Is coefficient alpha robust to non-normal data?

Yanyan Sheng[1]* and Zhaohui Sheng[2]

[1] Department of Educational Psychology and Special Education, Southern Illinois University, Carbondale, IL, USA
[2] Department of Educational Leadership, Western Illinois University, Macomb, IL, USA

Coefficient alpha has been a widely used measure by which internal consistency reliability is assessed. In addition to essential tau-equivalence and uncorrelated errors, normality has been noted as another important assumption for alpha. Earlier work on evaluating this assumption considered either exclusively non-normal error score distributions, or limited conditions. In view of this and the availability of advanced methods for generating univariate non-normal data, Monte Carlo simulations were conducted to show that non-normal distributions for true or error scores do create problems for using alpha to estimate the internal consistency reliability. The sample coefficient alpha is affected by leptokurtic true score distributions, or skewed and/or kurtotic error score distributions. Increased sample sizes, not test lengths, help improve the accuracy, bias, or precision of using it with non-normal data.

Keywords: coefficient alpha, true score distribution, error score distribution, non-normality, skew, kurtosis, Monte Carlo, power method polynomials

INTRODUCTION

Coefficient alpha (Guttman, 1945; Cronbach, 1951) has been one of the most commonly used measures today to assess internal consistency reliability despite criticisms of its use (e.g., Raykov, 1998; Green and Hershberger, 2000; Green and Yang, 2009; Sijtsma, 2009). The derivation of the coefficient is based on classical test theory (CTT; Lord and Novick, 1968), which posits that a person's observed score is a linear function of his/her unobserved true score (or underlying construct) and error score. In the theory, measures can be parallel (essential) tau-equivalent, or congeneric, depending on the assumptions on the units of measurement, degrees of precision, and/or error variances. When two tests are designed to measure the same latent construct, they are parallel if they measure it with identical units of measurement, the same precision, and the same amounts of error; tau-equivalent if they measure it with the same units, the same precision, but have possibly different error variance; essentially tau-equivalent if they assess it using the same units, but with possibly different precision and different amounts of error; or congeneric if they assess it with possibly different units of measurement, precision, and amounts of error (Lord and Novick, 1968; Graham, 2006). From parallel to congeneric, tests are requiring less strict assumptions and hence are becoming more general. Studies (Lord and Novick, 1968, pp. 87–91; see also Novick and Lewis, 1967, pp. 6–7) have shown formally that the population coefficient alpha equals internal consistency reliability for tests that are tau-equivalent or at least essential tau-equivalent. It underestimates the actual reliability for the more general congeneric test. Apart from essential tau-equivalence, coefficient alpha requires two additional assumptions: uncorrelated errors (Guttman, 1945; Novick and Lewis, 1967) and normality (e.g., Zumbo, 1999). Over the past decades, studies have well documented the effects of violations of essential tau-equivalence and uncorrelated errors (e.g., Zimmerman et al., 1993; Miller, 1995; Raykov, 1998; Green and Hershberger, 2000; Zumbo and Rupp,

2004; Graham, 2006; Green and Yang, 2009), which have been considered as two major assumptions for alpha. The normality assumption, however, has received little attention. This could be a concern in typical applications where the population coefficient is an unknown parameter and has to be estimated using the sample coefficient. When data are normally distributed, sample coefficient alpha has been shown to be an unbiased estimate of the population coefficient alpha (Kristof, 1963; van Zyl et al., 2000); however, less is known about situations when data are non-normal.

Over the past decades, the effect of departure from normality on the sample coefficient alpha has been evaluated by Bay (1973), Shultz (1993), and Zimmerman et al. (1993) using Monte Carlo simulations. They reached different conclusions on the effect of non-normal data. In particular, Bay (1973) concluded that a leptokurtic true score distribution could cause coefficient alpha to seriously underestimate internal consistency reliability. Zimmerman et al. (1993) and Shultz (1993), on the other hand, found that the sample coefficient alpha was fairly robust to departure from the normality assumption. The three studies differed in the design, in the factors manipulated and in the non-normal distributions considered, but each is limited in certain ways. For example, Zimmerman et al. (1993) and Shultz (1993) only evaluated the effect of non-normal error score distributions. Bay (1973), while looked at the effect of non-normal true score or error score distributions, only studied conditions of 30 subjects and 8 test items. Moreover, these studies have considered only two or three scenarios when it comes to non-normal distributions. Specifically, Bay (1973) employed uniform (symmetric platykurtic) and exponential (non-symmetric leptokurtic with positive skew) distributions for both true and error scores. Zimmerman et al. (1993) generated error scores from uniform, exponential, and mixed normal (symmetric leptokurtic) distributions, while Shultz (1993) generated them using exponential, mixed normal, and negative exponential

(non-symmetric leptokurtic with negative skew) distributions. Since the presence of skew and/or kurtosis determines whether and how a distribution departs from the normal pattern, it is desirable to consider distributions with varying levels of skew and kurtosis so that a set of guidelines can be provided. Generating univariate non-normal data with specified moments can be achieved via the use of power method polynomials (Fleishman, 1978), and its current developments (e.g., Headrick, 2010) make it possible to consider more combinations of skew and kurtosis.

Further, in the actual design of a reliability study, sample size determination is frequently an important and difficult aspect. The literature offers widely different recommendations, ranging from 15 to 20 (Fleiss, 1986), a minimum of 30 (Johanson and Brooks, 2010) to a minimum of 300 (Nunnally and Bernstein, 1994). Although Bay (1973) has used analytical derivations to suggest that coefficient alpha shall be robust against the violation of the normality assumption if sample size is large, or the number of items is large and the true score kurtosis is close to zero, it is never clear how many subjects and/or items are desirable in such situations.

In view of the above, the purpose of this study is to investigate the effect of non-normality (especially the presence of skew and/or kurtosis) on reliability estimation and how sample sizes and test lengths affect the estimation with non-normal data. It is believed that the results will not only shed insights on how non-normality affects coefficient alpha, but also provide a set of guidelines for researchers when specifying the numbers of subjects and items in a reliability study.

MATERIALS AND METHODS

This section starts with a brief review of the CTT model for coefficient alpha. Then the procedures for simulating observed scores used in the Monte Carlo study are described, followed by measures that were used to evaluate the performance of the sample alpha in each simulated situation.

