Random contamination and select response styles affecting measures of fit and reliability in factor analysis

Random Contamination and Select Response Styles Affecting Measures of Fit and Reliability in Factor Analysis

Allen G. Harbaugh (Boston University) & Min Liu (University of Hawaii)

Abstract
This research examines the effects of nonattending response pattern contamination and select response style patterns on measures of model fit (CFI) and internal reliability (Cronbach's α). A simulation study examines the effects resulting from percentage of contamination, number of manifest items measured and sample size. Initial results indicate that sample size very mildly affects CFI but does not influence α. Percent contamination decreases both CFI and α in a nearly linear fashion over a limited range of contamination. Finally, whereas an increase in the number of manifest items increases resilience to random contamination for α, the opposite was observed for CFI. An increase in the number of manifest items resulted in larger decreases in CFI. Implications are briefly discussed.

Key words: Likert-type data, reliability, model fit, nonattending response patterns, response styles

1. Objective

The purpose of this project is to systematically examine the effects of random contamination—a form of disingenuous response style—in survey responses. In particular, the goal is to assess the possible impact on both internal reliability measures (*e.g.*, Cronbach's α) and model-fit indices (*e.g.*, CFI & RMSEA) in the context of a confirmatory factor analysis (CFA). This project aims to answer the following research questions:

(1) How does the percentage of random (nonattending) response contamination (or of select response style patters or of a mixture of both) affect the CFI, RMSEA and α given different numbers of manifest items and samples sizes?
(2) How resilient are models with more manifest items to the effect of percent contamination?
(3) How resilient are models assessed with larger sample sizes to the effect of percent contamination?

2. Literature Review

With no intention of engaging in the decades-long debate over treating ordinal data as interval data, we acknowledge that it is common practice in social science research to treat Likert-type survey data as an interval-level measurement for psychometric investigation using exploratory or confirmatory factor analysis (Holgado-Tello, Chacón-Moscoso, Barbero-García & Vila-Abad, 2010). Indeed, there has been research conducted to suggest that such a strategy is not ideal, but also is probably not overly problematic (Babakus, Ferguson & Jöreskog, 1987; DiStefano, 2002). The focus here is not to call into question the use of the procedure, but to acknowledge that it is used frequently enough to warrant further investigation into other influential issues. For example, a review of the current literature does not readily reveal any research on the impact of random contamination on subsequent analyses of survey data in general, or of survey data used to measure latent constructs specifically.

Recent research by Liu and colleagues (Liu & Zumbo, 2007; Liu, Wu & Zumbo, 2010; Liu & Zumbo, 2012) has systematically examined the impact of outliers on Likert-type data in relation to measures of internal reliability and factor analysis model fit. However, comparable questions could be asked about random contamination of survey data or the presence of distinct response styles (e.g., extreme responders who treat Likert-type scales as dichotomous, choosing only the extreme options). In this research, it is assumed that an individual that responds one way on one item (e.g., nonattentively) will do so on all the items (e.g., all responses will be random). (Alternatively, a person may first respond attentively and then, perhaps due to fatigue, begin provide nonattending/random responses (N/RR). This was explored more fully by Barnette [1999], but will not be examined here.)

3. Method

A simulation experimental study will be conducted to examine the effect of N/RR contamination. This experimental design will generate data sets in which the following key factors are manipulated: sample size, number of manifest items in the congeneric model, and percent contamination of N/RRs. To simulate data closely modeling authentic data, values for the manipulated variables were chosen from personal experience, current theoretical understanding, and examination of patterns in authentic data sets. The choices of sample size are in line with other simulation studies in which a sample size of 200 would be considered small and 1000 would be considered large. The choice for the number of manifest items for this preliminary study were based on the conventional recommendation that single factors be measured by 3-5 items. Consequently, the values of 5 and 8 seemed to be reasonable compromises for this recommendation in which the numbers are far enough apart, but not too far beyond the recommended window. Finally, the choice for percentage of contamination was chosen based on percent contamination for a different response style. In particular, a variety of data sets were collected from friends and colleagues (and some were obtained via solicitation on ResearchGate). This data was used to assess a reasonable range for "naturally" occurring single-value responses (SVRs) in authentic data sets (be these responses genuine responses or examples of a disingenuous RS). Of the nearly 10,000 survey responses examined, 12.2% of them demonstrated the SVR pattern. As the items for the constructs measured all included reverse-coded items, this suggests some form of disingenuous responding (at least for those not responding with only the central value—but probably even for most of those—which occurred 9% for all of the SVRs). An examination revealed that single-value responding is not as straightforward to detect, because there are slight variations of the pattern. Consequently, random responding is even more difficult to detect. Thus, it seemed reasonable that an examination using a window of percentages near this value of 12.2% was reasonable for this initial exploration.

