

Educational and Psychological Measurement

<http://epm.sagepub.com/>

The Impact of Outliers on Cronbach's Coefficient Alpha Estimate of Reliability: Ordinal/Rating Scale Item Responses

Yan Liu, Amery D. Wu and Bruno D. Zumbo

Educational and Psychological Measurement 2010 70: 5 originally published online 2 September 2009

DOI: 10.1177/0013164409344548

The online version of this article can be found at:

<http://epm.sagepub.com/content/70/1/5>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Educational and Psychological Measurement* can be found at:

Email Alerts: <http://epm.sagepub.com/cgi/alerts>


Subscriptions: <http://epm.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://epm.sagepub.com/content/70/1/5.refs.html>

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
© 2010 SAGE Publications
DOI: 10.1177/0013164409344548
<http://epm.sagepub.com>


Yan Liu,¹ Amery D. Wu,¹ and Bruno D. Zumbo¹

Abstract

In a recent Monte Carlo simulation study, Liu and Zumbo showed that outliers can severely inflate the estimates of Cronbach's coefficient alpha for continuous item response data—visual analogue response format. Little, however, is known about the effect of outliers for ordinal item response data—also commonly referred to as Likert, Likert-type, ordered categorical, or ordinal/rating scale item responses. Building on the work of Liu and Zumbo, the authors investigated the effects of outlier contamination for binary and ordinal response scales. Their results showed that coefficient alpha estimates were severely inflated with the presence of outliers, and like the earlier findings, the effects of outliers were reduced with increasing theoretical reliability. The efficiency of coefficient alpha estimates (i.e., sample-to-sample variation) was inflated as well and affected by the number of scale points. It is worth noting that when there were no outliers, the alpha estimates were downward biased because of the ordinal scaling. However, the alpha estimates were, in general, inflated in the presence of outliers leading to positive bias.

Keywords

outliers, reliability, coefficient alpha, Likert item responses, rating scale

¹ University of British Columbia, Vancouver, British Columbia, Canada

Corresponding Author:

Bruno D. Zumbo, Department of ECPS, The University of British Columbia, Scarfe Building, 2125 Main Mall, Vancouver, British Columbia, Canada V6T 1Z4
Email: bruno.zumbo@ubc.ca

The concern over outliers has a long history since the first time Bernoulli (1777/1961) pointed out the problem of discordant observations. It has been shown that outliers may distort the parameter estimation, such as mean, correlation, and regression parameters, and even a single outlier can severely bias the descriptive and inferential statistics (Blair & Higgins, 1980; Cook & Weisberg, 1980; Huber, 1981; Lind & Zumbo, 1993; Stevens, 1984). Hence, an essential task of data preparation is to determine if outliers appear in the data and how they affect the study's results.

Cronbach's coefficient alpha is widely used as an index of reliability and frequently reported in social and behavioral studies (Cronbach, 2004; Zumbo & Rupp, 2004). Recently, Liu and Zumbo (2007) systematically investigated the impact of outliers on coefficient alpha estimates. They investigated the effect of both symmetric and asymmetric outliers on coefficient alpha for continuous item response data, that is, the visual analogue item response format. Their study has shown that symmetric outliers do not affect coefficient alpha estimation, but asymmetric outliers artificially inflate the estimates of coefficient alpha. The estimates of coefficient alpha can be inflated to as high as .95 with a population reliability of .40 (i.e., a bias of .55) when the proportion of outliers (e.g., 8%) and the level of asymmetry of contamination are very high. That is to say, a measure with a poor reliability could be misconstrued as highly reliable because of asymmetric outliers. Obviously, an understanding of the impact of outliers is crucial for an appropriate interpretation of coefficient alpha.

In the social and behavioral sciences, most items used in scales and questionnaires typically use the Likert-type item response format. To respond to a question, the participants are asked to choose one of a given number of ordered response categories running from, for example, *never* to *very often*. The data arising from this type of item are ordinal categorical scales. However, the studies of how outliers affect the reliability of measures for this type of item response scale are rare. In a simulation study, Barnette (1999) investigated eight patterns of nonattending observations for a 50-item test with a 7-point Likert item response scale. Note that "nonattending observations" refer to those participants who respond to questionnaires or surveys inattentively because of fatigue or their lack of interest in participation. His study showed that different response item patterns had differential systematic effects on coefficient alpha estimates. Some patterns resulted in inflated alpha coefficient, some resulted in deflated alpha, and some had little effect.

In another recent study, Zijlstra, van der Ark, and Sijtsma (2007) investigated the influence of outliers on four commonly used statistics including Cronbach's alpha for categorical scales using 10 real data sets. They examined two types of outliers. The first type of outliers was defined as an individual's frequency of rating the unpopular response categories. The second type of outliers is the number of weighted Guttman (1950) errors (i.e., a respondent answers a relatively difficult item correctly but an easier item incorrectly). Their overall findings revealed that the first type of outliers inflated alpha, but the second type deflated alpha.

Outliers and Contamination Models

In psychometrics, outliers can be investigated from two perspectives: (a) in the course of a computer simulation, selecting a respondent and altering that respondent's scores (which we refer to as *person outliers*) and (b) in a computer simulation forming a mixture of multivariate distributions (which we refer to as *item outliers*). Person outliers can be defined as the outlying respondents in terms of their response patterns across some or all items. These outlying respondents are sometimes referred to as aberrant respondents, misresponders, or person misfit in item response theory.

