What Is The Impact On Scale Reliability And Exploratory Factor Analysis Of A Pearson Correlation Matrix When Some Respondents Are Not Able To Follow The Rating Scale?

Michaela N. Gelin*

T. Mark. Beasley**

Bruno D. Zumbo*

* University of British Columbia, ** University Alabama at Birmingham

In B.D. Zumbo & C.O.Ochieng (Chairs), *The impact of Likert or rating scale mis-responding on commonly used statistical methods*. Symposium conducted at the 2003 annual meeting of the American Educational Research Association (AERA) in Chicago, Illinois

DRAFT

Contact Information:

Professor Bruno D. Zumbo
The University of British Columbia
Scarfe Building, 2125 Main Mall
Department of ECPS
(Program Area: Measurement, Evaluation, & Research Methodology)
Vancouver, B.C.
CANADA V6T 1Z4

e-mail: bruno.zumbo@ubc.ca

Summary

A computer simulation was designed to investigate the impact on exploratory factor analysis (EFA) of a Pearson matrix and the coefficient alpha estimate of scale reliability when some respondents are not able to follow the Likert-type rating scale (e.g., misresponding). In this study, misresponding was simulated by collapsing the upper two scale points for a given Likert scale, thereby reflecting the case wherein respondents are unable to discriminate between the top two Likert options (i.e., misresponders). Moreover, three factors were manipulated in the design: (1) the levels of Likert-type categorization (i.e., number of scale points), ranging from 3 to 9 points, (2) the percent of misresponders (10, 20, and 30), and (3) the percent of items misresponded to (10, 20, 30, 100). Latent responses for 100 000 simulees were generated and transformed to correspond to the various conditions in the simulation study. The results revealed a quadratic effect of the number of scale points with a diminishing effect after 4 scale points when a Likert-type scale is imposed on continuous data. Moreover, this effect held for coefficient alpha and the ratio of the first two eigenvalues. In addition, when a one-factor model is forced onto the data, the interpretation of that factor solution (via the salient loadings), is not affected by the number of scale points. Lastly, the type of misresponding we investigated, for either the proportion of misresponders or proportion of items misresponded to, had little to no effect on the estimated coefficient alpha, the decision on the number of factors to extract (i.e., retain), and the factor interpretations when one has items with 4 or more rating scale points.

What is the impact on scale reliability and exploratory factor analysis of a Pearson correlation matrix when some respondents are not able to follow the rating scale?

Given that rating scale response formats (Likert scales) are widely used in the social sciences to measure unobserved continuous variables, we need to pay attention to these measurement properties when analyzing social science data. Debate has existed in the literature regarding the statistical effects of the number of Likert-type categories (i.e., categorical data) on continuous concepts or variables (i.e., continuous data). For example, coarse categorization (e.g., small number of Likert-type categories) was previously found to result in large standard errors of Pearson's correlation (Bollen & Barb, 1981) and spurious factors in factor analysis and confirmation factor analysis (Green, Akey, Fleming, Hershberger, & Marquis, 1997; Johnson & Creech, 1983), whereas refined categorization (e.g., large number of Likert-type categories) was found to result in low standard errors in the correlation estimates (Green et al., 1997) and fewer spurious factors (Bollen & Barb, 1981).

Likewise, other studies were conducted on the effect of using rating scale items on the reliability of the scores from a test or measure (e.g., Bandalos & Enders, 1996; Birkett, 1989; Chang, 1994; Matell & Jacoby, 1971). Although there were some mixed and conflicting results, it was found that the resulting reliability coefficient was deemed to be independent of the number of rating scale points and that the reliability measures did not improve with the increased refinement of the rating scale points.

In addition, several studies have been conducted on the effect of using rating scales on identification of factors and components in factor analysis, confirmatory maximum likelihood factor analysis, and principal component analysis of Pearson correlation matrices (e.g., Babakus, Ferguson, & Joreskog, 1987; Bernstein & Teng, 1989; DiStefano, 2002; Green, Akey, Fleming, Hershberger, & Marquis, 1997; Muthen & Kaplan, 1985). Results from these studies indicate that rating scale data leads to misrepresentation of the underlying factors and wrong identification of the dimensionality of the latent variables in confirmatory factor analysis.

It should be noted that the chi-square statistic, eigenvalues, and Cronbach's alpha are all summary measures. For example, the chi-square statistic and the eigenvalues can be seen as sums of squares. Due to the Central Limit Theorem, these type of statistics tend to be rather "resistant" to deviations from expected values and to violations of assumptions (see, for example, Beasley, 1992; Franklin, S. B. , Gibson, D. J., Robertson, P.A., Pohlmann, J. T. & Fralish, J. S., 1995).

Surprisingly, the majority of research has focused on confirmatory factor analysis and little to no work has investigated the statistical effects of rating scales (e.g., categorical data) on exploratory factor analysis. Examining these effects is important because it is quite common in the social sciences for investigators to report the results of an exploratory factor analysis (EFA) and then the reliability estimate of their tests or measures.

Moreover, to our knowledge, no work has gone into studying the effect of what we refer to as “misresponding” (see the lead paper in this symposium by Zumbo & Ochieng, 2003) on exploratory factor analysis. Whereas previous research to date has focused on the case wherein all respondents are using the same rating scale and the same thresholds, this study will focus on sensitivity to the violation of the same-scale and same-threshold assumption. That is, this study manipulates, via simulation, (1) the levels of Likert-type categorization, (2) the proportion of misresponders, and (3) the proportion of items responded to differentially.