In this current initial report, 2 dependent variables (CFI and Cronbach's α) are examined (with the intention that the final report will include additional fit indices such as RMSEA). The 3 factors being manipulated here are (1) number of manifest items associated with the construct, (2) percent random contamination, and (3) sample size (with the intention that additional factors will be examined in subsequent analyses). For each analysis, percent contamination will range from 1 to 15%. Though this range may be smaller than that found in other comparable simulation studies, the inclusion of larger values began to introduce convergence issues in the model estimation process. Additionally, the smaller range seemed appropriate for this initial study as one purpose of this study is to determine if small amounts of contamination in Likert-type data have substantive consequences (though a larger range will be included in the final report). For these preliminary results, 3 complete simulations were run: 8 manifest items with sample size $n = 1000$, 8 manifest items with sample size $n = 200$, and 5 manifest items with sample size $n = 200$.

Likert-type data was simulated using a continuous underlying latent factor model. A response scale with granularity of 7 chosen. Likert-type scales with 5 and 7 categories appear frequently in the literature, and there is research to suggest that these may be the optimal number of categories based on perceptions of ease of use, interpretability and reliability (Preston & Colman, 2000). Additionally, the choice of 7 categories allowed for flexibility in assigning different means and variances to individual items in the simulation. (More specifically, means for each item were randomly selected from a uniform distribution ranging from 2 to 6 with a variance of $\sigma^2 = 1.6 - |\mu - 4|/2$. Variance was constrained thusly in an attempt to keep the resulting ordinal responses nearly-normally distributed.) By varying the means, this avoided a preponderance of central response patterns (all 4s) or mixtures of 2 response choices (all 3 and 4s or all 4 and 5s). A factor analysis model was used in which the standardized factor loadings ranged from 0.65 to 0.86. Once a continuous underlying variable was estimated, this was constrained to an ordinal measure by rounding. As the estimation method occasionally resulted in values below 1 or above 7, the data was censored at these values. (Thus, the τ estimates for the cut-points for the underlying latent distribution were arbitrarily fixed at {1.5, 2.5, …, 6.5}. With the additional—possibly unrealistic—normality assumptions for the purposes of this simulation, this should not greatly affect the generalization of these results.)

Once the data sets were generated, a contamination data set was generated consisting of random responses. The distribution of responses was nearly uniform with the addition of one constraint: responses on consecutive items were always distinct. Thus, if 1 was randomly chosen for the first response, the second response would be uniformly selected from 2 through 7. (This was an attempt to mirror the common misunderstanding of the nature of random data.) The percentage of contamination ranged from 1 to 15 by whole number increments.

Model parameters were then estimated using the sem package in R (Fox, 2006). Each data set was randomly assigned a different anchor item for the model estimation process (e.g., a different factor loading was chosen to be constrained to 1 for each analysis). For all data sets, if the first random choice encountered convergence issues (e.g., failure to converge or negative variance estimates), a different anchor item was selected. Via this protocol, all data sets resulted in convergent and interpretable models.

4. Results

The internal reliability (Cronbach's α) analysis demonstrated a noticeable effect for number of factors and percent contamination with little effect due to sample size. In all cases, the measure of reliability decreased in a nearly linear fashion as the percent contamination increased (though the linear trend may be a consequence of the small sampling range for the percent contamination). Analysis of change in manifest items is presented in Figure 1a. Results are as might be expected; survey constructs with fewer manifest items are more susceptible to effects of N/RRs. Analysis of change in sample size is presented in Figure 1b. There is no noticeable change in the amount of decrease in the internal reliability measure as the sample size changed from n = 200 to 1000 for 8 manifest items.

[insert Figures 1a & 1b]

The model fit (CFI) analysis demonstrated a noticeable effect for number of factors and percent contamination with little effect due to sample size. In all cases, the measure of model fit decreased in a nearly linear fashion as the percent contamination increased (though this may be a consequence of the small range of percentages sampled). Analysis of change in manifest items is presented in Figure 2a. Survey constructs with fewer manifest items are less susceptible to effects of N/RRs. Analysis of change in sample size is presented in Figure 2b. There is only a mildly noticeable change in the amount of decrease in the model fit measure as the sample size changed from n = 200 to 1000 for 8 manifest items. Results indicate larger samples may be more resilient to random contamination.