Alternatively, one can investigate outliers from an item perspective by examining different distributions of outlying responses across individuals for each item. Although the person versus item outliers distinction may seem subtle, it is important because it represents different outlier processes, that is, aberrant *respondents* versus aberrant *item responses*, which can also be thought of as row versus column outliers in a typical data matrix. The approach of aberrant item responses, which is typically considered in the multivariate robust statistics literature, was adopted by Liu and Zumbo (2007) and is also used in the present study whereas the former one was adopted by Barnette (1999) and Zijlstra et al. (2007).

There are many sources that can cause the presence of outliers. Liu and Zumbo (2007) summarized three categories of possible sources of outliers: (a) the errors that occur during data collection (e.g., data-recording errors) and errors in preparing data for analysis (e.g., typos); (b) the unpredictable measurement-related errors from participants, including participants' guessing, inattentiveness, which may be caused by fatigue, and misresponding, which happens when, for example, participants misunderstand the instructions; and (c) inclusion of participants who do not belong to the target population.

Over decades, a number of researchers have provided several forms of mathematical contamination models for outliers. The most common mathematical formulation is the *mixed contamination model*, which regards outliers as a contamination fraction in the distribution. It is also known as *mixed normal distributions* or the *mixture contamination model* (e.g., Barnett & Lewis, 1978; Mosteller & Tukey, 1968; Zumbo & Jennings, 2002). The present study adopts this contamination model mathematical formulation.

The mixed contamination model is characterized by adding some outlying data points from a contamination distribution (denoted as P_c) into a parent normal distribution (denoted as P_p). The proportion of sampling from the parent distribution is denoted as p . The proportion of sampling from a contamination distribution is usually a small fraction, defined as $1 - p$.

For this contamination model, the working hypothesis is a statement of the parent probability model, H , from which the data X_i ($i = 1, 2, \dots, n$) are drawn as independent observations. We can denote it as

$$H : X_i \in P_p \quad (i = 1, 2, \dots, n).$$

The alternative hypothesis assumes a mixture of data from both the parent and contamination distributions. The alternative hypothesis is denoted as

$$\bar{H} : X_i \in p * P_p + (1 - p) * P_c \quad (0 < 1 - p < 1; i = 1, 2, \dots, n).$$

For the alternative hypothesis, there are two possible types of outlier contamination: symmetric contamination and asymmetric contamination. The contamination is symmetric if the population is a mixture of $N(\mu, \sigma)$ and $N(\mu, b\sigma)$ components where b is a positive constant, and $b\sigma > \sigma$ (if $b\sigma < \sigma$, the contamination distribution will not constitute outliers but inliers). The contamination is asymmetric when the population is a mixture of $N(\mu, \sigma)$ and $N(\mu \pm a, \sigma)$ or $N(\mu \pm a, b\sigma)$ components where a is constant, and $a \neq 0$. In the present study, we only manipulated the constant a but not b considering that the standard deviation does not play a role in the effect of outliers on reliability, as revealed by the Liu and Zumbo's (2007) study.

Ordinal Response Scales

Since Likert (1932) introduced the Likert-type item response scale, great attention has been drawn to how the number of response categories affects the psychometric properties of commonly used statistics (e.g., reliability, Pearson product moment correlation, regression analysis, and structural equation models). Assuming a continuous unobserved variable underlies individuals' response process, the less precise categorization of this continuous variable into an ordinal scale has been shown to increase measurement error. However, it is not yet clear how it affects the associated psychometric properties (e.g., Krieg, 1999; Weng, 2004).

The investigation of outliers on Cronbach's alpha for the ordered categorical scales is more complicated than that for the continuous scales. Before examining the impact of outliers, we need to consider the effect of the number of response categories on Cronbach's alpha. Previous studies have shown inconsistent results for the effect of the number of scale points on Cronbach's alpha. Some studies concluded that Cronbach's alpha was not, or hardly, affected by the number of response categories (e.g., Aiken, 1983; Matell & Jacoby, 1971; Wong, Chuen, & Fung, 1993). Others, however, revealed that reliability estimates increased as the number of response categories increased compared with the 2-point response scales (Guilford, 1954; Lissitz & Green, 1975; Nunnally, 1978). Lissitz and Green (1975) did a simulation study and found that Cronbach's alpha increased when the number of response categories increased from two to five but leveled off after that. Jenkins and Taber's (1977) and Bandalos and Enders's (1996) simulation studies revealed similar results to those reported by Lissitz and Green (1975). In line with previous research, a recent study by Zumbo, Gadermann, and Zeisser (2007) has shown that when comparing coefficient alpha computed from ordinal item responses with that computed from the underlying continuous scale, the estimates of coefficient alpha were downward biased, but the magnitude of bias decreased as the theoretical reliability increased, that is, scales with higher theoretical reliability were less affected by the number of scale points.

It is important to keep in mind that the literature on ordinal responses conceptualizes item responding (and how one arrives at item responses) as a manifestation of a continuous unobserved underlying variable. That is, a continuum is assumed to be underlying the individuals' ordinal responses, and the observed responses are the manifestation of respondents' amount of the underlying continuum exceeding a certain number of latent thresholds on that same underlying continuum. Formally, the observed ordinal response for item j with C response categories, where $c = 0, 1, 2, \dots, C - 1$, is defined by the latent variable y_j^* such that

$$y_j = c, \quad \text{if } \tau_c < y_j^* < \tau_{c+1},$$

where τ_c and τ_{c+1} are the latent thresholds on the underlying latent continuum, which may be spaced at nonequal intervals and satisfy the constraint $-\infty = \tau_0 < \tau_1 < \dots < \tau_{c-1} < \tau_c < \infty$. It is worth mentioning at this point that the latent distribution does not necessarily have to be normally distributed although it is commonly assumed because of its well-understood mathematical properties.