The objective of this simulation is to compare the statistics produced by analyzing rating scale data to the same statistics that one would have obtained, with the same data, had they been able to conduct the statistical methodology using the continuous latent variable rather than rating scale responses. The statistics of interest were the factor loadings, eigenvalues, and resulting chi-square fit statistic from an EFA of a Pearson matrix, as well as the coefficient alpha estimate of scale reliability.

Methodology

The general methodology for this study was adopted from the introductory paper by Ochieng & Zumbo (2003) and is the same as the methodology in Rupp, Koh, & Zumbo's (2003) paper in this symposium. However, unlike the Rupp et al. study, we are using the Pearson correlation matrix throughout this paper.

Procedure

Latent responses for 100 000 simulees were generated from a well-fitting one-factor model with 10 variables that was based on the kinds of population factor loadings typically found in the social sciences (see Table 1). That is, the factor loadings range from moderate to high. From the factor loadings, a covariance matrix was created, followed by the corresponding correlation matrix, Table 2. Accordingly, these normally distributed continuous scores represent the (typically unobserved) latent scores from which the order responses were simulated. The resulting dataset served as the population from which the observed response conditions were generated. In this study, the continuous scores (which represent the unobserved latent variable) were manipulated to mimic responses on a rating scale. In other words, the simulation methodology mimics the *process* of responding to a rating scale format and then uses the responses as variables in the analyses.

Table 1. Factor loadings used for generating continuous data

Factor loadings
0.804
0.678
0.671
0.640
0.628
0.568
0.560
0.458
0.437
0.423

Table 2. Theoretical Population Correlation Matrix

Correlation Matrix									
1.000									
0.545	1.000								
0.539	0.455	1.000							
0.515	0.434	0.429	1.000						
0.505	0.426	0.421	0.402	1.000					
0.457	0.385	0.381	0.364	0.357	1.000				
0.450	0.380	0.376	0.358	0.352	0.318	1.000			
0.368	0.311	0.307	0.293	0.288	0.260	0.256	1.000		
0.351	0.296	0.293	0.280	0.274	0.248	0.245	0.200	1.000	
0.340	0.287	0.284	0.271	0.266	0.240	0.237	0.194	0.185	1.000

Study Design

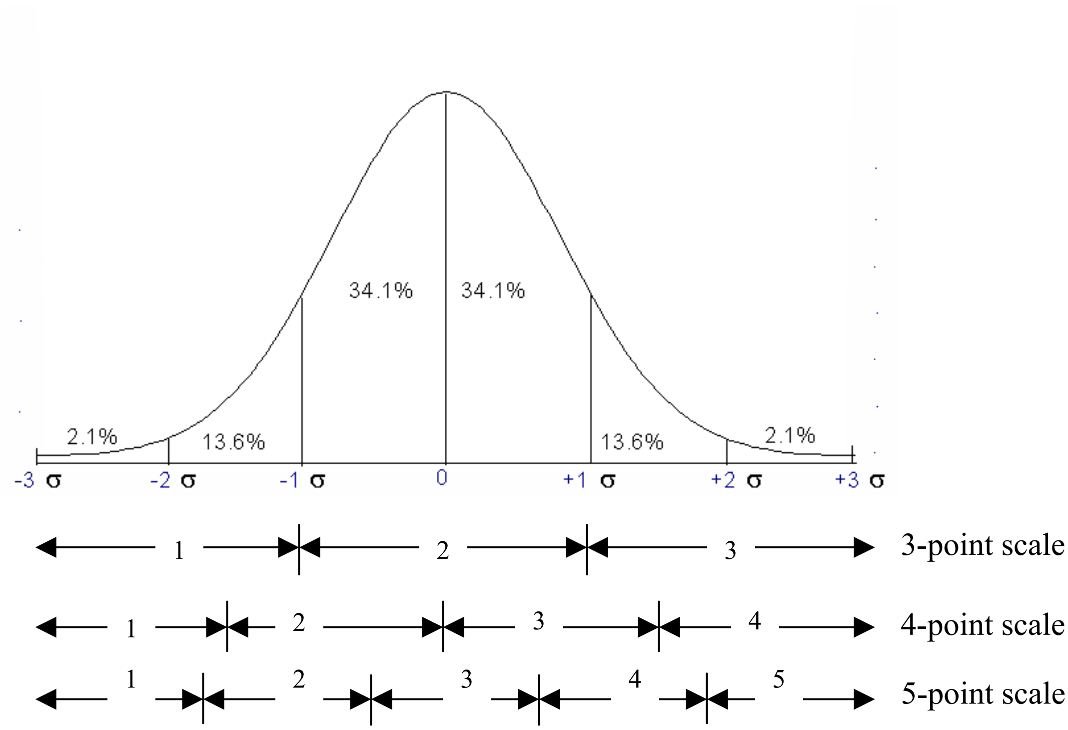
In this study, the levels of Likert-type categorization (e.g., number of scale points), ranging from three to nine points, and the response distribution of the rating scale variables were manipulated from the 100 000 continuous normally (mean=0, sd=1) distributed generated scores. This range of categories was chosen based on the typically scales commonly encountered in the social sciences. In addition, the proportion of respondents not able to use the rating scale as it was intended (i.e., misresponders) and hence responding differentially was varied in the simulation from ten, twenty, thirty, and one-hundred percent of the respondents. Misresponding was simulated in the case where respondents are not able to discern scale points at the top end of the rating scale. For example, this would occur if some respondents used *strongly agree* with *agree* throughout the instrument, thereby responding differentially. Within this type of misresponding the proportion of scale points at the top end of the rating scale that respondents were unable to discern varied from ten, twenty, and thirty. Consequently, some proportion of respondents were simulated to use the top one, two, or three scale points indiscriminately and randomly choose between them.