[insert Figures 2a & 2b]

5. Scientific Importance

A brief examination of the preliminary results indicates some key findings. First, as can be seen in Figure 1a, it appears that models with more manifest items are more robust to the effects of contamination by N/RRs. Additionally, the sample size appears to have little effect on this trend, as can be seen in Figure 1b. Regarding measures of fit, Figure 2a indicates that models with more manifest items are more susceptible to random response contamination. For example, the average model with 8 manifest items would fail to meet the CFI=0.90 acceptable cut-off level for good fit with as little as 5% random response contamination. Comparably, the models with 5 manifest items would this level of N/RR contamination would, on average, clear the moderate threshold of CFI=0.90, but are still relatively close to the better threshold of CFI=0.95. But, even for models with 5 manifest items, 9% N/RR contamination would result in the average model failing to reach the CFI=0.90 cut-off. Unlike the findings regarding internal reliability and sample size, sample size does appear to have a mildly protective feature against the effect of N/RR contamination. In particular, larger sample sizes appear to experience less decrease in CFI compared to smaller samples for comparable levels of N/RR contamination (see Figure 2b).

As survey research is quite prevalent in the social sciences, this research is valuable in providing a better understanding of potential issues that might arise when working with authentic data. In particular, this research provides the working researcher with an idea of how many manifest items should be chosen to measure a psychometric construct. Furthermore, this number can be appropriately adjusted based on an assessment of the likelihood of obtaining nonattending responses. Alternatively, a researcher that obtains less than desirable results despite working with a sound theory or well-established survey instrument may use this as a justification to locate and remove questionable responses.

Additionally, this research can be extended to other types of nonattending response contamination (Barnette, 1999). Future research can also examine additional dependent variables (e.g., SRMR, standardized factor loadings), measures of person fit, and other potentially influential factors (e.g., Likert-type scale granularity and complexity of multifactor models).

6. References

Babakus, E., Ferguson, C. E. Jr., & Jöreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research, 24*(2), 222–228. http://www.jstor.org/stable/3151512

Barnette, J. J. (1999). Nonattending respondent effects on internal consistency of self-administered surveys: A Monte Carlo simulation study. *Educational and Psychological Measurement, 59*(1), 38–46. doi:10.1177/0013164499591003

DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(3), 327–346. doi:10.1207/S15328007SEM0903_2

Fox, J. (2006). Structural equation modeling with the sem package in R. *Structural Equation Modeling, 13*(3), 465–486. doi:10.1207/s15328007sem1303_7

Holgado–Tello, F. P., Chacón–Moscoso, S., Barbero–García, I., & Vila–Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality and Quantity, 44*(1), 153–166. doi:10.1007/s11135-008-9190-y

Liu, Y., Wu, A. D., & Zumbo, B. D. (2010). The impact of outliers on Cronbach's coefficient alpha estimate of reliability: Ordinal/Rating scale item responses. *Educational and Psychological Measurement, 70*(1), 5–21. doi:10.1177/0013164409344548

Liu, Y., & Zumbo, B. D. (2007). The impact of outliers on Cronbach's coefficient alpha estimate of reliability. *Educational and Psychological Measurement, 67*(4), 620–634. doi:10.1177/0013164406296976

Liu, Y., & Zumbo, B. D. (2012). Impact of outliers arising from unintended and unknowingly included subpopulations on the decisions about the number of factors in exploratory factor analysis. *Educational and Psychological Measurement, 72*(3), 388–414. doi:10.1177/0013164411429821

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104*(1), 1–15. doi:10.1016/S0001-6918(99)00050-5