PRELIMINARIES

Coefficient alpha is typically associated with true score theory (Guttman, 1945; Cronbach, 1951; Lord and Novick, 1968), where the test score for person i on item j, denoted as X_{ij}, is assumed to be a linear function of a true score (t_{ij}) and an error score (e_{ij}):

$$X_{ij} = t_{ij} + e_{ij}, \qquad (1)$$

$i = 1, \ldots, n$ and $j = 1, \ldots, k$, where $E(e_{ij}) = 0$, $\rho_{te} = 0$, and $\rho_{e_{ij},e_{ij'}} = 0$. Here, e_{ij} denotes random error that reflects unpredictable trial-by-trial fluctuations. It has to be differentiated from systematic error that reflects situational or individual effects that may be specified. In the theory, items are usually assumed to be tau-equivalent, where true scores are restricted to be the same across items, or essentially tau-equivalent, where they are allowed to differ from item to item by a constant (v_j). Under these conditions (1) becomes

$$X_{ij} = t_i + e_{ij} \qquad (2)$$

for tau-equivalence, and

$$X_{ij} = t_i + v_j + e_{ij}, \qquad (3)$$

where $\Sigma_j v_j = 0$, for essential tau-equivalence.

Summing across k items, we obtain a composite score (X_{i+}) and a scale error score (e_{i+}). The variance of the composite scores is then the summation of true score and scale error score variances:

$$\sigma_{X_+}^2 = \sigma_t^2 + \sigma_{e_+}^2. \qquad (4)$$

The reliability coefficient, $\rho_{XX'}$, is defined as the proportion of composite score variance that is due to true score variance:

$$\rho_{XX'} = \frac{\sigma_t^2}{\sigma_{X_+}^2}. \qquad (5)$$

Under (essential) tau-equivalence, that is, for models in (2) and (3), the population coefficient alpha, defined as

$$\alpha = \frac{k}{k-1} \frac{\sum \sum_{j \neq j'} \sigma_{X_j X_{j'}}}{\sigma_{X_+}^2},$$

or

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{j=1}^k \sigma_j^2}{\sigma_{X_+}^2} \right), \qquad (6)$$

is equal to the reliability as defined in (5). As was noted, $\rho_{XX'}$ and α focus on the amount of random error and do not evaluate error that may be systematic.

Although the derivation of coefficient alpha based on Lord and Novick (1968) does not require distributional assumptions for t_i and e_{ij}, its estimation does (see Shultz, 1993; Zumbo, 1999), as the sample coefficient alpha estimated using sample variances s^2,

$$\hat{\alpha} = \frac{k}{k-1} \left(1 - \frac{\sum_{j=1}^k s_j^2}{s_{X_+}^2} \right), \qquad (7)$$

is shown to be the maximum likelihood estimator of the population alpha assuming normal distributions (Kristof, 1963; van Zyl et al., 2000). Typically, we assume $t_i \sim N(\mu_t, \sigma_t^2)$ and $e_{ij} \sim N(0, \sigma_e^2)$, where σ_e^2 has to be differentiated from the scale error score variance $\sigma_{e_+}^2$ defined in (4).

STUDY DESIGN

To evaluate the performance of the sample alpha as defined in (7) in situations where true score or error score distributions depart from normality, a Monte Carlo simulation study was carried out, where test scores of n persons ($n = 30, 50, 100, 1000$) for k items ($k = 5, 10, 30$) were generated assuming tau-equivalence and where the population reliability coefficient ($\rho_{XX'}$) was specified to be 0.3, 0.6, or 0.8 to correspond to unacceptable, acceptable, or very good reliability (Caplan et al., 1984, p. 306; DeVellis, 1991,

p. 85; Nunnally, 1967, p. 226). These are referred to as small, moderate, and high reliabilities in subsequent discussions. Specifically, true scores (t_i) and error scores (e_{ij}) were simulated from their respective distributions with $\sigma_e^2 = 1$, $\mu_t = 5$ and $\sigma_t^2 = \frac{\sigma_e^2 \rho_{XX'}}{(1 - \rho_{XX'})k}$. The observed scores ($X_{ij}$) were subsequently obtained using Eq. (2).

In addition, true score or error score distributions were manipulated to be symmetric (so that skew, γ_1, is 0) or non-symmetric ($\gamma_1 > 0$) with kurtosis (γ_2) being 0, negative or positive. It is noted that only positively skewed distributions were considered in the study because due to the symmetric property, negative skew should have the same effect as positive skew. Generating non-normal distributions in this study involves the use of power method polynomials. Fleishman (1978) introduced this popular moment matching technique for generating univariate non-normal distributions. Headrick (2002, 2010) further extended from third-order to fifth-order polynomials to lower the skew and kurtosis boundary. As is pointed out by Headrick (2010, p. 26), for distributions with a mean of 0 and a variance of 1, the skew and kurtosis have to satisfy $\gamma_2 \geqslant \gamma_1^2 - 2$, and hence it is not plausible to consider all possible combinations of skew and kurtosis using power method polynomials. Given this, six distributions with the following combinations of skew and kurtosis were considered:

1. $\gamma_1 = 0$, $\gamma_2 = 0$ (normal distribution);
2. $\gamma_1 = 0$, $\gamma_2 = -1.385$ (symmetric platykurtic distribution);
3. $\gamma_1 = 0$, $\gamma_2 = 25$ (symmetric leptokurtic distribution);
4. $\gamma_1 = 0.96$, $\gamma_2 = 0.13$ (non-symmetric distribution);
5. $\gamma_1 = 0.48$, $\gamma_2 = -0.92$ (non-symmetric platykurtic distribution);
6. $\gamma_1 = 2.5$, $\gamma_2 = 25$ (non-symmetric leptokurtic distribution).

A normal distribution was included so that it could be used as a baseline against which the non-normal distributions could be compared. To actually generate univariate distributions using the fifth-order polynomial transformation, a random variate Z is first generated from a standard normal distribution, $Z \sim N(0,1)$. Then the following polynomial,

$$Y = c_0 + c_1 Z + c_2 Z^2 + c_3 Z^3 + c_4 Z^4 + c_5 Z^5 \qquad (8)$$

is used to obtain Y. With appropriate coefficients ($c_0, \ldots, c_5$), Y would follow a distribution with a mean of 0, a variance of 1, and the desired levels of skew and kurtosis (see Headrick, 2002, for a detailed description of the procedure). A subsequent linear transformation would rescale the distribution to have a desired location or scale parameter. In this study, Y could be the true score (t_i) or the error score (e_{ij}). For the six distributions considered for t_i or e_{ij} herein, the corresponding coefficients are:

1. $c_0 = 0$, $c_1 = 1$, $c_2 = 0$, $c_3 = 0$, $c_4 = 0$, $c_5 = 0$;
2. $c_0 = 0$, $c_1 = 1.643377$, $c_2 = 0$, $c_3 = -0.319988$, $c_4 = 0$, $c_5 = 0.011344$;
3. $c_0 = 0$, $c_1 = 0.262543$, $c_2 = 0$, $c_3 = 0.201036$, $c_4 = 0$, $c_5 = 0.000162$;

4. $c_0 = -0.446924$, $c_1 = 1.242521$, $c_2 = 0.500764$, $c_3 = -0.184710$, $c_4 = -0.017947$, $c_5 = 0.003159$;
5. $c_0 = -0.276330$, $c_1 = 1.506715$, $c_2 = 0.311114$, $c_3 = -0.274078$, $c_4 = -0.011595$, $c_5 = 0.007683$;
6. $c_0 = -0.304852$, $c_1 = 0.381063$, $c_2 = 0.356941$, $c_3 = 0.132688$, $c_4 = -0.017363$, $c_5 = 0.003570$.