In essence, the independent variables were: proportion of respondents misresponding (i.e., misresponders), proportion of items misresponded to, and number of scale points. Therefore this is a 3x4x7 design, respectively. In addition, there are seven cells with no misresponding resulting in a total of 91 cells in the design. The dependent variables were the: factor loadings, eigenvalues, chi-square goodness-of-fit statistic and corresponding *p*-values of an EFA of a Pearson matrix, and the coefficient alpha estimate of scale reliability. All comparisons and analyses were made at the population analogue level in order to avoid sample-to-sample variability and focus on large-sample impact (a form of bias) (Zumbo & Ochieng, 2002; Ochieng, 2001).

Response pattern

The response distribution, also referred to as response pattern, was simulated with equal interval scale points resulting in a symmetric distribution of responses similar to that used by Bollen and Barb (1981) in their study of ordinal variables and the Pearson correlation. Accordingly, responses were assumed to be normally distributed across a standardized scale of z=-3 to z=3, and scale points were equally divided for each ordinal item response process in which the Likert scale points were simulated. Accordingly, for a given variable x with m categories, there are $m-1$ unknown thresholds. Figure 1 depicts the thresholds for a three, four, and five point Likert-type scale, followed by a table illustrating thresholds the Likert-type scales with three through nine scale points.

Figure 1. Construction of equal interval Likert-type responses on the standard normal distribution



Scale Points	Equal Interval Thresholds							
3	-1.0000	1.0000						
4	-1.5000	0.0000	1.5000					
5	-1.8000	-0.6000	0.6000	1.8000				
6	-2.0000	-1.0000	0.0000	1.0000	2.0000			
7	-2.1429	-1.2857	-0.4286	0.4286	1.2857	2.1429		
8	-2.2500	-1.5000	-0.7500	0.0000	0.7500	1.5000	2.2500	
9	-2.3333	-1.6667	-1.0000	-0.3333	0.3333	1.0000	1.6667	2.3333

Results

Factor Analysis

(i) *The number of factors (i.e., dimensionality of the item space)*

In an EFA, the first question one is confronted with is how many factors to retain (i.e., extract) in the factor analysis. In this study we focused on two common indices for deciding on the number of factors. The first index is a ratio of the first to the second eigenvalues that arise from a principal components analysis (PCA) of the correlation matrix. These are the initial eigenvalues before any rotation. Please note that although we used maximum likelihood EFA, a PCA is reported in the first steps of that analysis. For this index, two commonly used criterions for the ratio of the first two eigenvalues are a ratio of at least 3.0 or 4.0 to suggest a unidimensional factor structure. The second index is the chi-squared fit statistic that accompanies the EFA, in which a non-significant chi-square value indicates a good fit. We recorded the chi-squared value (and accompanying *p*-value) for the fit of a one-factor model to the correlation matrix.

For the continuous (typically unobserved) case the results were (recall that we generated data that fit a one-factor model) as expected:

Ratio of 1^{st} to 2^{nd} eigenvalues = 5.15

Chi-squared fit statistic for one-factor solution = 36.98, *df*=35, *p*=.377

(a) *The ratio of the first two eigenvalues*

- We can see from Table 3 and Figure 2 that the ratio of the first two eigenvalues suggest that if we were to use a criterion of three for the ratio, a one-factor model holds for all of the Likert-type categorizations and this model holds irrespective of the misresponding.

However, if a ratio of four were used, the 3-point Likert-type scale would result in the researcher rejecting a one-factor model.

- Finally, as the number of scale points increases, the ratio of the first two eigenvalues approaches the continuous case with a diminishing return beyond four scale points. Likewise, this quadratic relationship of the number of scale points is not affected by misresponding as shown in Figures 3 to 5.
- The above findings were supported in the statistical modeling of the simulation outcomes. We fit a general linear (regression) model with the additive effects of the three design factors (treating them as continuous variables) including a quadratic term for the number of scale points. The results only showed a statistically significant quadratic effect of the number scale points: model R-squared of 0.980, linear effect $F(1,86)=1065.7$ $p<.0001$, quadratic effect of $F(1,86)=587.7$, $p<.0001$, all other effects were statistically non-significant.

Table 3. Ratios of first two eigenvalues for a one-factor model by the number of scale points.

Proportion of respondents who are misresponders	**Proportion of items on which they misrespond**	**Number of scale points**						
		3	**4**	**5**	**6**	**7**	**8**	**9**
0	0	3.88	4.33	4.60	4.73	4.85	4.90	4.95
10	10	3.85	4.32	4.59	4.73	4.85	4.90	4.95
	20	3.84	4.32	4.59	4.73	4.85	4.90	4.95
	30	3.86	4.32	4.59	4.73	4.85	4.90	4.95
	100	3.92	4.31	4.58	4.72	4.85	4.90	4.95
20	10	3.82	4.31	4.59	4.73	4.85	4.89	4.95
	20	3.78	4.31	4.59	4.73	4.84	4.90	4.95
	30	3.78	4.32	4.59	4.73	4.84	4.90	4.95
	100	3.95	4.30	4.57	4.72	4.84	4.90	4.94
30	10	3.79	4.30	4.58	4.73	4.85	4.90	4.95
	20	3.71	4.31	4.59	4.73	4.84	4.90	4.95
	30	3.68	4.30	4.58	4.73	4.84	4.90	4.95
	100	3.96	4.27	4.56	4.71	4.84	4.89	4.94

Figure 2. Scatter diagram of the number of scale points by the ratio of the 1st and 2nd eigenvalues for the case wherein there are no misresponders