With the underlying variable model in mind, it is worth noting that there are three possible reliability coefficients in a study of ordinal responses (Zimmerman & Zumbo, 1993; Zumbo & Zimmerman, 1993).

- a. The theoretical (unobserved/latent variable) reliability $\rho_{xx'}^* = \text{var}(T^*) / [\text{var}(T^*) + \text{var}(e^*)]$, where $y^* = T^* + e^*$ assuming T^* and e^* are independent, and $\text{var}(\cdot)$ denotes variance of the variable in parentheses. This quantity is a variant of the theoretical reliability from classical test theory, for which many different methods of estimating it have been developed, for example, coefficient alpha, test-retest, and so on. The readers should recall that in classical test theory, a person's observed score equals to the sum of true score and measurement error, which is expressed as $y = T + e$ where y denotes the observed score, T the true score, and e the measurement error. Correspondingly, y^* , T^* , and e^* are theoretical parameters (as underlying variables to the response) because they can only be considered theoretically or observed in computer simulation studies.
- b. The observed population reliability, which is derived from the observed ordinal responses (i.e., the categorization of the underlying continuum into ordinal responses) of the entire population such that $\rho_{xx'} = \text{var}(T) / [\text{var}(T) + \text{var}(e)]$, where $y = T + e$ assuming T and e are independent, and the notation is the same as above.
- c. The observed sample reliability, which is derived from the same observed ordinal item responses of (b) but based on a sample drawn from the population.

The essential and subtle difference between (a) and (b) is that (a) is based on the "unobserved" underlying variable whereas (b) is based on the manifest or observed (ordinal) variable. When considering the performance of sample estimators, such as (c), it is conventional in statistical mathematics to compute the bias and the efficiency

(i.e., an estimate of the sample-to-sample variability with a smaller value indicating a better estimator). In the case of ordinal responses, however, the question arises as to what should be the “reference” or “target” for quantifying the bias and efficiency for the estimator. Should the bias and the efficiency be considered relative to the population reliability in (b) or the theoretical reliability in (a)? The two referents answer different research questions. For example, the bias of (c) relative to (b) would inform one how close, on average, the sample coefficient alpha computed from ordinal data in (c) would be to the population coefficient alpha in (b), which is also computed from ordinal data. The bias of (c) relative to (a) would inform one how close, on average, the sample coefficient alpha in (c) computed from ordinal data would be to the theoretical reliability in (a), which is defined for the underlying continuous distribution.

As Zumbo and Zimmerman (1993) reminded us, to study the effect of measurement scale (e.g., ordinal scaling), one needs to compare the sample estimate (c) to the underlying unobserved quantity in (a). Therefore, in the present study, we used the theoretical reliability, (a), as the reference reliability. We hypothesized, in the case of ordinal response scales, that not only the factors characterizing the contamination model (i.e., the proportion of outliers and the mean shift of contamination distribution) but also the number of response categories is one source of bias for the Cronbach’s alpha estimates.

Building on the recent study by Liu and Zumbo (2007), four factors were systematically manipulated in the present study: theoretical reliability, proportion of outliers, mean shift of the contamination distribution, and the number of response scale points. Furthermore, given the results of Zijlstra et al. (2007) from the first type of outliers (i.e., defined as an individual’s frequency of rating the unpopular response categories), which is akin to our item outliers approach, and Liu and Zumbo’s (2007) results for visual analogue scales, we anticipated that the coefficient alpha should also be inflated in this study. This, however, has not yet been investigated, and how the degree of inflation, the effect of sample-to-sample variability, and the effect of the number of response categories affect coefficient alpha was yet unknown.

Therefore, the purpose of this study is to investigate the impact of outliers on coefficient alpha for ordinal item response data. Note that we considered ordinal data from 2- to 7-point item response formats. In this context, we are considering binary data as a type of ordinal response, which is in line with how binary data are often considered in achievement and psychosocial measurement. To experimentally examine the impact of outliers, a Monte Carlo simulation was used in the present study. This study was designed to answer the following two interrelated research questions:

Research Question 1: How does the number of response categories affect the *bias* (and *efficiency*) of coefficient alpha estimates for different magnitudes of theoretical reliability without and with the presence of outliers?

Research Question 2: Assuming an outlier contamination model, how do the mean shift of the contamination distribution and proportion of outliers affect the *bias* (and *efficiency*) of coefficient alpha estimates for various numbers of scale points and magnitudes of theoretical reliability?

Method

The present study built on the findings and adapted the simulation design of Liu and Zumbo (2007). As mentioned above, we investigated outliers from an item perspective by examining different distributions of outlying responses across individuals for each item. In this section, we introduce the outliers and contamination models first, then describe the simulation of the data, and finally provide the analytical methods for investigating how outliers affect the Cronbach's coefficient alpha for ordinal scale item responses.