It is noted that the effect of the true score or error score distribution was investigated independently, holding the other constant by assuming it to be normal.

Hence, a total of 4 (sample sizes) × 3 (test lengths) × 3 (levels of population reliability) × 6 (distributions) × 2 (true or error score) = 432 conditions were considered in the simulation study. Each condition involved 100,000 replications, where coefficient alpha was estimated using Eq. (7) for simulated test scores (X_{ij}). The 100,000 estimates of α can be considered as random samples from the sampling distribution of $\hat{\alpha}$, and its summary statistics including the observed mean, SD, and 95% interval provide information about this distribution. In particular, the observed mean indicates whether the sample coefficient is biased. If it equals α, $\hat{\alpha}$ is unbiased; otherwise, it is biased either positively or negatively depending on whether it is larger or smaller than α. The SD of the sampling distribution is what we usually call the SE. It reflects the uncertainty in estimating α, with a smaller SE suggesting more precision and hence less uncertainty in the estimation. The SE is directly related to the 95% observed interval, as the larger it is, the more spread the distribution is and the wider the interval will be. With respect to the observed interval, it contains about 95% of $\hat{\alpha}$ around its center location from its empirical sampling distribution. If α falls inside the interval, $\hat{\alpha}$ is not significantly different from α even though it is not unbiased. On the other hand, if α falls outside of the interval, which means that 95% of the estimates differ from α, we can consider $\hat{\alpha}$ to be significantly different from α.

In addition to these summary statistics, the accuracy of the estimate was evaluated by the root mean square error (RMSE) and *bias*, which are defined as

$$\text{RMSE} = \sqrt{\frac{\sum (\hat{\alpha} - \alpha)^2}{100,000}}, \qquad (9)$$

and

$$\text{bias} = \frac{\sum (\hat{\alpha} - \alpha)}{100,000}, \qquad (10)$$

respectively. The larger the RMSE is, the less accurate the sample coefficient is in estimating the population coefficient. Similarly, the larger the absolute value of the *bias* is, the more bias the sample coefficient involves. As the equations suggest, RMSE is always positive, with values close to zero reflecting less error in estimating the actual reliability. On the other hand, *bias* can be negative or positive. A positive *bias* suggests that the sample coefficient tends to overestimate the reliability, and a negative *bias* suggests that it tends to underestimate the reliability. In effect, *bias* provides similar information as the observed mean of the sampling distribution of $\hat{\alpha}$.

RESULTS

The simulations were carried out using MATLAB (MathWorks, 2010), with the source code being provided in the Section "Appendix." Simulation results are summarized in **Tables 1–3** for conditions where true scores follow one of the six distributions specified in the previous section. Here, results from the five non-normal distributions were mainly compared with those from the normal

distribution to determine if $\hat{\alpha}$ was affected by non-normality in true scores. Take the condition where a test of 5 items with the actual reliability being 0.3 was given to 30 persons as an example. A normal distribution resulted in an observed mean of 0.230 and a SE of 0.241 for the sampling distribution of $\hat{\alpha}$ (see **Table 1**). Compared with it, a symmetric platykurtic distribution, with an observed mean of 0.234 and a SE of 0.235, did not differ much.

Table 1 | Observed mean and SD of the sample alpha ($\hat{\alpha}$) for the simulated situations where the true score (t_i) distribution is normal or non-normal.

n	k	Mean ($\hat{\alpha}$)						SD ($\hat{\alpha}$)					
		dist1	dist2	dist3	dist4	dist5	dist6	dist1	dist2	dist3	dist4	dist5	dist6
$\rho_{XX'} = 0.3$													
30	5	0.230	0.234	0.198	0.230	0.231	0.201	0.241	0.235	0.290	0.242	0.237	0.288
	10	0.231	0.234	0.199	0.229	0.233	0.202	0.229	0.223	0.278	0.230	0.224	0.276
	30	0.231	0.233	0.199	0.230	0.233	0.200	0.221	0.215	0.269	0.222	0.216	0.270
50	5	0.252	0.253	0.233	0.253	0.254	0.232	0.176	0.172	0.214	0.177	0.172	0.214
	10	0.252	0.256	0.232	0.252	0.254	0.233	0.166	0.161	0.205	0.166	0.162	0.204
	30	0.254	0.254	0.231	0.252	0.254	0.233	0.160	0.156	0.202	0.160	0.157	0.199
100	5	0.269	0.269	0.258	0.268	0.269	0.258	0.118	0.116	0.148	0.119	0.117	0.148
	10	0.268	0.269	0.257	0.269	0.270	0.258	0.112	0.109	0.143	0.112	0.110	0.142
	30	0.269	0.270	0.257	0.268	0.269	0.256	0.108	0.105	0.141	0.108	0.106	0.140
1000	5	0.282	0.282	0.281	0.282	0.282	0.281	0.036	0.035	0.048	0.036	0.035	0.048
	10	0.282	0.282	0.281	0.282	0.282	0.281	0.034	0.033	0.046	0.034	0.033	0.046
	30	0.282	0.282	0.281	0.282	0.282	0.281	0.033	0.032	0.045	0.033	0.032	0.045
$\rho_{XX'} = 0.6$													
30	5	0.549	0.556	0.479	0.549	0.554	0.482	0.142	0.125	0.239	0.142	0.131	0.238
	10	0.551	0.557	0.481	0.549	0.554	0.480	0.133	0.117	0.232	0.136	0.122	0.232
	30	0.550	0.557	0.480	0.550	0.555	0.481	0.129	0.112	0.230	0.131	0.118	0.229
50	5	0.563	0.567	0.517	0.563	0.566	0.517	0.103	0.092	0.179	0.104	0.095	0.180
	10	0.563	0.567	0.516	0.563	0.566	0.517	0.097	0.086	0.176	0.098	0.089	0.174
	30	0.563	0.567	0.516	0.563	0.566	0.518	0.093	0.082	0.174	0.094	0.086	0.172
100	5	0.572	0.574	0.545	0.572	0.573	0.546	0.069	0.062	0.128	0.070	0.065	0.126
	10	0.572	0.574	0.545	0.572	0.573	0.547	0.066	0.057	0.126	0.066	0.060	0.124
	30	0.572	0.574	0.545	0.572	0.573	0.546	0.063	0.055	0.124	0.064	0.058	0.122
1000	5	0.580	0.580	0.576	0.580	0.580	0.577	0.021	0.019	0.043	0.021	0.020	0.042
	10	0.580	0.580	0.577	0.580	0.580	0.576	0.020	0.018	0.042	0.020	0.018	0.042
	30	0.580	0.580	0.576	0.580	0.580	0.576	0.019	0.017	0.042	0.019	0.018	0.041
$\rho_{XX'} = 0.8$													
30	5	0.771	0.778	0.701	0.770	0.776	0.703	0.072	0.056	0.171	0.075	0.062	0.172
	10	0.771	0.778	0.702	0.770	0.776	0.702	0.068	0.052	0.167	0.070	0.057	0.169
	30	0.771	0.778	0.701	0.771	0.776	0.702	0.066	0.049	0.166	0.068	0.055	0.167
50	5	0.778	0.782	0.733	0.778	0.780	0.733	0.052	0.041	0.125	0.053	0.045	0.125
	10	0.778	0.782	0.733	0.778	0.781	0.733	0.049	0.038	0.123	0.050	0.042	0.123
	30	0.778	0.782	0.732	0.778	0.781	0.733	0.048	0.036	0.122	0.049	0.040	0.122
100	5	0.782	0.784	0.757	0.782	0.784	0.757	0.035	0.028	0.086	0.036	0.031	0.085
	10	0.783	0.784	0.757	0.782	0.784	0.757	0.033	0.026	0.085	0.034	0.028	0.084
	30	0.783	0.784	0.757	0.782	0.784	0.757	0.032	0.024	0.084	0.033	0.027	0.084
1000	5	0.786	0.787	0.783	0.786	0.787	0.783	0.011	0.009	0.028	0.011	0.009	0.027
	10	0.786	0.787	0.783	0.787	0.787	0.783	0.010	0.008	0.028	0.010	0.009	0.027
	30	0.787	0.787	0.783	0.786	0.787	0.784	0.010	0.007	0.027	0.010	0.008	0.027