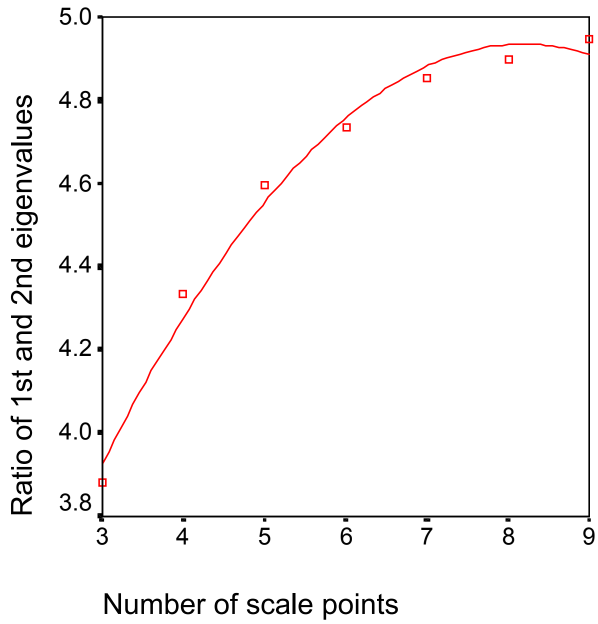


Figure 3. Scatter diagram of the number of scale points by the ratio of the 1st and 2nd eigenvalues for the case where 10% of respondents misrespond.

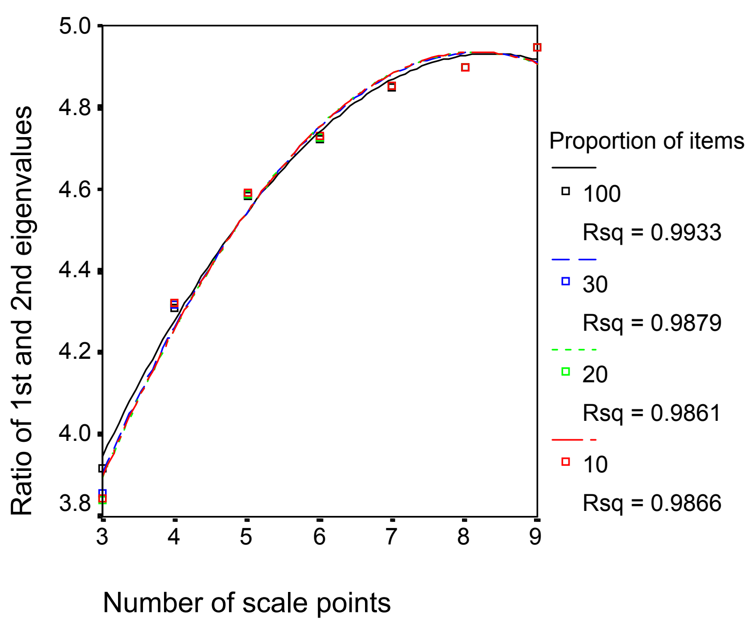


Figure 4. Scatter diagram of the number of scale points by the ratio of the 1st and 2nd eigenvalues for the case where 20% of respondents misrespond.

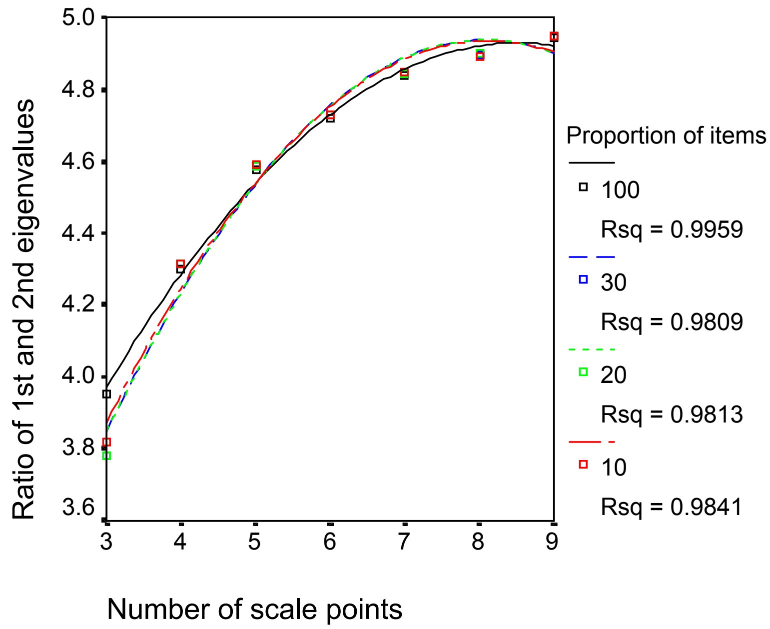
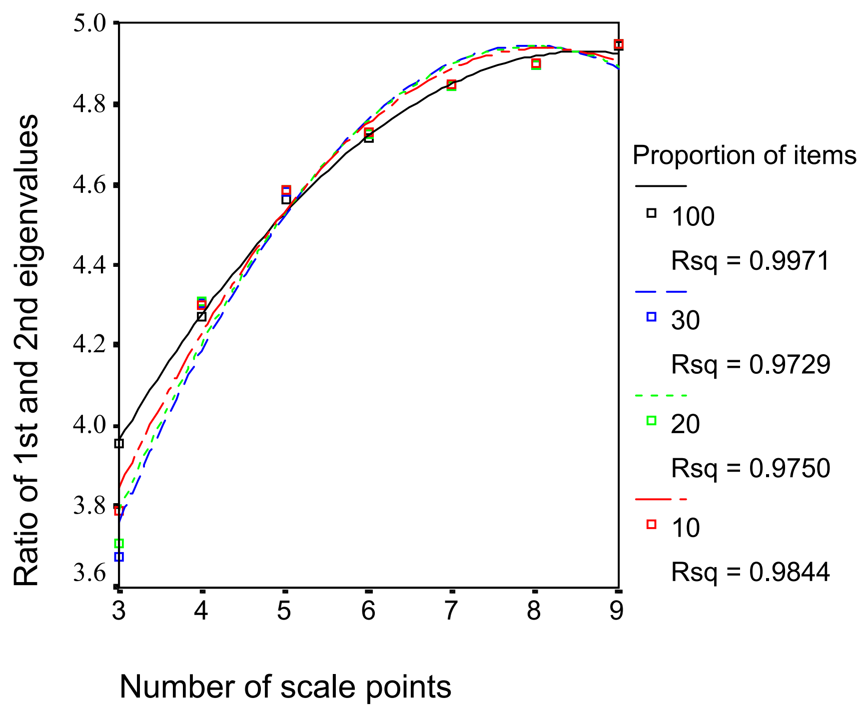