Simulation of Data

As in Liu and Zumbo (2007), a Monte Carlo simulation study was designed to answer the research questions. The item response data on a 14-item test were generated using an item common factor analysis model for each sample with a size of 100. For each mixed contamination population or each design cell, there were 100 replications. A coefficient alpha estimate was obtained for each replication in each cell, and then the average of 100 alpha estimates was calculated. The mixed contamination models were generated by varying the magnitude of four factors. Liu and Zumbo showed that sample size and standard deviation of the contamination distribution did not affect coefficient alpha estimates in their study of continuous item responses on a visual analogue scale. Thus, given that we simulated ordinal responses from an underlying continuous distribution (as described above), we assumed that these two factors were again unimportant and hence were not included in this study.¹ Hence, based on these findings, the present study included four factors: (a) the contamination proportions (1%, 8%, and 15%), (b) the mean shift of the sampling from the contamination distribution (P_c ; 0, 1.5, and 3), (c) the magnitude of theoretical reliability (.40, .60, .80, and .90), and (d) the number of scale points (2, 3, 4, 5, 6, and 7). Therefore, given the description of the design factors, the simulation study was a $3 \times 3 \times 4 \times 6$ completely crossed factorial design. The following paragraphs provide more detailed descriptions of the data simulation and the design.

The simulation of the data was conducted into two steps. In the first step, following Liu and Zumbo (2007), we generated the underlying continuous distributions using common factor analysis (a one-factor model). The formula to compute the theoretical reliability is as follows:

$$\rho_{xx'}^* = \frac{(\sum_{i=1}^m \lambda_i)^2}{(\sum_{i=1}^m \lambda_i)^2 + \sum_{i=1}^m \theta_{ii}}$$

where λ_i denotes factor loadings, θ_{ii} denotes the error variance derived from the common factor model, and m denotes the number of items. Using the above equation, we calculated and specified the factor loadings to be .213, .311, .471, and .625 to obtain the theoretical reliabilities of .40, .60, .80, and .90, respectively. All 14 items were generated with equal factor loadings. The underlying continuous distribution

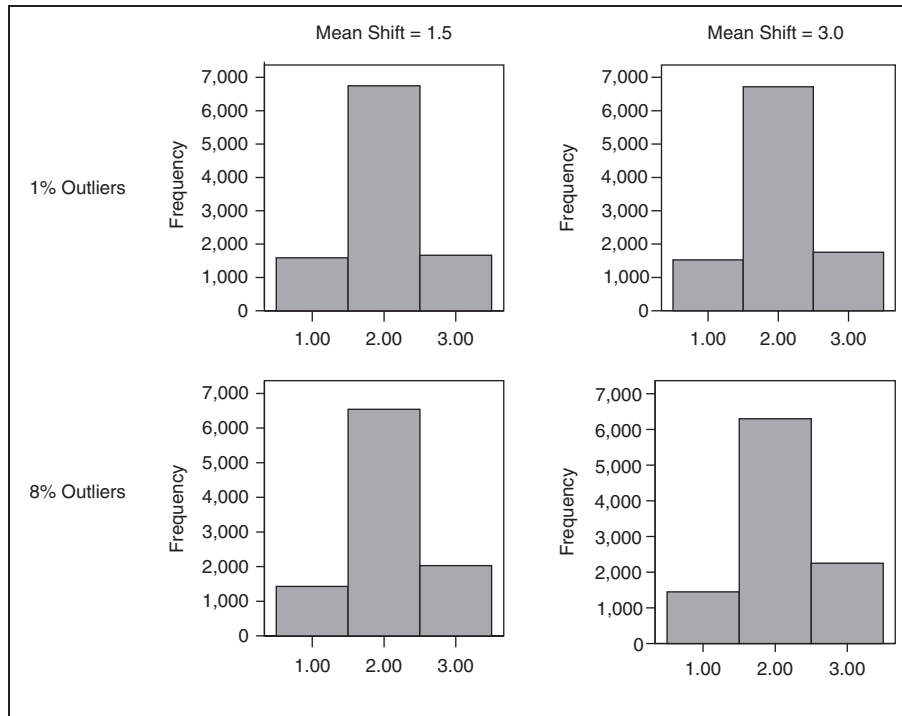


Figure 1. Demonstration of effects of the outlier proportion and the mean shift on the distributions of ordinal item response scales using an example of 3-data point scale

consisted of a parent population or a mixture of a parent population (i.e., mean shift = 0) and contamination population (i.e., mean shift = 1.5 or 3). Without the outliers, the underlying distribution is normally distributed whereas it becomes negatively skewed with the presence of outliers.

In short, in the first step, following Liu and Zumbo (2007), we used a covariance modeling approach (in our case an item factor analysis) to generate the underlying variates for the item response data. We began by creating covariance matrices based on the fundamental equations of factor analysis and then generated multivariate data based on these covariance matrices. The mixture of normal distributions comes into play when we generate multivariate data from the covariance matrices. At this point, we have the continuous underlying variates, the y^* , and can now turn to creating the ordinal (Likert or rating scale) item response data.

Therefore, in the second step, these underlying continuous distributions were transformed into ordinal item response scales by imposing the thresholds dividing the underlying continuum into intervals, as described in Zumbo et al. (2007). The number of item response categories ranges from 2 to 7. Figure 1 demonstrates how the degree of outlier proportion and the mean shift changed the symmetric distributions of the observed ordinal item responses using an example of 3-data point scale. With the

presence of outliers, the symmetric distributions of this 3-category response scale became negatively skewed, and the skewness increased with the increase of the mean shift and the proportion of outliers.

As in Liu and Zumbo (2007), the dependent variables in the simulation study are the bias and efficiency of coefficient alpha estimates. In this study, bias is defined as $Bias = E(\hat{\rho} - \rho)$ where $\hat{\rho}$ is an estimator of the parameter ρ . More specifically, $\hat{\rho}$ is the Cronbach's alpha estimates that were calculated based on the ordinal response data of the 100 replications, each with a sample size of 100, and ρ is the theoretical reliability—as described above, this is the theoretical reliability referencing the underlying response variable. The efficiency of the estimator is defined as $efficiency = E((\hat{\rho} - \rho)^2)$, which is the sample-to-sample variation in the estimates, that is, the ability to replicate the alpha estimates. The higher value of efficiency, the more difficult it is to replicate the findings. For the description of bias and efficiency of estimators, see Freund and Walpole (1980).

