A STUDY OF THE DISTRIBUTION OF SAMPLE COEFFICIENT ALPHA WITH THE HOPKINS SYMPTOM CHECKLIST: BOOTSTRAP VERSUS ASYMPTOTICS

KE-HAI YUAN
University of Notre Dame

CHARLES A. GUARNACCIA AND BERT HAYSLIP JR.
University of North Texas

Sample coefficient alpha is commonly reported for psychological measurement scales. However, how to characterize the distribution of sample coefficient alpha with the Likert-type scales typically used in social and behavioral science research is not clear. Using the Hopkins Symptom Checklist, the authors compare three characterizations of the distribution of the sample coefficient alpha: the existing normal-theory-based distribution, a newly proposed distribution based on fourth-order moments, and the bootstrap empirical distribution. Their study indicates that the normal-theory-based distribution has a systematic bias in describing the behavior of the sample coefficient alpha. The distribution based on fourth-order moments is better than the normal-theory-based one but is still not good enough with finite samples. The bootstrap automatically takes the sampling distribution and sample size into account; thus it is recommended for characterizing the behavior of sample coefficient alpha with Likert-type scales.

Keywords: *reliability; Cronbach's alpha; asymptotic distribution; multivariate kurtosis; quantile-quantile plot*

Sample coefficient alpha (Cronbach, 1951) is reported for most psychosocial measurement scales. Social scientists typically interpret alpha as measuring the sample internal-consistency reliability of items making up a scale or subscale. When the items are homogeneous, the coefficient alpha is the lower bound for the reliability coefficient of the composite score, and it is the reliability coefficient when the items are tau-equivalent (Lord & Novick, 1968; Zimmerman, 1972). Recently, Raykov (1998a) studied the situation when items are nonhomogeneous and found that coefficient alpha can also overestimate the reliability of a composite score at the population level. Even

Educational and Psychological Measurement, Vol. 63 No. 1, February 2003 5-23
DOI: 10.1177/0013164402239314

though improved estimation of the reliability coefficient of a composite score can be obtained by a confirmatory factor analysis of the item-level scores, in the applied literature coefficient alpha is still one of the most commonly reported characterizations of the accuracy of a measurement scale. Actually, sample coefficient alpha has been one of the most recommended estimates of reliability coefficient in standard textbooks on testing theory (Allen & Yen, 1979; Crocker & Algina, 1986; McDonald, 1999; Nunnally & Bernstein, 1994). In fact, the difference between coefficient alpha and the reliability coefficient is very minimal when the number of items in the composite score is large (Raykov, 1997b).

For a given scale, sample coefficient alpha differs from sample to sample due to sampling errors or changes in population characteristics. It is important to know when the difference is still within sampling errors not due to such population characteristics. Under the assumption of parallel items and normal data, Kristof (1963) and Feldt (1965) have obtained the exact distribution of the sample coefficient alpha. Based on the exact distribution, various inference issues for coefficient alpha have been formulated systematically by Feldt, Woodruff, and Salih (1987). Recently, van Zyl, Neudecker, and Nel (2000) obtained the asymptotic distribution of the sample coefficient alpha without assuming compound symmetry but still assuming item-level scores follow a multivariate normal distribution. In practice, item-level scores in the social sciences are typically of Likert-type scales. Thus, it is important to know how much guidance this normal-theory-based distribution can be when applied to data sets that are not normal.

Under some special conditions, Yuan and Bentler (2002) noted that the asymptotic distribution given by van Zyl et al. (2000) can still be valid for data sets with heterogeneous skewnesses and kurtoses. However, at present there is no available procedure to verify these conditions. When the distribution of a data set is unknown, the bootstrap procedure can yield inferences that automatically take the distribution shape of the data set into account. Because sample coefficient alpha is a smooth function of the sample covariance matrix, regularity conditions for bootstrap inference, as set in Bickel and Freedman (1981) and Mammen (1992) (see also Section 2.6 of Davison & Hinkley, 1997), are satisfied for coefficient alpha. Thus, the bootstrap procedure can be used to study the distribution of the sample coefficient alpha with Likert-type item-level scores. We can then use the bootstrap procedure to check the relevance of the normal-theory-based asymptotic distribution when applied to nonnormal data sets. In addition to studying the normal-theory-based asymptotic distribution of sample coefficient alpha, we also propose another asymptotic distribution to the sample coefficient alpha. Variance in this new distribution depends on the population fourth-order moments, which does not require the sampling distribution to be multivariate normal. These various procedures are introduced in the following section.

Following the introduction of these statistical procedures, we study the sample coefficient alpha for each of the five subscales of the Hopkins Symptom Checklist (HSCL) (Derogatis, Lipman, Rickels, Uhlenhuth, & Covi, 1974). The HSCL is a well-known psychological inventory commonly used in applied social science research and is related to the Symptom Checklist-90-R (SCL-90-R) (Derogatis, 1983). Because item-level scores on the HSCL are on a 4-point Likert-type scale, which is typical for psychological measurements, we use the bootstrap procedure to check the appropriateness of the normal-theory-based distribution for the sample coefficient alpha of the five subscales of the HSCL. In addition, we also contrast the newly proposed distribution to the bootstrap empirical distributions for the sample coefficient alpha of the five subscales. Standard errors as well as confidence intervals based on each procedure for the HSCL subscales are reported. Because distribution shapes, such as skewness and kurtosis, of the HSCL subscales may not be shared by other psychological inventories with Likert-type scales, the conclusions regarding the appropriateness of using asymptotic distributions for the HSCL subscales may not be generalized to other Likert-type scales. However, the statistical procedures and study conducted here can be applied equally to any other measurement scales.