dist1, Normal distribution for t_i; dist2, distribution with negative kurtosis for t_i; dist3, distribution with positive kurtosis for t_i; dist4, skewed distribution for t_i; dist5, skewed distribution with negative kurtosis for t_i; dist6, skewed distribution with positive kurtosis for t_i.

Table 2 | Root mean square error and bias for estimating α for the simulated situations where the true score (t_i) distribution is normal or non-normal.

n	k	RMSE						bias					
		dist1	dist2	dist3	dist4	dist5	dist6	dist1	dist2	dist3	dist4	dist5	dist6
$\rho_{XX'} = 0.3$													
30	5	0.251	0.244	0.308	0.252	0.247	0.305	−0.070	−0.066	−0.102	−0.070	−0.069	−0.100
	10	0.240	0.233	0.296	0.241	0.234	0.292	−0.069	−0.067	−0.101	−0.071	−0.067	−0.098
	30	0.232	0.226	0.287	0.232	0.226	0.288	−0.069	−0.067	−0.101	−0.070	−0.067	−0.101
50	5	0.182	0.178	0.224	0.183	0.178	0.224	−0.048	−0.047	−0.067	−0.047	−0.046	−0.068
	10	0.173	0.167	0.216	0.173	0.169	0.215	−0.048	−0.044	−0.068	−0.048	−0.046	−0.067
	30	0.166	0.162	0.213	0.167	0.164	0.210	−0.046	−0.046	−0.069	−0.048	−0.046	−0.067
100	5	0.122	0.120	0.154	0.123	0.121	0.154	−0.031	−0.031	−0.042	−0.032	−0.031	−0.042
	10	0.116	0.114	0.149	0.116	0.114	0.148	−0.032	−0.031	−0.043	−0.031	−0.031	−0.042
	30	0.112	0.109	0.147	0.113	0.110	0.147	−0.031	−0.030	−0.043	−0.032	−0.031	−0.044
1000	5	0.040	0.040	0.052	0.041	0.040	0.051	−0.018	−0.018	−0.019	−0.018	−0.018	−0.019
	10	0.038	0.038	0.050	0.039	0.038	0.050	−0.018	−0.018	−0.019	−0.018	−0.018	−0.020
	30	0.038	0.037	0.049	0.038	0.037	0.049	−0.018	−0.018	−0.019	−0.018	−0.018	−0.019
$\rho_{XX'} = 0.6$													
30	5	0.151	0.132	0.268	0.151	0.139	0.266	−0.051	−0.044	−0.121	−0.051	−0.046	−0.118
	10	0.142	0.125	0.261	0.145	0.131	0.261	−0.050	−0.043	−0.120	−0.051	−0.046	−0.120
	30	0.139	0.120	0.260	0.140	0.126	0.258	−0.050	−0.043	−0.120	−0.051	−0.045	−0.119
50	5	0.109	0.097	0.198	0.110	0.101	0.198	−0.037	−0.033	−0.083	−0.037	−0.035	−0.083
	10	0.104	0.092	0.195	0.105	0.096	0.193	−0.037	−0.033	−0.084	−0.037	−0.034	−0.083
	30	0.100	0.088	0.193	0.102	0.092	0.191	−0.037	−0.033	−0.084	−0.037	−0.034	−0.083
100	5	0.075	0.067	0.139	0.076	0.070	0.137	−0.028	−0.026	−0.055	−0.028	−0.027	−0.054
	10	0.071	0.063	0.137	0.072	0.066	0.135	−0.028	−0.026	−0.056	−0.028	−0.027	−0.053
	30	0.069	0.061	0.135	0.070	0.064	0.133	−0.028	−0.026	−0.055	−0.028	−0.027	−0.054
1000	5	0.029	0.028	0.049	0.029	0.028	0.049	−0.020	−0.020	−0.024	−0.020	−0.020	−0.024
	10	0.028	0.027	0.048	0.029	0.027	0.048	−0.020	−0.020	−0.023	−0.020	−0.020	−0.024
	30	0.028	0.026	0.048	0.028	0.027	0.048	−0.020	−0.020	−0.024	−0.020	−0.020	−0.024
$\rho_{XX'} = 0.8$													
30	5	0.078	0.060	0.197	0.080	0.066	0.198	−0.030	−0.022	−0.099	−0.030	−0.024	−0.097
	10	0.074	0.056	0.194	0.076	0.062	0.196	−0.029	−0.023	−0.098	−0.030	−0.024	−0.099
	30	0.072	0.053	0.193	0.074	0.060	0.193	−0.029	−0.022	−0.099	−0.029	−0.024	−0.098
50	5	0.057	0.045	0.142	0.058	0.049	0.142	−0.022	−0.018	−0.067	−0.023	−0.020	−0.067
	10	0.054	0.042	0.140	0.055	0.046	0.140	−0.022	−0.018	−0.067	−0.022	−0.020	−0.067
	30	0.052	0.040	0.140	0.054	0.044	0.139	−0.022	−0.018	−0.068	−0.022	−0.019	−0.067
100	5	0.039	0.032	0.096	0.040	0.035	0.095	−0.018	−0.016	−0.043	−0.018	−0.016	−0.043
	10	0.038	0.030	0.095	0.038	0.033	0.094	−0.018	−0.016	−0.043	−0.018	−0.016	−0.043
	30	0.037	0.029	0.094	0.037	0.032	0.094	−0.018	−0.016	−0.043	−0.018	−0.016	−0.043
1000	5	0.017	0.016	0.033	0.017	0.016	0.032	−0.014	−0.013	−0.017	−0.014	−0.013	−0.017
	10	0.017	0.016	0.032	0.017	0.016	0.032	−0.014	−0.013	−0.017	−0.014	−0.013	−0.017
	30	0.017	0.015	0.032	0.017	0.016	0.032	−0.014	−0.013	−0.017	−0.014	−0.013	−0.017

dist1, Normal distribution for t_i; dist2, distribution with negative kurtosis for t_i; dist3, distribution with positive kurtosis for t_i; dist4, skewed distribution for t_i; dist5, skewed distribution with negative kurtosis for t_i; dist6, skewed distribution with positive kurtosis for t_i.