Data Analysis

As in Liu and Zumbo (2007), a factorial ANOVA (with up to 3-way interactions) was carried out and η^2 is used for the interpretation of simulation results. Although our design is a $3 \times 3 \times 4 \times 6$ crossed factorial, a three-way (incomplete) ANOVA was carried out for two reasons. First, the three-way (incomplete) ANOVA model has explained 99.7% of total variance for bias and 98.1% for efficiency and hence it is not necessary to add four-way interactions into the model because the model already accounts for nearly all of the variation. Second, the model will run out of degrees of freedom by adding the four-way interactions. To make sure that the inclusion of four-way interactions would not affect the conclusions, we also conducted multiple regression analyses that included two-, three-, and four-way interactions in the model to examine whether we can obtain consistent results from these two analyses—the regression analyses treat the explanatory variables as continuous and hence the degrees of freedom problem is no longer an issue. The results from the regression analysis were consistent with those of the ANOVA, and the four-way interaction did not explain much more variance in the dependent variable over and above the three-way interaction. Thus, we only reported the results from the ANOVA.

In addition, we ordered the relative importance of all main effects and interactions. Like R^2 in a regression analysis, η^2 is used as the indication of proportion of explained variance. Any main effects or interactions that account for less than 1% of the explained variance were considered to have trivial effect and were not interpreted.

Results

The results for bias and efficiency are described respectively in this section. It is worth noting that the conditions with mean shift of zero are a special case for both bias and efficiency, which demonstrate how the categorizations of the underlying continuous scales into ordinal/rating scales affect the estimates of coefficient alpha without the

Table 1. Variable Ordering for the Bias of Coefficient Alpha Estimates

	Sum Square	η^2	Cumulative Percentage of η^2
Intercept	0.734		
MEAN	1.386	0.288	28.9
RELIAB	0.949	0.197	48.6
PROPT	0.667	0.139	62.5
RELIAB * MEAN	0.620	0.129	75.4
PROPT * MEAN	0.362	0.075	83.0
DATPOINT	0.335	0.070	90.0
PROPT * RELIAB	0.265	0.055	95.5
PROP * RELIAB * MEAN	0.148	0.031	98.6
RELIAB * DATPOINT	0.027	0.006	99.1
MEAN * DATPOINT	0.021	0.004	99.6
RELIAB * MEAN * DATPOINT	0.012	0.003	99.8
PROPT * DATPOINT	0.003	0.001	99.9
PROPT * MEAN * DATPOINT	0.003	0.001	100.0
PROPT * RELIAB * DATPOINT	0.003	0.001	100.0
Error	0.005		
Total	5.542		
Corrected total	4.808		

Note: MEAN denotes mean shift of the contamination distribution; RELIAB denotes the theoretical reliability; PROPT denotes the outlier proportions; DATPOINT denotes the number of scale points.

presence of outliers. In discussing the results, we denote the four experimental factors as proportion of outliers (PROPT), theoretical reliability (RELIAB), the mean shift of the contamination distribution (MEAN), and number of item response categories or data points (DATPOINT).

Bias of Cronbach's Coefficient Alpha

The ANOVA results showed that the model explained 99.7% of total variance (Table 1). The partition of variance suggested that the main effects of all four factors, two-way interaction of RELIAB * MEAN, RELIAB * PROPT, PROPT * MEAN, and three-way interaction of PROPT * RELIAB * MEAN, met the importance criterion of accounting for more than 1% of the variance. If the higher order interactions were statistically significant, the main effects and the lower order interactions were not interpreted; for example, main effects or two-way interactions are not easily interpretable in the presence of three-way interactions (Howell, 1977). Hence, we only interpreted the three-way interaction PROPT * RELIAB * MEAN.

Figure 2 shows three plots for the interactions of PROPT and RELIAB for each of MEAN of 0, 1.5, and 3. It should be noted that for the case of no outliers, which is the mean shift of zero and/or when the proportion of outlier contamination is zero, the estimates of coefficient alpha were downward biased for all levels of theoretical