Distributions of Sample Coefficient Alpha

For a sample $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ of dimension p, let $\mathbf{S} = (s_{ij})$ be the sample covariance matrix whose population counterpart is $\boldsymbol{\Sigma} = (\sigma_{ij})$. Then, the sample coefficient alpha is given by

$$\hat{\alpha} = \frac{p}{p-1}\left(1 - \sum_i s_{ii} \Big/ \sum_{ij} s_{ij}\right). \tag{1}$$

This $\hat{\alpha}$ is also the maximum likelihood estimate of coefficient alpha when data are multivariate normal. Let $\mathbf{1}$ be a vector of 1s with length p. The asymptotic distribution for $\hat{\alpha}$ obtained by van Zyl et al. (2000) is

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{\mathcal{L}} N(0, \omega^2), \tag{2a}$$

where $\alpha = p(1 - \Sigma_i\sigma_{ii}/\Sigma_{ij}\sigma_{ij})/(p-1)$ and

$$\omega^2 = 2p^2\, \{(\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1})\mathrm{tr}(\boldsymbol{\Sigma}^2) + (\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1})[\mathrm{tr}(\boldsymbol{\Sigma})]^2 - 2\mathrm{tr}(\boldsymbol{\Sigma})(\mathbf{1}'\boldsymbol{\Sigma}^2\mathbf{1})\} \,/\, [(p-1)^2(\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1})^3]. \tag{2b}$$

The variance in Equation 2b is obtained based on the normality assumption for the distribution of x_i. Notice that Equation 2b only involves the second-order moments of the population due to normality assumption. Because $\hat{\alpha}$ is a function of second-order sample moments, its variance should be given by fourth-order moments. This leads to another asymptotic distribution for $\hat{\alpha}$. Let $\mathbf{s} = \text{vech}(\mathbf{S})$ be the vector of nonduplicated elements of $\mathbf{S}$ by stacking its columns leaving out the elements above the diagonals, and $\mathbf{y}_i = \text{vech}[(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})']$ with $\bar{\mathbf{x}}$ being the sample mean of $\mathbf{x}_i$. We can then write $\mathbf{s}$ as

$$\mathbf{s} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{y}_i,$$

which is just the sample mean of $\mathbf{y}_i$. So the covariance matrix of $\sqrt{n}\mathbf{s}$ is estimated by $\mathbf{S}_y$, the sample covariance matrix of $\mathbf{y}_i$.

Let

$$g(\mathbf{s}) = \mathbf{a}'\mathbf{s}/(\mathbf{b}'\mathbf{s}),$$

where $\mathbf{a}$ is a vector of 1s and 0s with 1s corresponding to s_{ii} and 0s elsewhere, and $\mathbf{b}$ is a vector of 1s and 2s with 1s corresponding to s_{ii} and 2s elsewhere. Then we can rewrite Equation 1 as $\hat{\alpha} = p(1 - g(\mathbf{s}))/(p - 1)$. It is easy to see that the derivative of g at $\boldsymbol{\sigma} = \text{vech}(\boldsymbol{\Sigma})$ is given by

$$\mathbf{G} = \frac{1}{\mathbf{b}'\boldsymbol{\sigma}}\mathbf{a} - \frac{\mathbf{a}'\boldsymbol{\sigma}}{(\mathbf{b}'\boldsymbol{\sigma})^2}\mathbf{b}.$$

Applying a standard asymptotic technique (e.g., Ferguson, 1996) leads to

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{\mathcal{L}} N(0, \varphi^2), \tag{3a}$$

where

$$\varphi^2 = p^2\mathbf{G}'\boldsymbol{\Gamma}\mathbf{G}/(p - 1)^2 \tag{3b}$$

and $\boldsymbol{\Gamma} = \text{Cov}\{\text{vech}[(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})']\}$ is consistently estimated by $\mathbf{S}_y$.

Notice that both the distributions in Equations 2 and 3 are based on large-sample theory. The difference between Equations 2 and 3 is that Equation 2 depends on normality assumption, whereas Equation 3 does not. When data are approximately normal and sample size is large enough, Equation 2 should give a good approximation for the distribution of $\hat{\alpha}$. When sample size is large enough, Equation 3 may give a good approximation to the distribution of $\hat{\alpha}$ even when data are not normal. Yuan and Bentler (2002) characterized the conditions under which Equation 2 is valid when data are not normal. However, no procedure is available on how to verify these conditions with practical data. In such a situation, Equation 3 should be preferred for general nonnormal data. Because both Equations 2 and 3 use normal distribution to describe the behavior of $\hat{\alpha}$, the centers of the confidence intervals for α based on these two distributions are the same, whereas the lengths of the intervals are different. When sample size is not large enough, the approximation given by Equation 2 or 3 may not describe the distribution of $\hat{\alpha}$ to the desired precision. Especially, the distribution of $\hat{\alpha}$ may not be well approximated by a normal distribution for small to medium sample sizes.

When the distribution function F_0 of the underlying population is unknown, the empirical distribution function $\hat{F}$, which puts a probability of $1/n$ at each $\mathbf{x}_i$, is a consistent estimator of F_0. Bootstrap is a simulation technique based on independently drawing samples from $\hat{F}$ or sampling from $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ with replacement (Efron & Tibshirani, 1993). Because as a function of the sample covariance matrix $\mathbf{S}$, $\hat{\alpha}$ is continuously differentiable to an arbitrary order, bootstrap is thus a valid inference procedure for studying the distribution of $\hat{\alpha}$ (Bickel & Freedman, 1981; Mammen, 1992). Contrasting to the distributions in Equation 2 or 3 that are based on asymptotics, bootstrap approach does not assume that $\hat{\alpha}$ follows a normal distribution. One does not need to have a huge sample size to get quite accurate confidence intervals for α (Efron, 1987). The effect of finite sample size, which generally leads to larger standard errors than that based on asymptotics, is also automatically taken into account in the bootstrap procedure. In addition to studying the asymptotic distributions of Equations 2 and 3, we also use the bootstrap technique to study the distribution of $\hat{\alpha}$. Specifically, we contrast the standard errors and confidence intervals as well as the distribution shapes given by Equations 2 and 3 to those based on the bootstrap procedure.

Let $X^{(b)} = (\mathbf{x}_1^{(b)}, \mathbf{x}_2^{(b)}, \ldots, \mathbf{x}_n^{(b)})$ $b = 1, 2, \ldots B$ be independent bootstrap samples and coefficient $\hat{\alpha}$ evaluated at sample $X^{(b)}$ be $\hat{\alpha}_b$. Then, $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_B$ represents a bootstrap sample for $\hat{\alpha}$. Standard error of $\hat{\alpha}$, confidence intervals for α, as well as the distribution shape of $\hat{\alpha}$ can be obtained through this bootstrap sample. Specifically, the bootstrap standard error of $\hat{\alpha}$ is given by

$$\mathrm{SD_B} = [\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\alpha}_b - \overline{\alpha})^2]^{1/2}, \tag{4}$$

where $\overline{\alpha} = \Sigma_b \hat{\alpha}_b / B$. Let

$$\hat{\alpha}_{(1)} \leq \hat{\alpha}_{(2)} \leq \ldots \leq \hat{\alpha}_{(B)} \tag{5}$$

be the order statistics for the $\hat{\alpha}_b$ s. The empirical distribution that puts a probability of $1/B$ at each $\hat{\alpha}_{(b)}$ will be the bootstrap estimate for the distribution of $\hat{\alpha}$. The bootstrap percentile confidence interval for α with level 2β is given by

$$[\hat{\alpha}_{([B\beta])}, \quad \hat{\alpha}_{([B(1-\beta)])}],$$

where $[B\beta]$ is the integer part of $B\beta$.