Figure 5. Scatter diagram of the number of scale points by the ratio of the 1st and 2nd eigenvalues for the case where 30% of respondents misrespond.



(b) The Chi-square fit statistic from the EFA ML estimation

- Tables 4 and 5 list the Chi-squared statistic values and the corresponding *p*-values, respectively, for the chi-square test of model-data fit. Clearly, if one has four or more scale points, the Chi-squared test will correctly identify the model as a one-factor model and this is irrespective of misresponding.
- However, if the number of scale points is three the results are mixed. That is, if there is no misresponding the Chi-squared test is accurate. However, misresponding on 20% or 30% of the items results in an incorrect statistical decision irrespective of the proportion of misresponders.
- Note that these results for the Chi-squared test were "population analogues" and because the degrees of freedom were the same in all chi-squares the differences in outcomes of the statistical decision were not affected by the degrees of freedom (*df*=35, as in the continuous case) of the test.

Table 4. Chi-square values for a one-factor model by the number of scale points.

Proportion of respondents who are misresponders	**Proportion of items on which they misrespond**	**Number of scale points**						
		3	**4**	**5**	**6**	**7**	**8**	**9**
0	0	39.541	30.151	26.543	37.574	34.640	28.661	40.690
10	10	41.709	30.142	26.657	37.352	34.773	28.771	40.917
	20	59.041	29.485	26.972	37.754	34.799	28.730	40.965
	30	85.599	28.696	26.644	37.514	34.514	28.769	40.768
	100	38.711	29.464	27.177	37.125	34.514	28.713	40.949
20	10	37.783	30.783	26.370	37.750	34.643	28.730	40.894
	20	92.129	29.659	26.479	37.751	34.652	28.717	40.911
	30	179.894	31.271	26.891	37.941	34.560	28.905	40.806
	100	36.392	29.545	26.142	35.995	33.144	27.503	40.678
30	10	36.876	31.176	26.544	37.721	34.572	28.937	41.002
	20	148.626	30.850	26.861	37.670	34.525	28.790	41.112
	30	316.746	33.965	27.384	37.887	34.414	29.048	41.128
	100	38.168	30.629	26.341	37.224	33.915	27.956	41.459

Table 5. Chi-square *p*-values for a one-factor model by the number of scale points.

Proportion of respondents who are misresponders	**Proportion of items on which they misrespond**	**Number of scale points**						
		3	**4**	**5**	**6**	**7**	**8**	**9**
0	0	.274	.701	.847	.352	.485	.767	.234
10	10	.202	.702	.843	.362	.479	.762	.227
	20	**.007**	.731	.832	.345	.478	.764	.225
	30	**.000**	.765	.844	.355	.491	.762	.232
	100	.306	.732	.825	.371	.491	.764	.226
20	10	.343	.672	.853	.345	.485	.764	.227
	20	**.000**	.723	.849	.345	.485	.764	.227
	30	**.000**	.649	.835	.337	.489	.756	.230
	100	.404	.729	.861	.422	.558	.813	.235
30	10	.382	.653	.847	.346	.489	.755	.224
	20	**.000**	.669	.836	.348	.491	.761	.220
	30	**.000**	.518	.817	.339	.496	.750	.220
	100	.327	.679	.854	.367	.520	.795	.210

* Bold indicates significant χ^2 values.

(ii) *Preserving the order of the factor loadings from the ML estimation*

When one interprets the results of a factor analysis, a common strategy is to interpret the relative *ordering* of the factor loadings with an eye toward identifying salient variables to help define a factor. What this translates to in practice is that the ordering of the variables in terms of their factor loadings becomes a matter of concern.

To investigate this issue we created a data matrix of the simulation results wherein the rows were defined by the variables (items 1 to 10) being factored and the columns the various factor solutions (91, the number of cells in the simulation design). Spearman correlations were used to investigate whether the rank order of the factor loadings change with different numbers of scale points and when some proportion of responders cannot discriminate between the upper two scale points. The average Spearman rank correlation across the 91 outcomes was 0.999, and ranged between 0.997 and 1.000. This indicated that the order of the factor loadings changes very little (to none) when you impose Likert-type categories or have misresponders. The salient variables will therefore be unaltered.