On the other hand, a symmetric leptokurtic distribution resulted in a much smaller mean (0.198) and a larger SE (0.290), indicating that the center location of the sampling distribution of $\hat{\alpha}$ was further away from the actual value (0.3) and more uncertainty was involved in estimating α. With respect to the accuracy of the estimate, **Table 2** shows that the normal distribution had a RMSE of 0.251 and a *bias* value of −0.070. The platykurtic distribution gave rise to smaller but very similar values: 0.244 for RMSE and

−0.066 for *bias*, whereas the leptokurtic distribution had a relatively larger RMSE value (0.308) and a smaller *bias* value (−0.102), indicating that it involved more error and negative bias in estimating α. Hence, under this condition, positive kurtosis affected (the location and scale of) the sampling distribution of $\hat{\alpha}$ as well as the accuracy of using it to estimate α whereas negative kurtosis did not. Similar interpretations are used for the 95% interval shown in **Table 3**, except that one can also use the intervals to determine

Table 3 | Observed 95% interval of the sample alpha ($\hat{\alpha}$) for the simulated situations where the true score (t_i) distribution is normal or non-normal.

n	k	dist1		dist2		dist3		dist4		dist5		dist6	
		LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
$\rho_{XX'} = 0.3$													
30	5	−0.351	0.580	−0.329	0.577	−0.490	0.635	−0.356	0.580	−0.342	0.576	−0.481	0.637
	10	−0.323	0.563	−0.305	0.556	−0.457	0.630	−0.328	0.561	−0.308	0.558	−0.450	0.624
	30	−0.303	0.550	−0.285	0.545	−0.435	0.618	−0.303	0.551	−0.286	0.547	−0.435	0.616
50	5	−0.155	0.528	−0.143	0.524	−0.252	0.587	−0.155	0.529	−0.147	0.524	−0.255	0.583
	10	−0.136	0.512	−0.115	0.508	−0.233	0.576	−0.134	0.514	−0.123	0.510	−0.229	0.573
	30	−0.116	0.505	−0.106	0.500	−0.219	0.571	−0.119	0.504	−0.109	0.501	−0.216	0.568
100	5	0.005	0.469	0.013	0.465	−0.062	0.522	0.004	0.469	0.010	0.466	−0.062	0.521
	10	0.020	0.457	0.027	0.454	−0.050	0.515	0.021	0.458	0.025	0.455	−0.046	0.512
	30	0.030	0.452	0.039	0.447	−0.044	0.512	0.028	0.451	0.035	0.447	−0.040	0.508
1000	5	0.208	0.350	0.211	0.349	0.186	0.374	0.208	0.350	0.210	0.348	0.186	0.373
	10	0.213	0.346	0.215	0.344	0.189	0.371	0.213	0.346	0.214	0.345	0.190	0.371
	30	0.215	0.343	0.217	0.342	0.192	0.369	0.215	0.344	0.217	0.343	0.192	0.369
$\rho_{XX'} = 0.6$													
30	5	0.212	0.754	0.258	0.742	−0.088	0.836	0.206	0.754	0.239	0.746	−0.086	0.834
	10	0.231	0.743	0.277	0.730	−0.067	0.833	0.219	0.744	0.261	0.734	−0.071	0.832
	30	0.239	0.737	0.289	0.723	−0.063	0.831	0.235	0.737	0.273	0.727	−0.059	0.828
50	5	0.325	0.723	0.357	0.713	0.111	0.809	0.322	0.724	0.348	0.718	0.105	0.807
	10	0.338	0.716	0.371	0.704	0.118	0.807	0.335	0.716	0.358	0.707	0.122	0.801
	30	0.349	0.711	0.377	0.697	0.127	0.806	0.343	0.710	0.370	0.702	0.130	0.801
100	5	0.417	0.689	0.439	0.680	0.270	0.770	0.416	0.689	0.430	0.683	0.274	0.768
	10	0.426	0.682	0.448	0.672	0.277	0.768	0.426	0.684	0.440	0.676	0.280	0.768
	30	0.432	0.678	0.452	0.668	0.283	0.767	0.430	0.679	0.446	0.672	0.286	0.764
1000	5	0.537	0.619	0.541	0.616	0.492	0.660	0.537	0.620	0.539	0.617	0.493	0.659
	10	0.539	0.617	0.544	0.613	0.494	0.660	0.539	0.617	0.543	0.615	0.495	0.658
	30	0.541	0.616	0.546	0.612	0.494	0.658	0.540	0.616	0.544	0.613	0.495	0.657
$\rho_{XX'} = 0.8$													
30	5	0.596	0.875	0.646	0.864	0.281	0.930	0.590	0.875	0.630	0.868	0.274	0.928
	10	0.607	0.869	0.655	0.857	0.292	0.929	0.598	0.869	0.641	0.861	0.283	0.926
	30	0.612	0.866	0.663	0.852	0.300	0.927	0.604	0.867	0.645	0.858	0.291	0.926
50	5	0.656	0.860	0.688	0.849	0.436	0.917	0.653	0.860	0.677	0.853	0.433	0.914
	10	0.664	0.855	0.696	0.844	0.444	0.916	0.660	0.856	0.686	0.848	0.439	0.913
	30	0.667	0.853	0.700	0.840	0.444	0.915	0.664	0.853	0.690	0.845	0.443	0.913
100	5	0.704	0.842	0.723	0.833	0.562	0.896	0.703	0.842	0.717	0.837	0.564	0.896
	10	0.708	0.838	0.728	0.829	0.567	0.896	0.706	0.840	0.722	0.833	0.568	0.895
	30	0.711	0.836	0.731	0.827	0.569	0.896	0.710	0.837	0.725	0.830	0.568	0.894
1000	5	0.764	0.807	0.769	0.803	0.726	0.837	0.764	0.807	0.768	0.804	0.728	0.836
	10	0.766	0.805	0.771	0.802	0.728	0.836	0.766	0.806	0.769	0.803	0.728	0.836
	30	0.766	0.805	0.772	0.801	0.728	0.836	0.766	0.805	0.770	0.802	0.729	0.836

dist1, Normal distribution for t_i; dist2, distribution with negative kurtosis for t_i; dist3, distribution with positive kurtosis for t_i; dist4, skewed distribution for t_i; dist5, skewed distribution with negative kurtosis for t_i; dist6, skewed distribution with positive kurtosis for t_i; LB, lower bound; UB, upper bound.

whether the sample coefficient was significantly different from α as described in the previous section.