Figure 1a

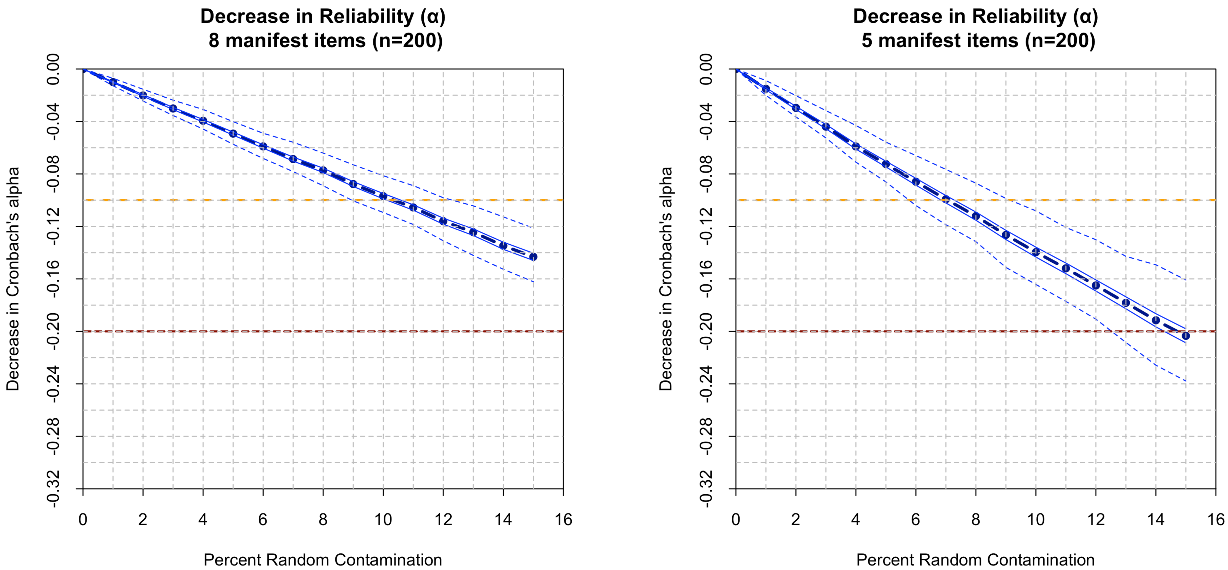
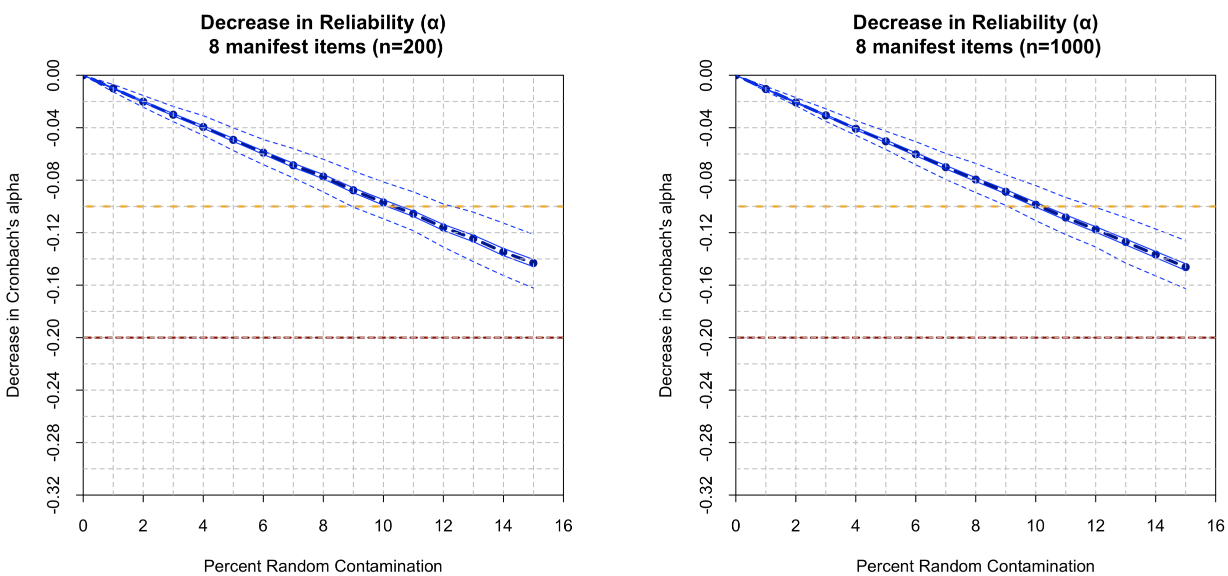


Figure 1b



Results obtained from 500 simulations per analysis. Narrow band enclosing line indicates the 95% confidence interval for the mean decrease. Dashed lines enclosing line indicate 1st and 3rd quartiles for the simulated data sets' decrease in α.