When the empirical distribution of $\hat{\alpha}$ decided by Equation 5 is not symmetric, Efron (1987) developed a bias corrected and accelerated (BC_a) confidence interval that is superior to the bootstrap percentile confidence interval. One needs to calculate two additional numbers to obtain the BC_a interval. The bias correction number is calculated as

$$\hat{z}_0 = \Phi^{-1}(\#\{\hat{\alpha}_b < \hat{\alpha}\}/B),$$

where Φ is the cumulative distribution function of the standard normal distribution. The acceleration number $\hat{\alpha}$ is estimated by

$$\hat{a} = -\frac{\sum_{b=1}^{n}(\hat{\alpha}_b - \overline{\alpha})^3}{6\left\{\sum_{b=1}^{n}(\hat{\alpha}_b - \overline{\alpha})^2\right\}^{3/2}},$$

which is a form of the sample skewness of the $\hat{\alpha}_b$. Let $z^{(\beta)}$ be the 100βth percentile of the standard normal distribution and

$$\beta_1 = \Phi(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\beta)}}{1 - \hat{a}(\hat{z}_0 + z^{(\beta)})}), \ \beta_2 = \Phi(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\beta)}}{1 - \hat{a}(\hat{z}_0 + z^{(1-\beta)})}).$$

The BC_a confidence interval of Efron (1987) for α is

$$[\hat{\alpha}_{([B\beta_1])}, \quad \hat{\alpha}_{([B\beta_2)])}].$$

Notice that when $\hat{z}_0 = \hat{\alpha} = 0$, the BC_a interval is identical to the percentile interval.

Both the percentile and the BC_a intervals are transformation respecting. For example, when $f(\alpha)$ is a monotonic function of α, then the percentile confidence interval for $f(\alpha)$ is given by

$$[\,f(\hat{\alpha}_{([B\beta])}), \quad F(\hat{\alpha}_{([B(1-\beta)])})].$$

Suppose the exact distribution of $\hat{\alpha}$ is available and one can construct an exact confidence interval for α. The BC_a interval can approximate the exact confidence interval to the order of $1/n$, whereas the percentile confidence interval as well as the interval based on Equation 3, or Equation 2 when data are normal, can only approximate the exact confidence interval to the order of $1/\sqrt{n}$. So, without knowing the exact distribution of $\hat{\alpha}$ in general, the BC_a is the preferred confidence interval for α. More interpretation and discussion about BC_a are given in Efron (1987) and Efron and Tibshirani (1993).

Distributions of Coefficient Alpha for the HSCL

Source of HSCL Data

The HSCL has been widely used in psychosocial research for a quarter century. The HSCL is similar to the longer SCL-90-R (Derogatis, 1983), which is a measure of more severe forms of psychopathology. The HSCL is a measure of psychological symptoms common among outpatients in therapy or counseling treatment settings. The HSCL asks participants to rate how much each of 58 symptoms has bothered or distressed them during the past week. The measure uses a four-point Likert-type scale response format of 1 = *not at all*, 2 = *a little bit*, 3 = *quite a bit*, 4 = *extremely*. The overall 58-item scale and the five factors—Somatization (12 items), Obsessive-Compulsive (8 items), Interpersonal Sensitivity (7 items), Depression (11 items), and Anxiety (6 items)—are thought to demonstrate good alpha reliability and predictive validity (Derogatis et al., 1974).

The sample used here is from a previously reported study of coping with bereavement (Guarnaccia & Hayslip, 1998; Guarnaccia, Hayslip, & Landry, 1999; Hayslip, Ragow-O'Brien, & Guarnaccia, 1999). In this study, adults who had attended a funeral of a relative or a close friend within approximately the past 2 years were contacted through a variety of sources and asked to volunteer. Of those contacted, 419 interested individuals completed and returned a questionnaire packet of usable data. This funeral survey questionnaire included the HSCL and a host of other measures about circumstances related to the death, the funeral, and also the study volunteer's psychological adjustment and personality characteristics. This data set was chosen as an appropriate example of psychological research using a commonly used Likert-type scale, the HSCL, with participants whose scores are both elevated above those of the general population and have a nonnormal distribution. These distributional characteristics likely reflect variations in such psychological factors as emotional closeness to the deceased person, time since the death, and differential bereavement coping (Guarnaccia et al., 1999; Hayslip et al., 1999).

Distribution of Item Scores

We know the item-level scores for the five subscales are not normally distributed; yet it is more informative to know how far these distributions are from multivariate normality. We will use Mardia's (1970, 1974) multivariate skewness and kurtosis to characterize the departure from normality of these subscales. When data are normal, the standardized multivariate skewness asymptotically follows a chi-square distribution with degrees of freedom df =

$p(p + 1)(p + 2)/6$, and the standardized multivariate kurtosis asymptotically follows the standard normal distribution. The standardized multivariate kurtosis is part of the default output of popular software (e.g., Bentler, 1995) because of its direct impact on the analysis of sample covariances or correlations.

Sample coefficient alpha, standardized multivariate skewness, and kurtosis for each of the five subscales are reported in Table 1. The sample coefficient alphas in Table 1 are comparable with those reported in Derogatis et al. (1974). Regarding distribution shapes, none of the item scores for the subscales has significant multivariate skewness when referring to $\chi^2_{p(p+1)(p+2)/6}$. All the multivariate kurtoses are quite large. These results suggest that, although reasonably symmetric compared to the multivariate normal distribution, the item-level scores have heavy tails compared to a multivariate normal distribution.