Reliability

For the continuous (typically unobserved) case the coefficient alpha was 0.840 (recall that we generated data that fit a one-factor model):

- From Table 6 we can see that: (a) there was very little effect of the number of scale points on coefficient α because in the no-misresponders case (i.e., the situation wherein we are only looking at the effect of the number of Likert scale points) the alpha values varied between 0.77 and 0.83, and (b) from Figure 6 we can see that there was a quadratic effect of the number of scale points such that with more scale points there is a diminishing return.
- Furthermore, from Figures 7 through 9 we can see that the proportion of items that were misresponded to appears to have little to no effect because the pattern we see for the number of Likert scale points was the same irrespective of the proportion of item that were responded to differentially.
- Given that there was no effect of the proportion of items responded to differentially, we collapsed over that factor of the design and graphed the quadratic effect of the number of scale points for the various proportions of misresponders in Figure 10. Again, the effect was minimal except for the clear quadratic effect.
- The above findings were supported in the statistical modeling of the simulation outcomes. That is, we fit a general linear (regression) model with the additive effects of the three factors (treating them as continuous variables) including a quadratic term for the number of scale points. The distribution of reliability coefficients was symmetric and

not near the bounds of coefficient α (i.e., zero and one) therefore no transformation of the data was necessary for the dependent variable α. The results only showed a statistically significant quadratic effect of the number scale points: model R-squared of 0.975, linear effect $F(1,86)=960.3$, $p<.0001$, quadratic effect of $F(1,86)=561.8$, $p<.0001$.

Table 6. Coefficient alpha estimates of scale reliability for a one-factor model by the number of scale points.

Proportion of respondents who are misresponders	Proportion of items on which they misrespond	Number of scale points						
		3	4	5	6	7	8	9
0	0	.7733	.8020	.8155	.8226	.8276	.8298	.8322
10	10	.7727	.8019	.8154	.8226	.8276	.8298	.8322
	20	.7722	.8017	.8154	.8226	.8276	.8298	.8322
	30	.7720	.8017	.8154	.8226	.8276	.8299	.8323
	100	.7752	.8008	.8148	.8223	.8274	.8297	.8322
20	10	.7723	.8018	.8154	.8226	.8276	.8298	.8322
	20	.7716	.8017	.8154	.8226	.8276	.8299	.8322
	30	.7710	.8016	.8155	.8226	.8276	.8299	.8323
	100	.7769	.7999	.8144	.8220	.8272	.8297	.8321
30	10	.7718	.8018	.8154	.8226	.8276	.8299	.8322
	20	.7706	.8016	.8154	.8226	.8276	.8299	.8323
	30	.7696	.8014	.8154	.8227	.8276	.8299	.8323
	100	.7767	.7983	.8137	.8216	.8270	.8296	.8320

Figure 6. Scatter diagram of the number of scale points by α for the case wherein there are no misresponders.

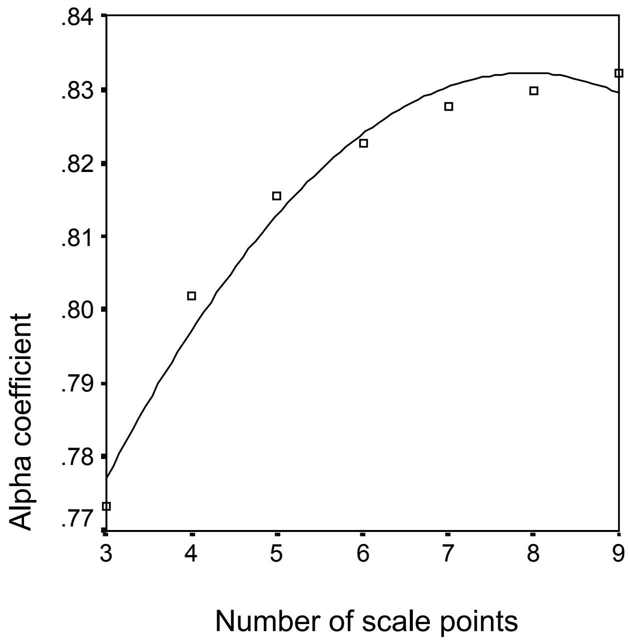


Figure 7. Scatter diagram of the number of scale points by α for the case where 10% of respondents misrespond.

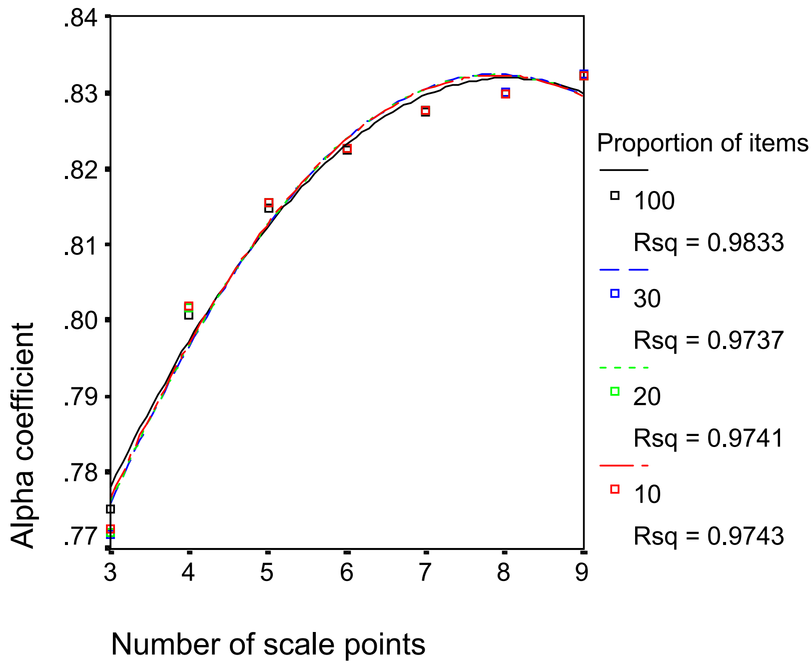


Figure 8. Scatter diagram of the number of scale points by α for the case where 20% of respondents misrespond.

