variable models be used to detect and correct such a bias at least to a certain extent?
- Is there evidence that this kind of bias has an impact on the analysis of real-world data?

The rest of the article will first describe the model and the estimation method. This is followed by a description of the simulation and its results. Finally, a real-world data set will be analyzed.

This article will follow the notational conventions suggested by Carifio and Perla[3] that state that a Likert scale is a combination of Likert items. An individual Likert item is a test item that allows respondents to express e.g. agreement with a given statement by choosing one of several (often five or seven) options displayed linearly and either numbered or described verbally.

The model

The Likert scales that are considered here typically are designed to measure a latent construct by several Likert items that ask respondents to rate their agreement with a given statement. So there is a natural zero (no agreement) and possible values of agreement are in some interval $[0, c]$. Respondents typically have a choice of five or seven grades, but in some realizations, also more fine-grained answers are possible (especially in online surveys, it is often possible to select a percentage by using a pointing device). In order to have a unified presentation, it is assumed that all these responses to Likert items are linearly transformed to the unit interval [0,1]. We will consider responses of n individuals on l Likert scales, each consisting of m Likert items. Thus, the raw data to be simulated and analyzed consists of the following data:

$$x_{i,k}^{(j)} \in [0,1],\ i = 1..n,\ j = 1..l, k = 1..m$$

The latent structure to be recovered is defined by a matrix of “true” values that each individual has for each construct:

$$\xi_i^{(j)} \in [0,1],\ i = 1..n,\ j = 1..l$$

If there was no multiplicative and additive response style transformation, then the reflexive congeneric measurement model (e.g., Kline[22]) would be simply

$$x_{i,k}^{(j)} = \lambda_k^{(j)} \xi_i^{(j)} + \epsilon_{i,k}^{(j)}$$

Here $\epsilon_{i,k}^{(j)}$ is an error variable such that the $\epsilon_{\cdot,k}^{(j)}$ have mean 0 and are uncorrelated. Widely used is also the special case of a τ-equivalent measurement model where all $\lambda_k^{(j)} = 1$.

In the presence of a multiplicative response style, it is assumed that some individuals tend to express a judgement by giving larger ratings than others, despite their true values should be equal if they could be measured. It is assumed that this is an individual trait of expressiveness, or an individual factor that multiplies the “true” value when answering the Likert scale. The measurement model that will be discussed in this paper is, in the general case, congeneric with intercepts $\beta_k^{(j)}$ and loadings $\lambda_k^{(j)}$:

$$x_{i,k}^{(j)} = \mu_i \lambda_k^{(j)} \xi_i^{(j)} + \beta_k^{(j)} + \epsilon_{i,k}^{(j)}$$

Note that this model is not identified as there is a “gauge freedom” because one could multiply the $\xi_i^{(j)}$ for all j by the same factor if μ_i or $\lambda_k^{(j)}$ is divided by this factor simultaneously. To solve this problem, one can either fix some loadings or restrict them to certain intervals. This will not fix them uniquely, but here we are only interested in correlations between latent variables, and these are not affected by such linear transformations. Therefore, the response style variables are restricted to $\mu_i \in [1-\delta_\mu, 1+\delta_\mu]$ with a fixed number $\delta_\mu > 0$ that is expected to be large enough to capture the expected range of response style factors (in the simulations below, $\delta_\mu = 0.5$ will be used). Moreover, the latent variables will be constrained to an interval, too: $\xi_i^{(j)} \in [0,1]$. For the loadings, it suffices to restrict them to some sensible interval, e.g., $\lambda_k^{(j)} \in [-10, 10]$, $\beta_k^{(j)} \in [-1, 1]$. All this guarantees that the solution will be contained in some compact space, and the remaining freedom within does not affect correlations calculated from estimates of latent variables.

The model parameters are estimated by weighted least squares, i.e.,

$$\underset{\{\mu_i, \lambda_k^{(j)}, \xi_i^{(j)}, \beta_k^{(j)}\}}{argmin} \sum_{j=1}^{l} \sum_{k=1}^{m} w_k^{(j)} \cdot \sum_{i=1}^{n} \epsilon_{i,k}^{(j)2}$$

With weights $w_k^{(j)} > 0$. Initially, uniform weighted least squares (ULS) can be used with $w_k^{(j)} = 1$. Then, some iterations of weight selection can be appended to get a weighted least squares (WLS) approximation by using $w_k^{(j)} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\epsilon_{i,k}^{(j)2}}$ from the estimates of the foregoing iteration step as an estimate for the inverse variance of $\epsilon_k^{(j)}$.