Distribution of $\hat{\alpha}$

We will compare the different distribution shapes of Equations 2 and 3 with the empirical distribution of $\hat{\alpha}$ estimated by Equation 5 to see how good these asymptotic distributions are in describing the empirical behavior of $\hat{\alpha}$. Based on $B = 1{,}000$ bootstrap replications, Equation 5 was obtained for each of the subscales. Corresponding quantiles of Equations 2 and 3 based on $\hat{\alpha}(\hat{\omega}) \sim N(\hat{\alpha}, \hat{\omega}^2/n)$ and $\hat{\alpha}(\hat{\varphi}) \sim N(\hat{\alpha}, \hat{\varphi}^2/n)$ can be easily generated by

$$\alpha_b(\hat{\omega}) = \hat{\alpha} + \Phi^{-1}(\mathrm{b}/(B + 1))\hat{\omega}/\sqrt{n}$$

and

$$\alpha_b(\hat{\varphi}) = \hat{\alpha} + \Phi^{-1}(\mathrm{b}/(B + 1))\hat{\varphi}/\sqrt{n},$$

respectively. Quantile-quantile (QQ) plots, with $\hat{\alpha}_{(b)}$ on the vertical axis against $\alpha_b(\hat{\omega})$ and $\alpha_b(\hat{\varphi})$ on the horizontal axis, are shown in Figures 1 to 5 corresponding to the five subscales.

Bootstrap quantile $\hat{\alpha}_{(b)}$ versus $\alpha_b(\hat{\omega})$ for the HSCL Somatization subscale is given in Figure 1(a). Figure 1(a) shows that the left tail of the empirical distribution of $\hat{\alpha}$ is much heavier than that of the normal distribution in Equation 2, whereas the right tail of $\hat{\alpha}$ is somewhat heavier than that of Equation 2. This implies that the distribution of $\hat{\alpha}$ cannot be described well by the normal-theory-based formula. Contrasting to the distribution of $\hat{\alpha}$ given in Equation 2, Equation 3 describes the behavior of $\hat{\alpha}$ much better, as indicated by Figure 1(b), especially in the middle part of the distribution. However, the tails of $\hat{\alpha}$ cannot be described well by Equation 3 either. Actually, Figure 1(b) suggests that the distribution of $\hat{\alpha}$ may not be described by a normal distribution.

Table 1
Sample Coefficient Alpha, Multivariate Skewness, and Multivariate Kurtosis of Hopkins Symptom Checklist Subscales

Subscale	Alpha	Skewness	Kurtosis
Somatization (12 items)	.849	31.476	86.146
Obsessive-Compulsive (8 items)	.870	1.813	32.554
Interpersonal Sensitivity (7 items)	.847	0.128	27.890
Depression (11 items)	.898	6.569	60.218
Anxiety (6 items)	.821	0.048	64.775

Figure 2 gives the QQ plots corresponding to the HSCL Obsessive-Compulsive subscale. Similar to that in Figure 1, Equation 2 cannot describe the distribution of $\hat{\alpha}$ well. The distribution in Equation 3 describes the behavior of $\hat{\alpha}$ much better in its middle; the right tail of $\hat{\alpha}$ is only slightly lighter than that of Equation 3.

QQ plots for the HSCL Interpersonal Sensitivity subscale in Figure 3 also suggest that Equation 2 cannot describe the distribution of $\hat{\alpha}$ well. For this subscale, Equation 3 offers a better description of the upper part of the distribution of $\hat{\alpha}$.

Similarly, Figure 4 indicates that the distribution of $\hat{\alpha}$ for the HSCL Depression subscale cannot be described by the distribution in Equation 2. The distribution in Equation 3 describes the behavior of $\hat{\alpha}$ better for this subscale but not good enough. For example, most of the quantiles of $\hat{\alpha}$ are below the corresponding ones for Equation 3.

The distribution description for $\hat{\alpha}$ by Equation 2 or 3 is the worst for the HSCL Anxiety subscale, as indicated in Figure 5. All the quantiles of $\hat{\alpha}$ are below the corresponding ones for Equation 3 for this subscale. This is likely because of the fewer number of items, six, making up this subscale.

QQ plots in Figures 1 to 5 suggest that the distribution in Equation 2 has a systematic bias in describing the behavior of $\hat{\alpha}$, due to more variability in $\hat{\alpha}$ than $\omega/\sqrt{n}$. The distribution in Equation 3 corrects this bias to some degree. However, the left tail of $\hat{\alpha}$ is heavier than that of Equation 3, and the right tail of $\hat{\alpha}$ is lighter than that of Equation 3. The plots also implicitly suggest that the distributions for the item-level scores of the five HSCL subscales do not satisfy the robustness conditions for the normal-theory-based results applied to nonnormal data (Yuan & Bentler, 2002).

Standard Errors and Confidence Intervals

Standard errors of $\hat{\alpha}$ based on Equations 2, 3, and 4 are given in Table 2. In comparing these three types of standard errors, those based on Equation 2 are

(text continues on p. 19)