Guided by these interpretations, one can make the following observations:

1. Among the five non-normal distributions considered for t_i, skewed or platykurtic distributions do not affect the mean or the SE for $\hat{\alpha}$ (see **Table 1**). They do not affect the accuracy

or bias in estimating α, either (see **Table 2**). On the other hand, symmetric or non-symmetric distributions with positive kurtosis tend to result in a much smaller average of $\hat{\alpha}$ with a larger SE (see **Table 1**), which in turn makes the 95% observed interval wider compared with the normal distribution (see **Table 3**). In addition, positive kurtosis tends to involve more bias in underestimating α with a reduced accuracy (see **Table 2**).

2. Sample size (n) and test length (k) play important roles for $\hat{\alpha}$ and its sampling distribution, as increased n or k tends to result in the mean of $\hat{\alpha}$ that is closer to the specified population reliability ($\rho_{XX'}$) with a smaller SE. We note that n has a larger and more apparent effect than k. Sample size further helps offset the effect of non-normality on the sampling distribution of $\hat{\alpha}$. In particular, when sample size gets large, e.g., $n = 1000$, departure from normal distributions (due to positive kurtosis) does not result in much different mean of $\hat{\alpha}$ although the SE is still slightly larger compared with normal distributions (see **Table 1**).

3. Increased n or k tends to increase the accuracy in estimating α while reducing bias. However, the effect of non-normality (due to positive kurtosis) on resulting in a larger estimating error and bias remains even with increased n and/or k (see **Table 2**). It is also noted that for all the conditions considered, $\hat{\alpha}$ has a consistently negative bias regardless of the shape of the distribution for t_i.

4. The 95% observed interval shown in **Table 3** agrees with the corresponding mean and SE shown in **Table 1**. It is noted that regardless of the population distribution for t_i, when n or k gets larger, $\hat{\alpha}$ has a smaller SE, and hence a narrower 95% interval, as the precision in estimating α increases. Given this, and that all intervals in the table, especially those for $n = 1000$, cover the specified population reliability ($\rho_{XX'}$), one should note that although departure from normality affects the accuracy, bias, and precision in estimating α, it does not result in systematically different $\hat{\alpha}$. In addition, when the actual reliability is small (i.e., $\rho_{XX'} = 0.3$), the use of large n is suggested, as when $n < 1000$, the 95% interval covers negative values of $\hat{\alpha}$. This is especially the case for the (symmetric or non-symmetric) distributions with positive kurtosis. For these distributions, at least 100 subjects are needed for $\hat{\alpha}$ to avoid relatively large estimation error when the actual reliability is moderate to large. For the other distributions, including the normal distribution, a minimum of 50 subjects is suggested for tests with a moderate reliability (i.e., $\rho_{XX'} = 0.6$), and 30 or more subjects are needed for tests with a high reliability (i.e., $\rho_{XX'} = 0.8$; see **Table 2**).

In addition, results for conditions where error scores depart from normal distributions are summarized in **Tables 4–6**. Given the design of the study, the results for the condition where e_{ij} followed a normal distribution are the same as those for the condition where the distribution for t_i was normal. For the purpose of comparisons, they are displayed in the tables again. Inspections of these tables result in the following findings, some of which are quite different from what are observed from **Tables 1–3**:

1. Symmetric platykurtic distributions or non-symmetric leptokurtic distributions consistently resulted in a larger mean but not a larger SE of $\hat{\alpha}$ than normal distributions (see **Table 4**). Some of the means, and especially those for non-symmetric leptokurtic distributions, are larger than the specified population reliability ($\rho_{XX'}$). This is consistent with the positive bias values in **Table 5**. On the other hand, symmetric leptokurtic, non-symmetric, or non-symmetric platykurtic distributions tend to have larger SE of $\hat{\alpha}$ than the normal distribution (see **Table 4**).

2. Sample size (n) and test length (k) have different effects on $\hat{\alpha}$ and its sampling distribution. Increased n consistently results in a larger mean of $\hat{\alpha}$ with a reduced SE. However, increased k may result in a reduced SE, but it has a negative effect on the mean in pushing it away from the specified population reliability ($\rho_{XX'}$), especially when $\rho_{XX'}$ is not large. In particular, with larger k, the mean of $\hat{\alpha}$ decreases to be much smaller for the non-normal distributions that are leptokurtic, non-symmetric, or non-symmetric platykurtic; but it increases to exceed $\rho_{XX'}$ for symmetric platykurtic or non-symmetric leptokurtic distributions. It is further observed that with increased n, the difference between non-normal and normal distributions of e_{ij} on the mean and SE of $\hat{\alpha}$ reduces. This is, however, not observed for increased k (see **Table 4**).

3. The RMSE and *bias* values presented in **Table 5** indicate that non-normal distributions for e_{ij}, especially leptokurtic, non-symmetric, or non-symmetric platykurtic distributions tend to involve larger error, if not bias, in estimating α. In addition, when k increases, RMSE or *bias* does not necessarily reduce. On the other hand, when n increases, RMSE decreases while *bias* increases. Hence, with larger sample sizes, there is more accuracy in estimating α, but bias is not necessarily reduced for symmetric platykurtic or non-symmetric leptokurtic distributions, as some of the negative bias values increase to become positive and non-negligible.

4. The effect of test length on the sample coefficient is more apparent in **Table 6**. From the 95% observed intervals for $\hat{\alpha}$, and particularly those obtained when the actual reliability is small to moderate (i.e., $\rho_{XX'} \leq 0.6$) with large sample sizes (i.e., $n = 1000$), one can see that when test length gets larger (e.g., $k = 30$), the intervals start to fail to cover the specified population reliability ($\rho_{XX'}$) regardless of the degree of the departure from the normality for e_{ij}. Given the fact that larger sample sizes result in less dispersion (i.e., smaller SE) in the sampling distribution of $\hat{\alpha}$ and hence a narrower 95% interval, and the fact that increased k pushes the mean of $\hat{\alpha}$ away from the specified reliability, this finding suggests that larger k amplifies the effect of non-normality of e_{ij} on $\hat{\alpha}$ in resulting in systematically biased estimates of α, and hence has to be avoided when the actual reliability is not large. With respect to sample sizes, similar patterns arise. That is, the use of large n is suggested when the actual reliability is small (i.e., $\rho_{XX'} = 0.3$), especially for tests with 30 items, whereas for tests with a high reliability (i.e., $\rho_{XX'} = 0.8$), a sample size of 30 may be sufficient. In addition, when the actual reliability is moderate, a minimum of 50 subjects is needed for $\hat{\alpha}$ to be fairly accurate for short tests ($k \leq 10$), and at least 100 are suggested for longer tests ($k = 30$; see **Table 5**).