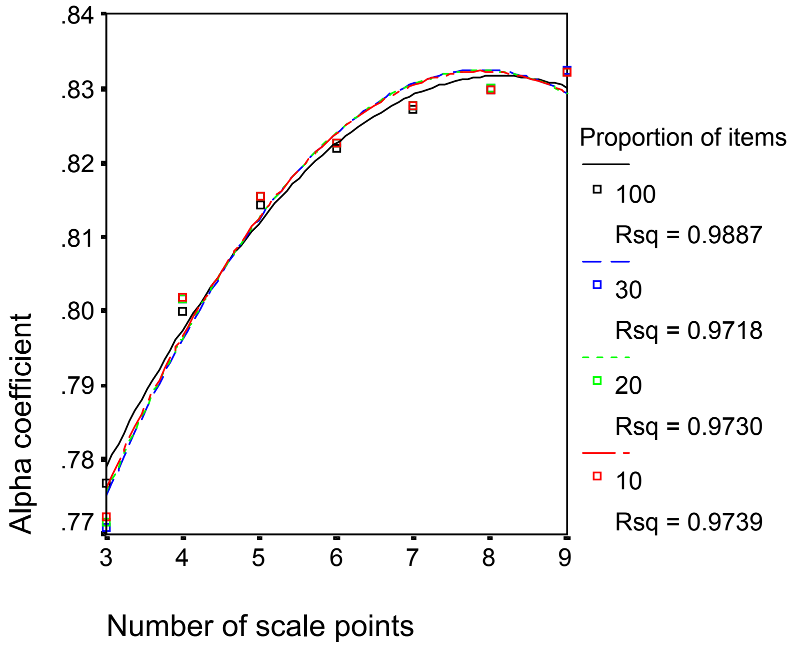


Figure 9. Scatter diagram of the number of scale points by α for the case where 30% of respondents misrespond.

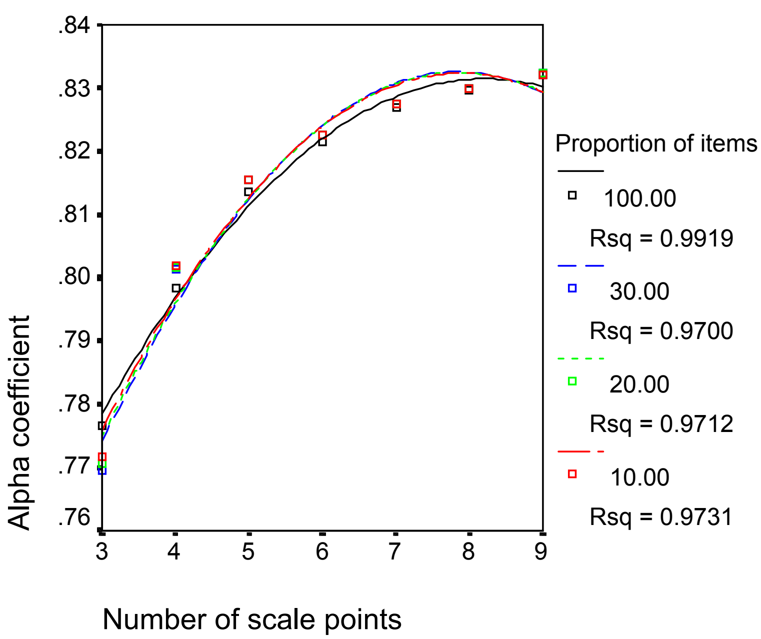
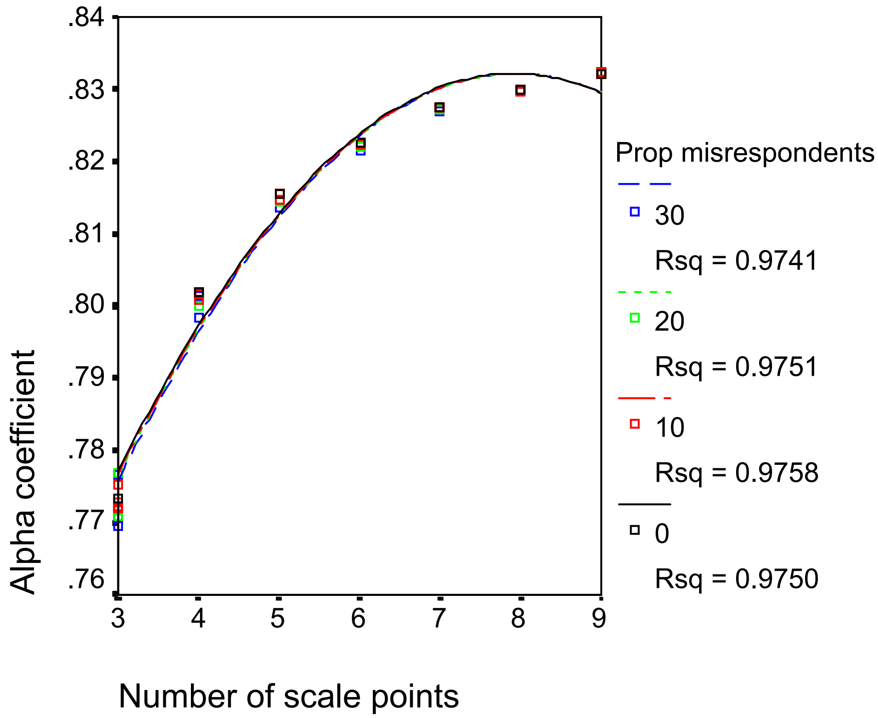


Figure 10. The plot depicting the effect of the proportion of misresponders, collapsed over the proportion of items responded to differentially.