Simulation studies

Simulations of data for the class of models described in the preceding section are straightforward. As input, one needs the sample size n, the number of constructs, and the number of Likert items per Likert scale. Moreover, a small number δ' that defines the maximal strength of the multiplicative bias, and of course the distributions of latent variables and error variances.

In the simulation reported here, four latent variables have been chosen to be multivariate normal with mean $(0.5, 0.5, 0.5, 0.5)$ and covariance

$$\sigma^2 = \begin{pmatrix} 0.02 & 0.004 & 0.001 & 0.006 \\ & 0.015 & 0.001 & 0.004 \\ & & 0.004 & 0.005 \\ & & & 0.01 \end{pmatrix}.$$

For each of the four Likert scales, 6 Likert scales have been generated by adding errors and rounding to Likert scale values. Error variables are chosen to be independent with a mean of 0 and are normally distributed with standard deviations between 0.03 and 0.1. All weights $\lambda_k^{(j)}$ have been chosen to be 1. The multiplicative factor is chosen to be uniformly distributed $\mu \sim U\left(1 - \delta', 1 + \delta'\right), \delta' = 0.5$. From these data, the simulated responses have been calculated and rounded to the s values of the Likert items ($s = 7$).

From a data sample of size n, one can calculate several estimates of the correlation matrix R. The easiest one is the correlation matrix over sum score scales $x^{(j)} := \sum_{k=1}^{m} x_{.,k}^{(j)} \in \mathrm{R}^n$, $\hat{R}_{sum} := cor\left(x^{(j)}, x^{\left(j'\right)}\right)_{j,j'}$. The second approach is to estimate model (1) with a standard SEM package like lavaan[23].

To estimate the non-linear model (2), several techniques can be applied. I found that in this application, least square estimation outperforms Bayesian estimation, so it was applied.

Initial values for parameters for latent variables were chosen as the first observed variable in its measurement model. Full details are given in the implementation (R code is in the attached files).

A total of 100 samples for sample size $n = 300$ were generated and estimated. Estimations were carried out on a MacBook (M1Pro, 32GB RAM) with R4.4, lavaan 0.6.18, and optimx 2023-10.21. Estimating one sample of size 300 took about one minute for all four methods together. All files are available from https://myweb.rz.uni-augsburg.de/~oldenbre/LikertCorrectionFiles.zip.

The root mean square error (RMSE) is an established measure for the deviations between estimates and true values. It was calculated independently for the 10 off-diagonal entries in the correlation matrix, and then the average of these was calculated and displayed in Table 1.

Table 1. Simulation results

	Mean absolute error				RMSE			
	Sum	Sem	ULS	WLS	Sum	Sem	ULS	WLS
***n* = 100**	0.31	0.34	0.26	0.23	0.31	0.34	0.28	0.25
***n* = 300**	0.31	0.34	0.21	0.25	0.31	0.34	0.22	0.26

The results show that ULS and WLS typically get better estimates than the traditional methods. However, the resulting RMSE are still far from perfect. A closer look shows that the traditional methods tend to overestimate the correlation, while ULS and WLS tend to underestimate the correlations, i.e., they overcompensate the bias by attributing too much common factor to the bias. One can bring down this effect by specifying a smaller value for δ_μ, but as there is no obvious strategy to choose a best value for this parameter, this optimization was not done here (and neither in the real-world application below).

Taken together, these results allow us to give affirmative answers to research questions 1 and 2.

Real-world data

A systematic study of the presence of a multiplicative response bias in data collected in research in the human sciences is, of course, beyond the scope of this publication. However, several data sets have been investigated, and all showed more or less strong bias. A typical example will be reported in detail. The data[24] is publicly available and measures three constructs, namely fear $\xi^{(1)}$ (with 7 five-point Likert items), dependence $\xi^{(2)}$ (5 Likert items), and anxiety $\xi^{(3)}$ (7 Likert items) of 1317 individuals.