Figure 2a

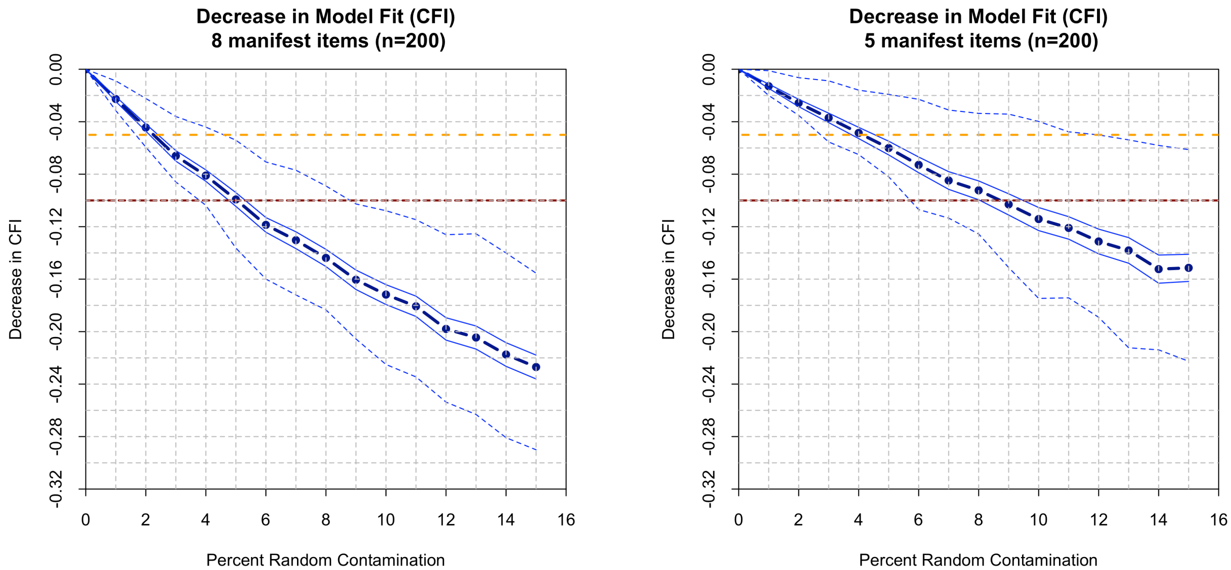
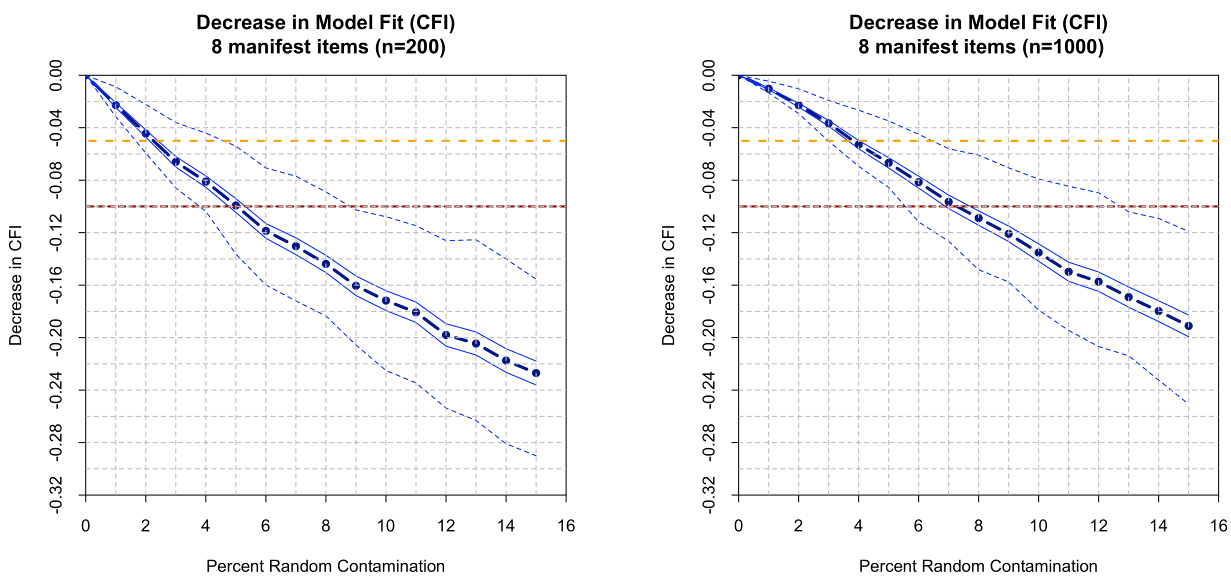


Figure 2b



Results obtained from 500 simulations per analysis. Narrow band enclosing line indicates the 95% confidence interval for the mean decrease. Dashed lines enclosing line indicate 1st and 3rd quartiles for the simulated data sets' decrease in CFI.