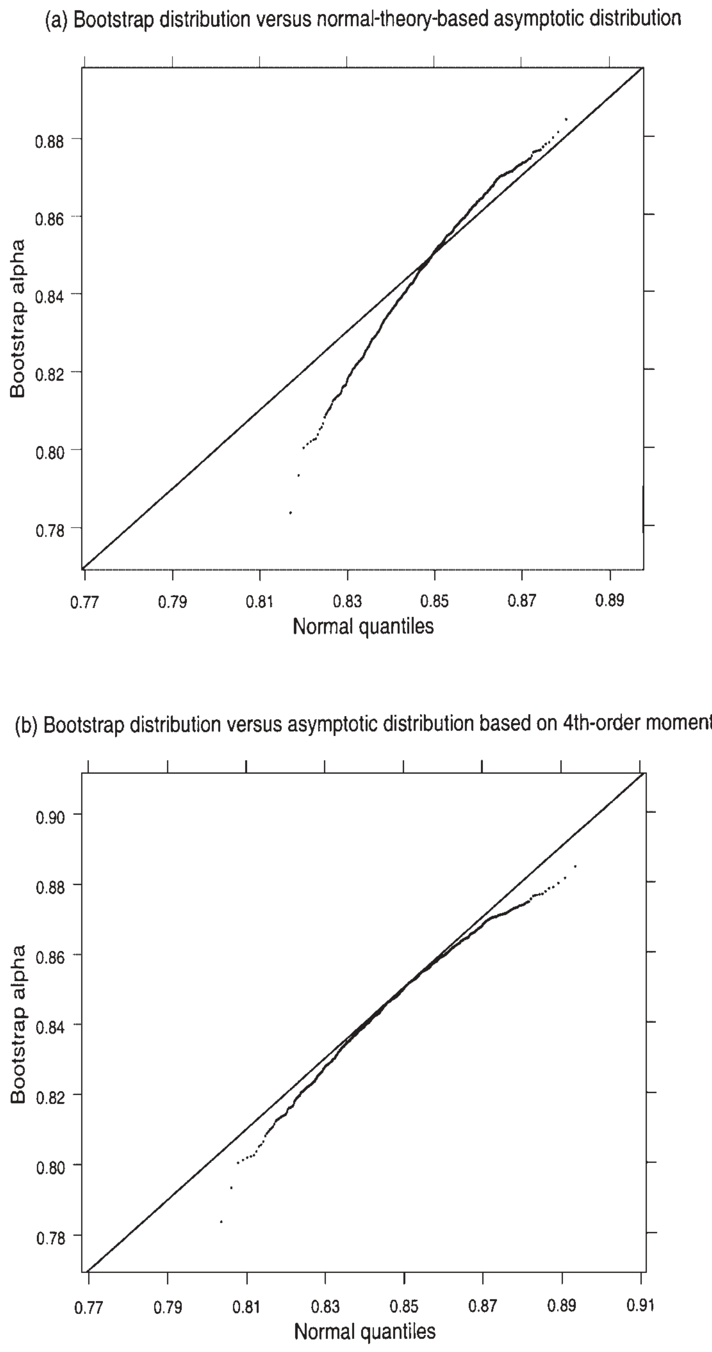

Figure 1. Plots of $\hat{\alpha}_{(b)}$ versus quantiles of $N(\hat{\alpha}, \hat{\omega}^2/n)$ and $N(\hat{\alpha}, \hat{\varphi}^2/n)$ (12-item Somatization subscale).

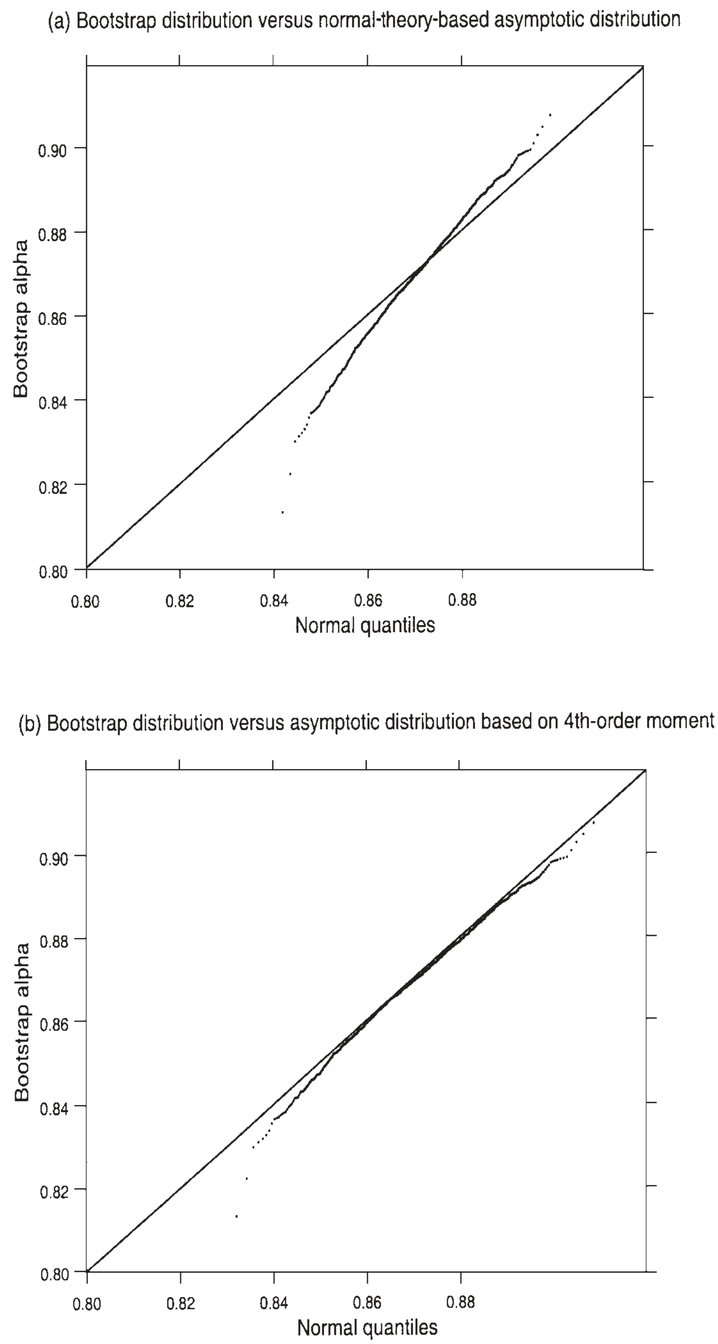

Figure 2. Plots of $\hat{\alpha}_{(b)}$ versus quantiles of $N(\hat{\alpha}, \hat{\omega}^2/n)$ and $N(\hat{\alpha}, \hat{\varphi}^2/n)$ (8-item Obsessive-Compulsive subscale).