Given the above results, we see that non-normal distributions for true or error scores do create problems for using coefficient alpha to estimate the internal consistency reliability. In particular, leptokurtic true score distributions that are either symmetric or skewed result in larger error and negative bias in estimating population α with less precision. This is similar to Bay's (1973) finding, and we see in this study that the problem remains even after increasing sample size to 1000 or test length to 30, although the effect is getting smaller. With respect to error score

Table 4 | Observed mean and SD of the sample alpha ($\hat{\alpha}$) for the simulated situations where the error score (e_{ij}) distribution is normal or non−normal.

n	k	Mean ($\hat{\alpha}$)						SD ($\hat{\alpha}$)					
		dist1	dist2	dist3	dist4	dist5	dist6	dist1	dist2	dist3	dist4	dist5	dist6
$\rho_{XX'} = 0.3$													
30	5	0.230	0.255	0.215	0.206	0.213	0.313	0.241	0.233	0.257	0.252	0.250	0.223
	10	0.231	0.295	0.158	0.174	0.185	0.367	0.229	0.207	0.256	0.248	0.248	0.195
	30	0.231	0.371	0.103	0.155	0.139	0.460	0.221	0.177	0.258	0.244	0.249	0.160
50	5	0.252	0.279	0.232	0.231	0.237	0.324	0.176	0.169	0.191	0.183	0.181	0.166
	10	0.252	0.316	0.180	0.198	0.213	0.380	0.166	0.150	0.187	0.180	0.178	0.143
	30	0.254	0.390	0.128	0.181	0.164	0.474	0.160	0.128	0.188	0.176	0.181	0.116
100	5	0.269	0.295	0.245	0.247	0.255	0.332	0.118	0.113	0.130	0.124	0.122	0.115
	10	0.268	0.331	0.196	0.216	0.229	0.390	0.112	0.101	0.128	0.121	0.120	0.098
	30	0.269	0.403	0.146	0.198	0.182	0.484	0.108	0.086	0.127	0.119	0.122	0.078
1000	5	0.282	0.308	0.254	0.261	0.269	0.338	0.036	0.035	0.040	0.038	0.037	0.036
	10	0.282	0.343	0.208	0.231	0.244	0.398	0.034	0.031	0.039	0.037	0.037	0.030
	30	0.282	0.414	0.161	0.213	0.197	0.493	0.033	0.026	0.039	0.036	0.037	0.024
$\rho_{XX'} = 0.6$													
30	5	0.549	0.550	0.565	0.552	0.551	0.571	0.142	0.140	0.163	0.141	0.140	0.159
	10	0.551	0.560	0.547	0.543	0.545	0.586	0.133	0.127	0.157	0.141	0.139	0.132
	30	0.550	0.615	0.452	0.482	0.500	0.669	0.129	0.102	0.180	0.160	0.156	0.093
50	5	0.563	0.564	0.574	0.565	0.564	0.579	0.103	0.100	0.121	0.103	0.101	0.118
	10	0.563	0.573	0.559	0.557	0.560	0.595	0.097	0.092	0.115	0.102	0.100	0.099
	30	0.563	0.625	0.472	0.499	0.518	0.676	0.093	0.074	0.131	0.115	0.112	0.068
100	5	0.572	0.573	0.579	0.574	0.574	0.584	0.069	0.068	0.084	0.069	0.068	0.082
	10	0.572	0.582	0.567	0.567	0.570	0.600	0.066	0.062	0.078	0.069	0.067	0.068
	30	0.572	0.633	0.484	0.511	0.530	0.681	0.063	0.050	0.088	0.078	0.075	0.046
1000	5	0.580	0.581	0.583	0.582	0.582	0.588	0.021	0.021	0.026	0.021	0.021	0.026
	10	0.580	0.589	0.574	0.576	0.578	0.605	0.020	0.019	0.024	0.021	0.020	0.021
	30	0.580	0.638	0.496	0.522	0.540	0.686	0.019	0.015	0.027	0.024	0.023	0.014
$\rho_{XX'} = 0.8$													
30	5	0.771	0.771	0.777	0.772	0.772	0.779	0.072	0.070	0.094	0.072	0.070	0.092
	10	0.771	0.771	0.776	0.773	0.772	0.777	0.068	0.067	0.081	0.068	0.067	0.080
	30	0.771	0.782	0.760	0.763	0.766	0.798	0.066	0.058	0.085	0.076	0.073	0.057
50	5	0.778	0.778	0.782	0.779	0.779	0.783	0.052	0.051	0.070	0.052	0.051	0.070
	10	0.778	0.778	0.781	0.779	0.779	0.784	0.049	0.048	0.060	0.049	0.048	0.059
	30	0.778	0.788	0.768	0.771	0.774	0.802	0.048	0.042	0.060	0.054	0.052	0.042
100	5	0.782	0.783	0.785	0.783	0.783	0.787	0.035	0.034	0.049	0.035	0.034	0.049
	10	0.783	0.783	0.785	0.784	0.784	0.787	0.033	0.033	0.041	0.033	0.033	0.041
	30	0.783	0.792	0.774	0.777	0.779	0.805	0.032	0.028	0.040	0.036	0.034	0.029
1000	5	0.786	0.787	0.788	0.787	0.787	0.789	0.011	0.010	0.015	0.011	0.011	0.015
	10	0.786	0.787	0.788	0.788	0.788	0.791	0.010	0.010	0.013	0.010	0.010	0.013
	30	0.787	0.795	0.779	0.781	0.784	0.807	0.010	0.009	0.012	0.011	0.010	0.009

dist1, Normal distribution for e_{ij}; dist2, distribution with negative kurtosis for e_{ij}; dist3, distribution with positive kurtosis for e_{ij}; dist4, skewed distribution for e_{ij}; dist5, skewed distribution with negative kurtosis for e_{ij}; dist6, skewed distribution with positive kurtosis for e_{ij}.

distributions, unlike conclusions from previous studies, departure from normality does create problems in the sample coefficient alpha and its sampling distribution. Specifically, leptokurtic, skewed, or non-symmetric platykurtic error score distributions tend to result in larger error and negative bias in estimating population α with less precision, whereas platykurtic or non-symmetric leptokurtic error score distributions tend to have increased positive bias when sample size, test length, and/or the actual reliability increases. In addition, different from conclusions made by Bay