In a first analysis, only the 370 cases without missing data have been kept, i.e., the data is a matrix $x_{ik},\ i = 1..n = 370,\ k = 1..19$. The model was chosen as congeneric, as no reason for τ-equivalence was given:

$$x_{ik} = a_k \cdot \mu_i \cdot \xi_i^{(j(k))} + b_k + \epsilon_{ik}$$

Here $j(k)$ is 1 for $1 \le k \le 7,\ 2$ for $8 \le k \le 12$ and 3 for $k \ge 13$. The constraints have been set to $\xi_i^{(j)} \in [0, 1],\ \mu_i \in [0.5, 1.5], |a_k| \le 2, -0.1 \le b_k \le 1$.

The correlation matrices of the constructs have been calculated from sum scores, SEM (using lavaan with ML estimation), and the above-given model estimated with ULS and WLS least squares. The results are given in this order in the following four correlation matrices.

$$\textit{Sum scores}: \begin{pmatrix} 1 & 0.104 & 0.403 \\ & 1 & -0.011 \\ & & 1 \end{pmatrix}, \textit{SEM}: \begin{pmatrix} 1 & 0.130 & 0.480 \\ & 1 & -0.017 \\ & & 1 \end{pmatrix}$$

$$\textit{ULS}: \begin{pmatrix} 1 & 0.079 & 0.373 \\ & 1 & -0.047 \\ & & 1 \end{pmatrix}, \textit{WLS}: \begin{pmatrix} 1 & 0.084 & 0.299 \\ & 1 & -0.042 \\ & & 1 \end{pmatrix}$$

The differences between the largest and smallest estimates of the same correlation are up to 0.18, and therefore, they cannot be ignored.

In a second analysis, the full set of size 1317 has been analyzed. For sum score correlation, pairwise complete observations were used. For the SEM estimation, the lavaan option missing = "fiml" was used (full information maximum likelihood). In the least squares estimation, summands with missing data were omitted. Results are:

$$\textit{Sum scores}: \begin{pmatrix} 1 & -0.108 & 0.458 \\ & 1 & -0.163 \\ & & 1 \end{pmatrix}, \textit{SEM}: \begin{pmatrix} 1 & -0.142 & 0.602 \\ & 1 & -0.226 \\ & & 1 \end{pmatrix}$$

$$\textit{ULS}: \begin{pmatrix} 1 & -0.155 & 0.393 \\ & 1 & -0.161 \\ & & 1 \end{pmatrix}, \textit{WLS}: \begin{pmatrix} 1 & -0.159 & 0.303 \\ & 1 & -0.201 \\ & & 1 \end{pmatrix}$$

In this case, the difference between the largest and smallest estimates of the same correlation is up to 0.3. However, for the correlations involving the second construct, estimates are much closer together. The reason for this is unclear but may be related to the orientation of the scale. It should be noted that lavaan reported no errors in the estimation and fit indices are good (CFI=0.933, TLI=0.923, RMSEA=0.048, SRMR=0.044).

Summarizing, one can affirmatively answer the third research question.

Conclusions

The simulation study indicated that multiplicative response style bias can affect estimated correlations and that this effect can be reduced (although not completely) by estimating a non-linear model with non-normal latent variables that takes the multiplicative effect into account. Data from real-world studies suggests that the effect can occur in reality and may lead to biased conclusions, maybe even to false ones.

From these results, a clear recommendation can be deduced: One should estimate correlations of latent variables both by traditional means such as sum scores and SEM models, but also by non-linear models as explained above. If the estimated correlations do not differ too much, one may conclude that no substantial multiplicative response bias effect affects the estimation. On the other hand, if they differ, then one should be cautious and interpret the data preferably with the correlations estimated by the non-linear model.