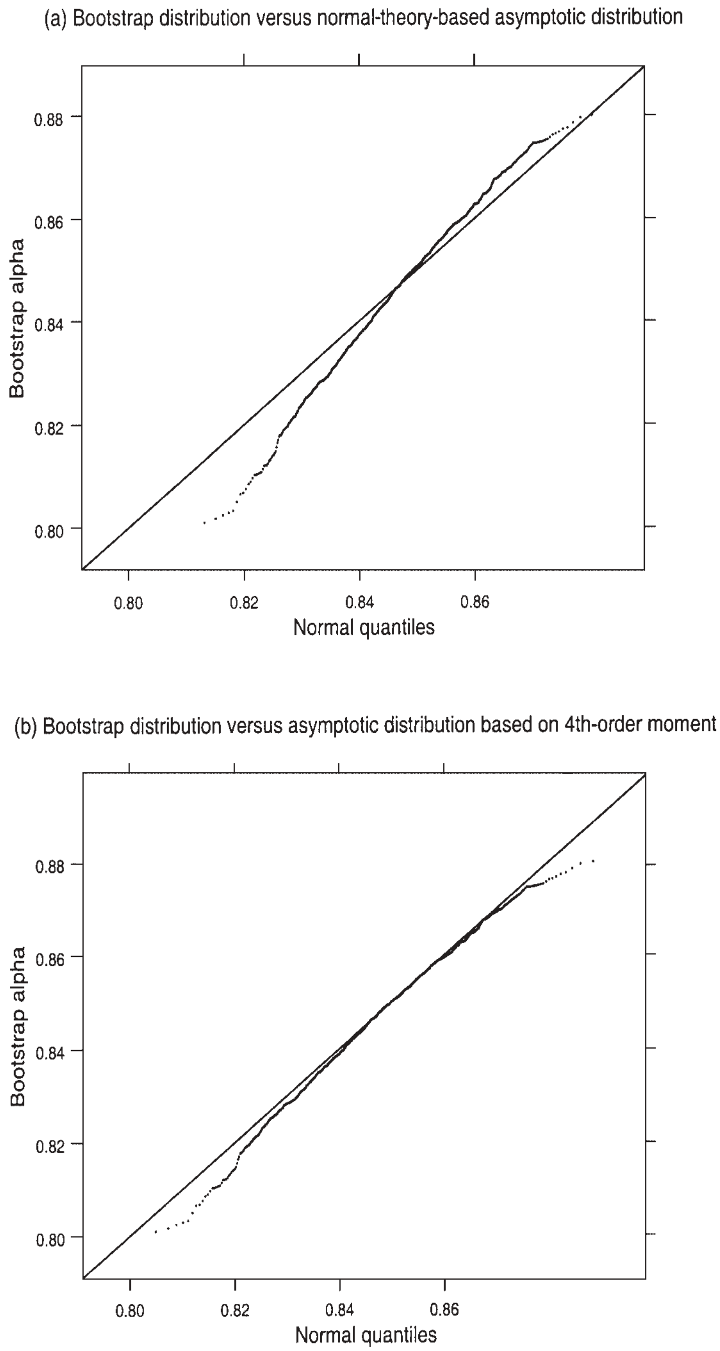

Figure 3. Plots of $\hat{\alpha}_{(b)}$ versus quantiles of $N(\hat{\alpha}, \hat{\omega}^2/n)$ and $N(\hat{\alpha}, \hat{\varphi}^2/n)$ (7-item Interpersonal Sensitivity subscale).

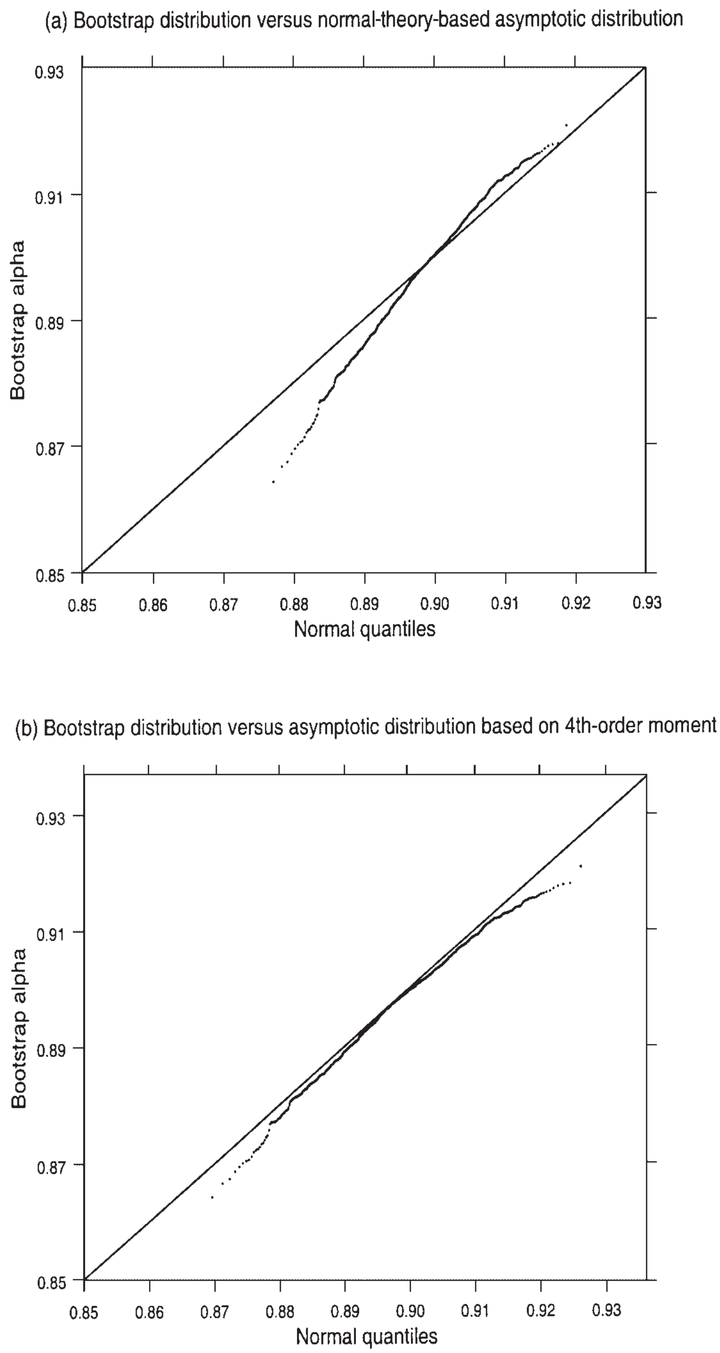

Figure 4. Plots of $\hat{\alpha}_{(b)}$ versus quantiles of $N(\hat{\alpha}, \hat{\omega}^2/n)$ and $N(\hat{\alpha}, \hat{\varphi}^2/n)$ (11-item Depression subscale).