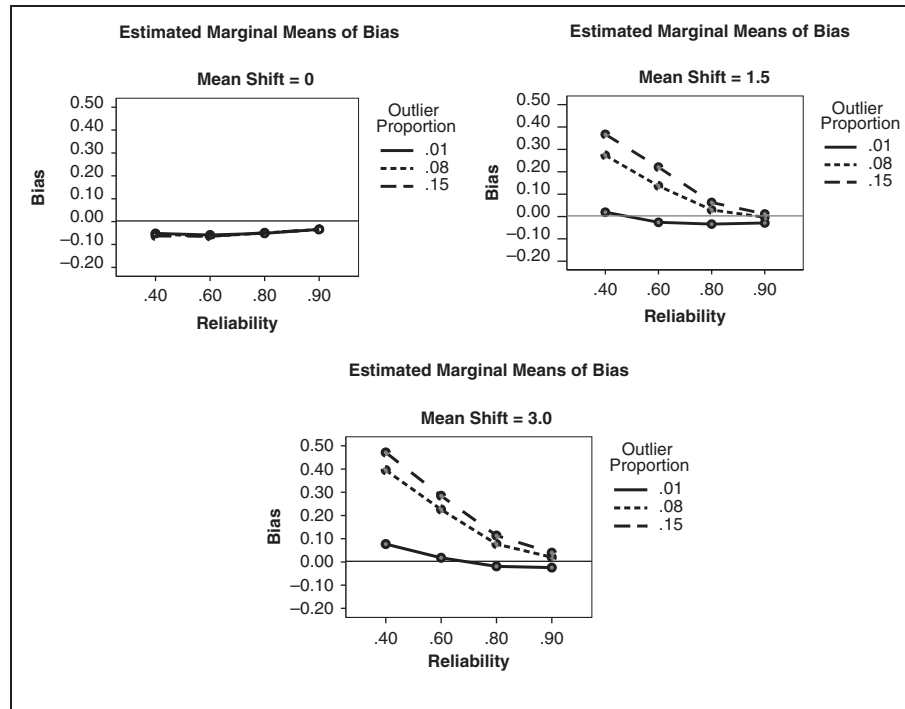


Figure 2. Three-way interaction effect on bias of coefficient alpha estimates (PROPT * RELIAB * MEAN)

reliability, which echoes the previous findings by Zumbo et al. (2007). However, such negative bias tapered off with increasing positive bias resulting from increasing mean shift and outlier proportion to an extent that the negative bias became zero and continued to increase positively. The estimates of coefficient alpha were severely inflated to .90 when the theoretical reliability was as low as .40. However, increasing theoretical reliability reduced such effect. For example, the theoretical reliability of .90 actually resulted in little bias, which is similar to the previous findings for continuous response scales by Liu and Zumbo (2007).

In summary, the results of bias of coefficient alpha indicate that mean shift and proportion of outliers (i.e., outlier contamination) inflate coefficient alpha estimates and that the effect decreases as theoretical reliability increases.

Efficiency of Cronbach's Coefficient Alpha

Table 2 presents the ANOVA results of efficiency of coefficient alpha, which shows that the model explained 98.1% of the total variance. Two of the three-way interactions met the criterion of accounting for at least 1% of the variance, which are PROPT * RELIAB * MEAN and RELIAB * MEAN * DATPOINT. The interaction of PROPT *

Table 2. Variable Ordering for the Efficiency of Coefficient Alpha Estimates

	Sum Square	η^2	Cumulative Percentage of η^2
Intercept	0.179		
RELIAB	0.200	0.321	32.2
RELIAB * MEAN	0.097	0.156	47.8
PROPT * RELIAB	0.069	0.111	58.9
MEAN	0.068	0.109	69.8
PROPT	0.052	0.083	78.1
PROP * RELIAB * MEAN	0.040	0.065	84.6
PROPT * MEAN	0.031	0.050	89.6
RELIAB * DATPOINT	0.015	0.024	92.0
MEAN * DATPOINT	0.014	0.023	94.2
RELIAB * MEAN * DATPOINT	0.011	0.018	96.0
DATPOINT	0.007	0.011	97.1
PROPT * DATPOINT	0.006	0.010	98.1
PROPT * RELIAB * DATPOINT	0.005	0.008	99.0
PROPT * MEAN * DATPOINT	0.004	0.006	99.6
Error	0.003	0.005	100.0
Total	0.802		
Corrected total	0.623		

Note: MEAN denotes mean shift of the contamination distribution; RELIAB denotes the theoretical reliability; PROPT denotes the outlier proportions; DATPOINT denotes the number of scale points.

RELIAB * MEAN shows that the efficiency of coefficient alpha was inflated in the presence of outliers, and the magnitude of inflation increased with the increase in mean shift of outliers and the proportions of outliers. However, the magnitude of inflation in efficiency decreased with the increase of theoretical reliability. This was consistent with the findings in the previous study for the continuous response data by Liu and Zumbo (2007). The plot of PROPT * RELIAB * MEAN is not presented here because this interaction presented a similar pattern as Figure 2 with the same interaction for bias.

Figure 3 shows the interaction of RELIAB * MEAN * DATPOINT, which indicates that the binary response scale resulted in the least inflation in efficiency compared with other response scales, and the values of efficiency increased as the number of scale points increased. However, increasing theoretical reliability led to less inflation in efficiency with theoretical reliability of .80 and .90 resulting in little or no inflation. Different from bias, the number of response categories does not play a role on efficiency in the case of no outliers.

Conclusion and Discussion

Coefficient alpha is a commonly reported reliability index used in the various fields of the social, behavioral, and health sciences. A previous study has shown that outliers can severely inflate the estimates of coefficient alpha for continuous item response