References

1. ^*Likert R (1932). "A technique for the measurement of attitudes". Archives of Psychology. 22 (140): 1-55.*
2. ^*DeVellis RF (2016). Scale Development: Theory and Applications. (4th ed.). Sage Publications.*
3. a, b*Carifio J, Perla R (2007). "Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their anti-dotes". Journal of Social Sciences. 3 (3): 106–116. doi:10.3844/jssp.2007.106.116.*
4. ^*Norman G (2010). "Likert scales, levels of measurement and the “laws” of statistics". Adv in Health Sci Educ. 15: 625–632. doi:10.1007/s10459-010-9222-y.*
5. ^*James LR, Demaree R, Wolf G (1984). "Estimating within-group interrater reliability with and without response bias". Journal of Applied Psychology. 69 (1): 85–98.*
6. ^*Chung J, Monroe GS (2003). "Exploring social desirability bias". Journal of Business Ethics. 44 (4): 291–302.*
7. ^*Kreitchmann RS, Abad FJ, Ponsoda V, Nieto MD, Morillo D (2019). "Controlling for Response Biases in Self-Report Scales: Forced-Choice vs. Psychometric Modeling of Likert Items". Front. Psychol. 10: 2309. doi:10.3389/fpsyg.2019.02309.*
8. ^*Pimentel J (2019). "Some Biases in Likert Scaling Usage and its Correction". International Journal of Sciences: Basic and Applied Research (IJSBAR). 45 (1): 183–191.*
9. a, b*Van Vaerenbergh Y, Thomas TD (2013). "Response styles in survey research: a literature review of antecedents, consequences, and remedies". Int. J. Public Opin. Res. 25: 195–217. doi:10.1093/ijpor/eds021.*
10. ^*Johnson T, Kulesa P, Cho YI, Shavitt S (2005). "The Relation Between Culture and Response Styles: Evidence From 19 Countries". Journal of Cross-Cultural Psychology. 36 (2): 264-277. doi:10.1177/0022022104272905.*
11. ^*Baumgartner H, Steenkamp J-BEM (2001). "Response Styles in Marketing Research: A Cross-National Investigation". Journal of Marketing Research. 38 (2): 143-156. doi:10.1509/jmkr.38.2.143.18840.*
12. ^*Liu M, Harbaugh AG, Harring JR, Hancock G, Gregory R (2017). "The Effect of Extreme Response and Non-extreme Response Styles on Testing Measurement Invariance". Frontiers in Psychology. 8. doi:10.3389/fpsyg.2017.00726.*
13. ^*Douven I (2018). "A Bayesian perspective on Likert scales and central tendency". Psychonomic Bulletin & Review. 25 (3): 1203–1211. doi:10.3758/s13423-017-1344-2.*
14. ^*van de Vijver FJR (2018). "Capturing Bias in Structural Equation Modeling". In: Eldad Davidov, Peter Schmidt, Jaak Billiet, Bart Meuleman (Eds). Cross-Cultural Analysis. Routledge.*
15. ^*Moors G (2003). "Diagnosing response style behavior by means of a latent-class factor approach. Socio-demographic correlates of gender role attitudes and perceptions of ethnic discrimination reexamined". Qual. Quant. 37: 277–302. doi:10.1023/A:1024472110002.*
16. ^*Greenleaf EA (1992). "Improving Rating Scale Measures by Detecting and Correcting Bias Components in Some Response Styles". Journal of Marketing Research. 29 (2): 176-188.*

17. ^*Rossi PE, Gilula Z, Allenby GM (2001). "Overcoming scale usage heterogeneity: A Bayesian hierarchical approach". Journal of the American Statistical Association. 96 (453): 20-31.*
18. ^*Böckenholt U (2016). "Measuring Response Styles in Likert Items". Psychological Methods. 22 (1): 69–83. doi:10.1037/met0000106.*
19. ^*Falk CF, Cai L (2016). "A flexible full-information approach to the modeling of response styles". Psychological Methods. 21 (3): 328.*
20. ^*Henninger M, Meiser T (2020a). "Different approaches to modeling response styles in Divide-by-Total Item Response Theory models (Part I): A model integration". Psychological Methods. 25: 560-576. doi:10.1037/met0000249.*
21. ^*Henninger M, Meiser T (2020b). "Different approaches to modeling response styles in Divide-by-Total Item Response Theory models (Part II): Applications and novel extensions". Psychological Methods. 25: 577-595. doi:10.1037/met0000268.*
22. ^*Kline RB (2015). Principles and Practice of Structural Equation Modeling. 4th ed. New York: Guilford Press.*
23. ^*Rosseel Y (2012). "lavaan: An R Package for Structural Equation Modeling". Journal of Statistical Software. 48 (2): 1-36. http://www.jstatsoft.org/v48/i02/.*
24. ^*Jahan N, Hossain M, Ahammed B, Shohel TA, Shovo T, Khan B, Islam MN (2020). "A psychometric evaluation of fear of COVID-19 scale (FCV-19S) in 4-point Likert-scale: Its reliability and validity". https://doi.org/10.7910/DVN/ABNXB5, Harvard Dataverse, V1.*