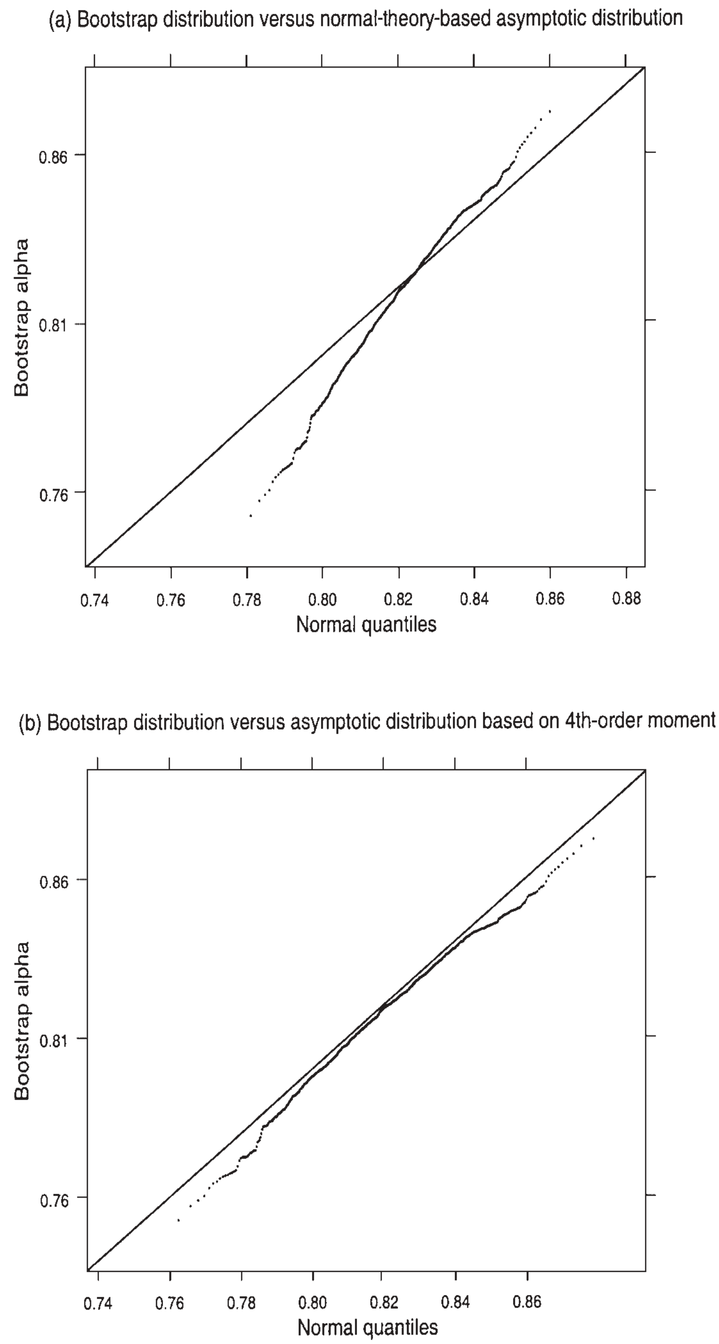

Figure 5. Plots of $\hat{\alpha}_{(b)}$ versus quantiles of $N(\hat{\alpha}, \hat{\omega}^2/n)$ and $N(\hat{\alpha}, \hat{\varphi}^2/n)$ (6-item Anxiety subscale).

the smallest. But they are by no means the optimal ones. This is because item-level scores for all the subscales have highly significant multivariate kurtosis, whereas Equation 2 only uses a multivariate normal distribution to describe their behavior. Standard errors based on Equation 3 take into account the heavy tails of the item-level scores through $\mathbf{\Gamma}$, but they are still somewhat smaller than the corresponding bootstrap standard errors. This is because the distribution characterized by Equation 3 is based on asymptotics, which is generally too optimistic for standard errors of parameters estimated with finite samples (Yuan & Bentler, 1997). These aspects are also reflected in confidence intervals for α, as reported in Table 3. For example, the confidence intervals based on Equation 2 are the shortest, but their coverage probability may not approach 0.90. Because Equation 3 cannot describe the tail behavior of $\hat{\alpha}$ well, confidence intervals based on it are different from bootstrap confidence intervals. From Figures 1 to 5, we know the distribution of $\hat{\alpha}$ is not symmetric for any of the five subscales; there also exist differences between the two type of bootstrap confidence intervals. Because BC_a interval has the best statistical property that approaches the exact confidence interval based on the exact distribution of $\hat{\alpha}$, the intervals in the last column of Table 3 are the recommended ones for predicting the size of the αs.

Conclusion and Discussion

Coefficient alpha is one of the most commonly reported reliabilities of scores in the social sciences. Several authors have characterized the distribution of sample coefficient alpha under various conditions. Because psychological scales at the item level are often of a Likert-type format, it is important to have a method to study the distribution of the sample coefficient alpha in such situations. It is also interesting to know whether the distribution of $\hat{\alpha}$ under normality assumption can still be used to infer the population alpha for a Likert-type scale. Using the bootstrap technique, we found that the normal-theory-based distribution does not describe well the behavior of $\hat{\alpha}$ for the HSCL scale. Even though the newly proposed distribution for $\hat{\alpha}$ in Equation 3 does not depend on the shape of the sampling distribution and is justified by large sample theory, it cannot describe the tail behavior of $\hat{\alpha}$ with the desired precision due to its slow convergence. The sample size of the HSCL in our study is $n = 419$, a size typical for applied social science data sets. For a larger n, the distribution in Equation 3 will describe the behavior of $\hat{\alpha}$ better. However, the discrepancy between Equations 2 and 5 is not because of sample size but because of the nonnormal nature of the underlying sampling distribution with such Likert-type scales.

When the exact confidence interval is available, which will be the case when $\mathbf{\Sigma}$ is of compound symmetry and the sample follows a multivariate normal distribution (Feldt, 1965; Kristof, 1963; van Zyl et al., 2000), then one can obtain the exact confidence interval for α. The exact confidence interval

Table 2
Three Types of Standard Errors × 10^2 of Sample Coefficient Alpha of Hopkins Symptom Checklist Subscales

Subscale	$\hat{\omega}/\sqrt{n}$	$\hat{\varphi}/\sqrt{n}$	SD_B
Somatization	1.060	1.511	1.535
Obsessive-Compulsive	0.955	1.284	1.309
Interpersonal Sensitivity	1.125	1.400	1.458
Depression	0.700	0.949	0.949
Anxiety	1.326	1.947	1.963

Table 3
Four Types of Confidence Intervals for Coefficient Alpha of Hopkins Symptom Checklist Subscales

Subscale	Based on Equation 2	Based on Equation 3	Percentile	BC_a
Somatization	[.831, .866]	[.824, .873]	[.820, .870]	[.822, .871]
Obsessive-Compulsive	[.855, .886]	[.849, .891]	[.847, .890]	[.850, .893]
Interpersonal Sensitivity	[.828, .865]	[.824, .870]	[.821, .869]	[.822, .870]
Depression	[.886, .909]	[.882, .914]	[.881, .912]	[.882, .913]
Anxiety	[.799, .842]	[.789, .853]	[.784, .847]	[.790, .851]

is generally superior to that based on the bootstrap technique (Efron, 1988; Rasmussen, 1988; Strube, 1988). The compound symmetry assumption allows for the fact that the asymptotic distribution based on normality can approximate the exact distribution of $\hat{\alpha}$ very well (van Zyl et al., 2000). However, the conditions of normality and compound symmetry may not hold for item-level scores of Likert-type scales. Our results indicate that inferences about coefficient alpha with the HSCL subscales based on the normal-theory-based asymptotic distribution will be misleading. Yuan and Bentler (2002) showed that the distribution of $\hat{\alpha}$ in Equation 2 is robust for a large class of distributions with heterogeneous skewnesses and kurtoses. The above results imply that the item-level Likert-type scores of the HSCL scale do not belong to the class of distributions characterized in Yuan and Bentler (2002).