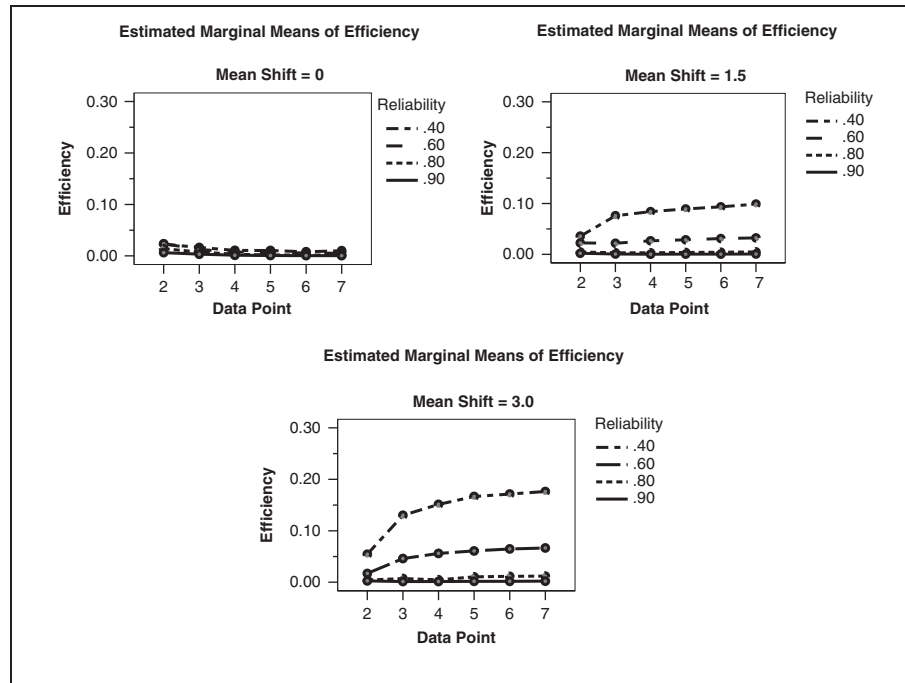


Figure 3. Three-way interactions on efficiency of coefficient alpha estimates (RELIAB * MEAN * DATPOINT)

data (Liu & Zumbo, 2007). The fact that ordinal data are not continuous and hence not normally distributed might further complicate the impact of outliers on coefficient alpha. However, little was known about whether and how outliers affect the estimate of coefficient alpha for ordinal item response data. Given the prevalence of binary and ordinal data in many fields of studies, this article was greatly needed to augment the literature.

Using a computer simulation methodology, the present study investigated the impact of outliers on bias and efficiency of Cronbach's coefficient alpha. Similar to the previous findings for the continuous response scale, the present study reveals that for ordinal response scales, outliers can inflate the estimates of coefficient alpha from .40 to as high as .90 and increase the sample-to-sample variation, which makes coefficient alpha difficult to replicate from sample to sample. This makes alpha very sample dependent.

To summarize our findings, first, increasing the proportions of contamination and mean shift, which characterizes outliers in the present study, increases the inflation of bias and efficiency of coefficient alpha. However, this effect is buffered by increasing theoretical reliability, which is, unfortunately, unknown to the day-to-day researcher. Second, in the context of no outliers, the estimates of coefficient alpha are downward

biased because of the ordinal scaling; such bias decreases with increasing number of scale points, whereas the efficiency of alpha is not affected by the ordinal scaling. In the presence of outliers, the number of scale points did not affect the bias of coefficient alpha but inflated the efficiency of coefficient alpha when increasing the number of scale points. The inflation effect of efficiency is also buffered by increasing theoretical reliability.

Readers should be cautious not to interpret the second finding as suggesting that fewer scale points is preferable to more scale points to reduce the sample-to-sample variation problem in the presence of outliers. Previous studies showed that fewer scale points are more likely to downwardly bias the coefficient alpha in the cases of no outliers because they provide less information (i.e., precision) and, hence, more measurement error (Bandalos & Enders, 1996; Jenkins & Taber, 1977; Lissitz & Green, 1975). This may, in turn, lead to less accurate parameter estimates and standard errors when the data are used for other analyses.

Although the number of scale points did not show their effect in the presence of other factors in the present study (i.e., mean shift of P_c , the proportion of outliers, and the theoretical reliability), readers need to be aware that this study did not present exhaustive outlier conditions, and hence, the number of scale points may affect coefficient alpha in other outlier conditions. More research is encouraged to examine how the number of scale points affect coefficient alpha when varying proportions of outliers spread in the ending categories of a various number of scale points. The outlying data points for fewer scale points are distributed in a different way from those for more scale points. For example, for three scale points, all outliers can only go to the ending category, whereas for seven scale points, outliers may spread out in the last two or three categories. The different patterns in distributing outliers for varying number of scale points may bring different magnitude of sample variances, which will correspondingly affect coefficient alpha.

In addition, the skewness of the observed ordinal distributions, because of categorizing the underlying continuous distribution with unequal intervals divided by the thresholds, is always a concern for data analysis of ordinal response scales. Zumbo et al. (2007) found that the estimates of coefficient alpha appear more biased when the ordinal observed distributions are skewed. However, the present study did not manipulate the effect of skewness because it is confounded with the effect of scaling the underlying continuous response variable into the observed ordinal variable.

It should be noted that in the data analysis phase of this study we did not transform the dependent variable “efficiency” even though the distribution of efficiency was quite skewed—the data were censored at zero because of the nonnegative values of efficiency. In cases like this, data should usually be transformed to be more normal before hypothesis testing. However, the purpose of our analyses was not to conduct hypothesis tests but to additively decompose the total variance, which is descriptive in nature. Furthermore, the transformation would make the modeling results difficult to interpret because we would have to consider the natural log (or square root) of efficiency, which is not understood in the statistical mathematics literature—in short,

the transformed variable needs to make sense in the research literature and is not just a statistical nicety (Osborne, 2002).

Our final remarks are to caution researchers that outliers can deteriorate the estimates of coefficient alpha for continuous as well as binary and ordinal data. Researchers should check their data for outliers even if the sample estimate of the coefficient alpha is high. As suggested by Lind and Zumbo (1993) and Wilcox (1992, 1998, 2005), robust estimates for coefficient alpha could be a possible solution to the outlier problem, which will provide relatively more accurate estimates of alpha and less efficiency.