Conclusions

- The effect of Likert scale points: there is a quadratic effect of the number of scale points with diminishing effect after 4 scale points. This effect holds for coefficient alpha and the ratio of the first two eigenvalues. The number of Likert scale points have no effect on the decision of the number of factors to retain as long as the criterion was a ratio of eigenvalues of three. If the criterion for the ratio is four, however, this decision rule behaves like the maximum likelihood Chi-squared test with incorrect decisions at three scale points, but correct decisions at four or more scale points. If a one-factor model is forced onto the data, the interpretation of that factor solution (via the salient loadings), is not effected by the number of scale points.

- The type of misresponding we investigated, for either proportion of misresponders and proportion of items misresponded to, has little to no effect on the estimated coefficient alpha, the decision on the number of factors to extract (i.e., retain), and the factor interpretations when one has items with four or more rating scale points.
- In future research our conceptualization of misresponding can also be expanded to include a variety of different forms. For example, there may be language and cultural barriers, when a instrument developed in America is translated to another language (e.g., the idea of a Likert scale instead of agree/disagree was very cumbersome for Chinese respondents; also, it is debatable whether children can distinguish 5 to 9 levels of agreement. Thus future work in this vein may use higher proportions of misresponders.
- Future work could also manipulate the type of misresponders. In this study, the top 2 categories were collapsed. But if there is misunderstanding at one end of the scale it could just as easily happen at both ends of the scale. Thus, we could collapse the top 2 and the bottom 2 categories. Also, the mispesponders could be equally likely to choose the outer or more central response. This could be done with a random choice mechanism. You can think of this as the probability of choosing the more extreme of the 2 collapsed options as (p=.5). We could simulate "middle of the road" misresponders who tend to choose the more centraly located option (say p=.25). Or we could simulated "extreme" misresponders who tend to choose the more extreme response (say p=.75). With symmetric misresponding (i.e., collapsing categories on both ends) there are several other posibilities. Misresponders who tend to choose the more "positve" or more "negative" option.

References

Babakus, E., Ferguson, C.E, & Joreskog, K.G.(1987). The sensitivity of confirmatory maximum likelihood factor analysis to violation of measurement scale and distributional assumptions. *Journal of Marketing Research, 24*, 222-228.

Bandalos, D.L., & Enders, C. (1996). The effects of nonnormality and number of response categories on reliability. *Applied Measurement in Education, 9*, 151-160.

Beasley, T. M. (August, 1992). *Order statistic confidence interval estimates for eigenvalues from non-normal variables*. Paper presented at the meeting of the American Psychological Association. Washington, DC.

Bernstein, I.H. & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin, 105*, 3, 464-477.

Birkett, N.J. (1989). Selecting the number of response categories for a Likert-type scale. In N. Rothwell (Chair), Cognitive aspects of survey methodology. Proceedings from the Survey Research Methods, American Statistical Association, USA, 488-492.

Bollen, K.A., & Barb, K. (1981). Pearson's r and coarsely categorized measures. *American Sociological Review, 46*, 232-239.

Chang, L. (1994). A psychometric evaluation of 4-point and 6-point Likert type scales in relation to reliability and validity. *Applied Psychological Measurement, 18*, 205-215.

DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling, 9,* 327-346.

Green, S.B., Akey, T.M., Fleming, K.K., Hershberger, S.L. & Marquis, J.G. (1997). Effect of the number of scale points on chi-square fit indices in confirmatory factor analysis. *Structural Equation Modeling, 4,* 108-120.

Johnson, D.R., & Creech, J.C. (1983). Ordinal measures in multiple indicator models: A simulation study of categorization error. American Sociological Review, 48, 598-407.

Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. *Educational and Psychological Measurement, 31*, 657-674.

Muthen, B. & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38*, 171-189.

Ochieng, C. O. (2001). *Implications of Using Likert Data in Multiple Regression Analysis*. Unpublished Doctoral Dissertation, University of British Columbia.

Franklin, S. B. , Gibson, D. J., Robertson, P.A., Pohlmann, J. T. and Fralish, J. S. (1995). Parallel analysis: a method for determining significant principal components. Journal of Vegetation Science, 6, 99-106.

Rupp, A. A, Koh, K., & Zumbo, B. D. (2003, April). *What is the impact on exploratory factor analysis results of a polychoric correlation matrix from LISREL/PRELIS and EQS when some respondents are not able to follow the rating scale?* Presented in the organized Symposium "The impact of using ordered responses such as Likert or rating scale data with commonly found statistical methods". Annual Meeting of the American Educational Research Association (AERA), Chicago, Illinois.

Zumbo, B. D., & Ochieng, C. O. (2002). *The Effects of Various Configurations of Likert, Ordered Categorical, or Rating Scale Data on Ordinal Logistic Regression Pseudo R-squared Measure of Fit: The case of the cumulative logit model*. Annual Meeting of the American Educational Research Association (AERA), New Orleans, Louisiana.

Zumbo, B. D. & Ochieng, C. O. (2003, April). *An overview of the symposium and the methodology*. Presented in the organized Symposium "The impact of using ordered responses such as Likert or rating scale data with commonly found statistical methods". Annual Meeting of the American Educational Research Association (AERA), Chicago, Illinois.