As a nonparametric technique, bootstrap not only takes into account the shape of the sampling distribution but also the effect of finite sample size. Comparing the two types of bootstrap confidence intervals, the percentile interval will not approach the exact confidence interval well because it does not use all information about the distribution of $\hat{\alpha}_{(b)}$ in Equation 5 (Efron & Tibshirani, 1993). The BC_a interval takes the distribution shape of $\hat{\alpha}$ into account through $(\hat{z}_0, \hat{a})$. The BC_a confidence interval for α is comparable to the interval based on the exact distribution of $\hat{\alpha}$ when known. Our recommendation is to use bootstrap and BC_a confidence intervals characterizing the distribution of $\hat{\alpha}$ and the size of α with Likert-type scales.

In this article, we examined the distribution of sample coefficient alpha. When items of a subscale in a sample are not tau-equivalent, the sample coefficient alpha may not be an accurate estimate of the reliability coefficient of the subscale. Under a more general condition of congeneric item-level scores, one may factor analyze the item scores and obtain an improved reliability estimate based on the estimated factor loadings and error variances (Raykov, 1997a). Bootstrap technique can also be used to obtain the standard error and confidence interval of this improved reliability estimate (Raykov, 1998b). Further discussion of coefficient alpha when the classical parallel condition is violated can be found in Green, Lissitz, and Mulaik (1977); McDonald (1981); Cortina (1993); and more recently by Li, Rosenthal, and Rubin (1996); Komaroff (1997); and Raykov (1998a). Hakstian and Barchard (2000) discussed inferential issues for coefficient alpha under sampling of both subjects and conditions.

References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory.* Pacific Grove, CA: Brooks-Cole.

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Bickel, P. J., & Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics*, *9*, 1196-1217.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*, 98-104.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Holt, Rinehart & Winston.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge, UK: Cambridge University Press.

Derogatis, L. R. (1983). *SCL-90-R: Administration, scoring & procedures manual-II* (2nd ed.). Towson, MD: Clinical Psychometric Research.

Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenhuth, E. H., & Covi, L. (1974). The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory. *Behavioral Science*, *19*, 1-15.

Efron, B. (1987). Better bootstrap confidence intervals (with discussion). *Journal of the American Statistical Association*, *82*, 171-200.

Efron, B. (1988). Bootstrap confidence intervals: Good or bad? *Psychological Bulletin*, *104*, 293-296.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Feldt, L. S. (1965). The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. *Psychometrika*, *30*, 357-370.

Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement*, *11*, 93-103.

Ferguson, T. (1996). *A course in large sample theory*. London: Chapman & Hall.

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, *37*, 827-837.

Guarnaccia, C. A., & Hayslip, B., Jr. (1998). Factor structure of the Bereavement Experience Questionnaire: The BEQ-24, a revised short form. *Omega: Journal of Death and Dying*, *37*, 301-314.

Guarnaccia, C. A., Hayslip, B., Jr., &. Landry, L. (1999). Influence of perceived preventability of the death and emotional closeness to the deceased: A test of Bugen's model. *Omega: Journal of Death and Dying*, *39*, 261-276.

Hakstian, A. R., & Barchard, K. A. (2000). Toward more robust inferential procedures for coefficient alpha under sampling of both subjects and conditions. *Multivariate Behavioral Research*, *35*, 427-456.

Hayslip, B., Jr., Ragow-O'Brien, D., & Guarnaccia, C. A. (1999). The relationship of cause of death to attitudes towards funerals and bereavement adjustment. *Omega: Journal of Death and Dying*, *38*, 275-290.

Komaroff, E. (1997). Effect of simultaneous violations of essential tau-equivalence and uncorrelated error on coefficient alpha. *Applied Psychological Measurement*, *21*, 337-348.

Kristof, W. (1963). The statistical theory of stepped-up reliability when a test has been divided into several equivalent parts. *Psychometrika*, *28*, 221-238.

Li, H., Rosenthal, R., & Rubin, D. B. (1996). Reliability of measurement in psychology: From Spearman-Brown to maximal reliability. *Psychological Methods*, *1*, 98-107.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Mammen, E. (1992). *When does bootstrap work? Asymptotic results and simulations. Lecture notes in statistics 77*. New York: Springer-Verlag.

Mardia, K. V. (1970). Measure of multivariate skewness and kurtosis with applications. *Biometrika*, *57*, 519-530.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya Ser. B*, *36*, 115-128.

McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, *34*, 100-117.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Rasmussen, J. L. (1988). "Bootstrap confidence intervals: Good or bad": Comments on Efron (1988) and Strube (1988) and further evaluation. *Psychological Bulletin*, *104*, 297-299.

Raykov, T. (1997a). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, *21*, 173-184.

Raykov, T. (1997b). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research*, *32*, 329-353.

Raykov, T. (1998a). Coefficient alpha and composite reliability with interrelated nonhomogeneous items. *Applied Psychological Measurement*, *22*, 375-385.

Raykov, T. (1998b). A method for obtaining standard errors and confidence intervals of composite reliability for congeneric items. *Applied Psychological Measurement*, *22*, 369-374.

Strube, M. J. (1988). Bootstrap type I error rates for the correlation coefficient: An examination of alternative procedures. *Psychological Bulletin*, *104*, 290-292.

van Zyl, J. M., Neudecker, H., & Nel, D. G. (2000). On the distribution of the maximum likelihood estimator of Cronbach's alpha. *Psychometrika*, *65*, 271-280.

Yuan, K.-H., & Bentler, P. M. (1997). Improving parameter tests in covariance structure analysis. *Computational Statistics and Data Analysis*, *6*, 177-198.

Yuan, K.-H., & Bentler, P. M. (2002). On robustness of the normal-theory based asymptotic distributions of three reliability coefficient estimates. *Psychometrika, 67,* 251-259.

Zimmerman, D. W. (1972). Test reliability and the Kuder-Richardson formulas: Derivation from probability theory. *Educational and Psychological Measurement*, *32*, 939-954.