Even without outliers, Cronbach's alpha may not be the appropriate estimator with ordinal item response data because alpha will underestimate the (theoretical underlying) reliability, especially in the binary case where the downward bias could be large. This finding is in line with those of Bandalos and Enders (1996), Jenkins and Taber (1977), and Lissitz and Green (1975). Therefore, the ordinal coefficient alpha newly developed by Zumbo et al. (2007) is recommended for binary and ordinal data. This new statistic has been demonstrated in their study to be an accurate and stable estimator for the theoretical reliability regardless of the number of scale points and the skewed distribution of the ordinal data. However, future research needs to investigate if outliers have an effect on the estimates of this new coefficient.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Financial Disclosure/Funding

The authors received no financial support for the research and/or authorship of this article.

Note

1. The reader will find that our assumption is borne out in the end because of the amount of variation the remaining factors account for in our statistical analysis of the simulation results below.

References

- Aiken, L. R. (1983). Number of response categories and statistics on a teacher rating scale. *Educational and Psychological Measurement, 43*, 397-401.
- Bandalos, D. L., & Enders, C. K. (1996). The effects of nonnormality and number of response categories on reliability. *Applied Measurement in Education, 9*, 151-160.
- Barnett, V., & Lewis, T. (1978). *Outliers in statistical data*. New York: John Wiley.
- Barnette, J. (1999). Nonattending respondent effects on internal consistency of self-administered surveys: A Monte Carlo simulation study. *Educational and Psychological Measurement, 59*, 38-46.

- Bernoulli, D. (1961). The most probable choice between several discrepant observations and the formation there from of the most likely induction. *Biometrika*, 48, 3-13. (Original work published 1777)
- Blair, R. C., & Higgins, J. J. (1980). The power of *t* and Wilcoxon statistics: A comparison. *Evaluation Review*, 4, 645-656.
- Cook, R. D., & Weisberg, S. (1980). Characterizations of an empirical influence function for detecting influential cases in regression. *Technometrics*, 22, 495-507.
- Cronbach, L. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, 64, 391-418.
- Freund, J. E., & Walpole, R. E. (1980). *Mathematical statistics* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.
- Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.
- Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Clausen (Eds.), *Measurement and prediction* (pp. 60-90). Princeton, NJ: Princeton University Press.
- Howell, D. C. (1997). *Statistical methods for psychology*. Belmont, CA: Duxbury Press.
- Huber, P. J. (1981). *Robust statistics*. New York: John Wiley.
- Jenkins, G. D., & Taber, T. D. (1977). A Monte Carlo study of factors affecting three indices of composite scale reliability. *Journal of Applied Psychology*, 62, 392-398.
- Krieg, E. F. (1999). Biases induced by coarse measurement scales. *Educational and Psychological Measurement*, 59, 749-766.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 1-55.
- Lissitz, R. W., & Green, S. B. (1975). Effect of the number of scale points on reliability: A Monte Carlo approach. *Journal of Applied Psychology*, 60, 10-13.
- Lind, J. C., & Zumbo, B. D. (1993). The continuity principle in psychological research: An introduction to robust statistics. *Canadian Psychology*, 34, 407-412.
- Liu, Y., & Zumbo, B. D. (2007). The impact of outliers on Cronbach's coefficient alpha estimate of reliability: Visual analogue scales. *Educational and Psychological Measurement*, 67, 620-634.
- Matell, M., & 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.
- Mosteller, F., & Tukey, J. W. (1968). Data analysis, including statistics. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology: Vol. 2. Research methods* (pp. 80-203). Reading, MA: Addison-Wesley.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Osborne, J. (2002). Notes on the use of data transformations. *Practical Assessment, Research & Evaluation*, 8. Retrieved October 21, 2008, from <http://PAREonline.net/getvn.asp?v=8&n=6>
- Stevens, J. P. (1984). Outliers and influential data points in regression analysis. *Psychological Bulletin*, 95, 334-344.
- Weng, L. J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, 64, 956-972.
- Wilcox, R. R. (1992). Robust generalizations of classical test reliability and Cronbach's alpha. *British Journal of Mathematical and Statistical Psychology*, 45, 239-254.

- Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist*, *53*, 300-314.
- Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd Ed.). San Diego, CA: Academic Press.
- Wong, C. S., Chuen, K. C., & Fung, M. Y. (1993). Differences between odd and even number of response scale: Some empirical evidence. *Chinese Journal of Psychology*, *35*, 75-86.
- Zijlstra, W. P., van der Ark, L. A., & Sijtsma, K. (2007). Outlier detection in test and questionnaire data. *Multivariate Behavioral Research*, *42*, 531-555.
- Zimmerman, D. W., & Zumbo, B. D. (1993). Relative power of parametric and nonparametric statistical methods. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Vol. 1. Methodological issues* (pp. 481-517). Hillsdale, NJ: Lawrence Erlbaum.
- Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, *6*, 21-29.
- Zumbo, B. D., & Jennings, M. (2002). The robustness of validity and efficiency of the related samples *t*-test in the presence of outliers. *Psicologica*, *23*, 415-450.
- Zumbo, B. D., & Rupp, A. A. (2004). Responsible modelling of measurement data for appropriate inferences: Important advances in reliability and validity theory. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 73-92). Thousand Oaks, CA: Sage.
- Zumbo, B. D., & Zimmerman, D. W. (1993). Is the selection of statistical methods governed by level of measurement? *Canadian Psychology*, *34*, 390-400.

This document was created with Win2PDF available at <http://www.win2pdf.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.
This page will not be added after purchasing Win2PDF.