(1973) and Shultz (1993), an increase in test length does have an effect on the accuracy and bias in estimating reliability with the sample coefficient alpha when error scores are not normal, but it is in an undesirable manner. In particular, as is noted earlier, increased test length pushes the mean of $\hat{\alpha}$ away from the actual reliability, and hence causes the sample coefficient alpha to be significantly different from the population coefficient when the actual reliability is not high (e.g., $\rho_{XX'} \leq 0.6$) and the sample size is large (e.g., $n = 1000$). This could be due to the fact that e_{ij} is involved

Table 5 | Root mean square error and bias for estimating α for the simulated situations where the error score (e_{ij}) distribution is normal or non−normal.

n	k	RMSE						bias					
		dist1	dist2	dist3	dist4	dist5	dist6	dist1	dist2	dist3	dist4	dist5	dist6
$\rho_{XX'} = 0.3$													
30	5	0.251	0.237	0.271	0.269	0.264	0.224	−0.070	−0.045	−0.085	−0.094	−0.087	0.013
	10	0.240	0.208	0.293	0.278	0.273	0.206	−0.069	−0.005	−0.142	−0.126	−0.115	0.067
	30	0.232	0.191	0.325	0.284	0.297	0.226	−0.069	0.071	−0.197	−0.145	−0.162	0.160
50	5	0.182	0.170	0.202	0.195	0.192	0.167	−0.048	−0.022	−0.068	−0.069	−0.063	0.024
	10	0.173	0.151	0.222	0.207	0.198	0.164	−0.048	0.016	−0.120	−0.103	−0.088	0.080
	30	0.166	0.157	0.255	0.212	0.226	0.209	−0.046	0.090	−0.172	−0.119	−0.136	0.174
100	5	0.122	0.113	0.141	0.135	0.130	0.119	−0.031	−0.006	−0.055	−0.053	−0.045	0.032
	10	0.116	0.106	0.165	0.148	0.140	0.133	−0.032	0.031	−0.105	−0.084	−0.071	0.090
	30	0.112	0.134	0.200	0.157	0.170	0.200	−0.031	0.103	−0.154	−0.102	−0.118	0.184
1000	5	0.040	0.035	0.061	0.054	0.048	0.052	−0.018	0.008	−0.046	−0.039	−0.031	0.038
	10	0.038	0.053	0.100	0.079	0.067	0.102	−0.018	0.043	−0.092	−0.070	−0.056	0.098
	30	0.038	0.117	0.144	0.095	0.109	0.194	−0.018	0.114	−0.139	−0.087	−0.103	0.193
$\rho_{XX'} = 0.6$													
30	5	0.151	0.148	0.166	0.149	0.149	0.161	−0.051	−0.050	−0.035	−0.048	−0.050	−0.029
	10	0.142	0.133	0.165	0.152	0.149	0.133	−0.050	−0.040	−0.053	−0.057	−0.055	−0.014
	30	0.139	0.103	0.233	0.199	0.185	0.116	−0.050	0.015	−0.148	−0.118	−0.100	0.069
50	5	0.109	0.106	0.124	0.108	0.107	0.120	−0.037	−0.036	−0.026	−0.035	−0.036	−0.021
	10	0.104	0.096	0.123	0.111	0.107	0.099	−0.037	−0.027	−0.041	−0.043	−0.040	−0.005
	30	0.100	0.078	0.183	0.153	0.139	0.102	−0.037	0.025	−0.128	−0.101	−0.082	0.076
100	5	0.075	0.073	0.087	0.074	0.073	0.084	−0.028	−0.027	−0.022	−0.026	−0.026	−0.016
	10	0.071	0.065	0.085	0.076	0.074	0.068	−0.028	−0.018	−0.033	−0.033	−0.031	0.000
	30	0.069	0.060	0.146	0.118	0.103	0.093	−0.028	0.033	−0.116	−0.089	−0.070	0.081
1000	5	0.029	0.028	0.032	0.028	0.028	0.029	−0.020	−0.019	−0.017	−0.018	−0.018	−0.012
	10	0.028	0.022	0.036	0.032	0.030	0.022	−0.020	−0.011	−0.026	−0.024	−0.022	0.005
	30	0.028	0.041	0.108	0.082	0.064	0.087	−0.020	0.038	−0.104	−0.078	−0.060	0.086
$\rho_{XX'} = 0.8$													
30	5	0.078	0.076	0.097	0.077	0.076	0.095	−0.030	−0.029	−0.023	−0.028	−0.029	−0.021
	10	0.074	0.073	0.085	0.073	0.073	0.084	−0.029	−0.029	−0.024	−0.027	−0.028	−0.023
	30	0.072	0.060	0.094	0.085	0.080	0.057	−0.029	−0.018	−0.040	−0.037	−0.034	−0.003
50	5	0.057	0.055	0.073	0.057	0.055	0.072	−0.022	−0.022	−0.018	−0.021	−0.021	−0.017
	10	0.054	0.053	0.063	0.053	0.053	0.062	−0.022	−0.022	−0.019	−0.021	−0.021	−0.016
	30	0.052	0.044	0.068	0.061	0.058	0.042	−0.022	−0.012	−0.032	−0.029	−0.026	0.002
100	5	0.039	0.038	0.051	0.039	0.038	0.050	−0.018	−0.017	−0.015	−0.017	−0.017	−0.013
	10	0.038	0.037	0.044	0.037	0.037	0.043	−0.018	−0.017	−0.015	−0.016	−0.016	−0.013
	30	0.037	0.030	0.048	0.043	0.040	0.029	−0.018	−0.008	−0.026	−0.024	−0.021	0.005
1000	5	0.017	0.017	0.020	0.017	0.017	0.019	−0.014	−0.013	−0.012	−0.013	−0.013	−0.011
	10	0.017	0.016	0.017	0.016	0.016	0.016	−0.014	−0.013	−0.012	−0.012	−0.012	−0.010
	30	0.017	0.010	0.024	0.022	0.019	0.012	−0.014	−0.005	−0.021	−0.019	−0.016	0.007

dist1, Normal distribution for e_{ij}; dist2, distribution with negative kurtosis for e_{ij}; dist3, distribution with positive kurtosis for e_{ij}; dist4, skewed distribution for e_{ij}; dist5, skewed distribution with negative kurtosis for e_{ij}; dist6, skewed distribution with positive kurtosis for e_{ij}.

in each item, and hence an increase in the number of items would add up the effect of non-normality on the sample coefficient.

DISCUSSION

In practice, coefficient alpha is often used to estimate reliability with little consideration of the assumptions required for the sample coefficient to be accurate. As noted by Graham (2006, p. 942), students and researchers in education and psychology are often unaware of many assumptions for a statistical procedure,

and this situation is much worse when it comes to measurement issues such as reliability. In actual applications, it is vital to not only evaluate the assumptions for coefficient alpha, but also understand them and the consequences of any violations.

Normality is not commonly considered as a major assumption for coefficient alpha and hence has not been well investigated. This study takes the advantage of recently developed techniques in generating univariate non-normal data to suggest that different from conclusions made by Bay (1973), Zimmerman et al. (1993),

Table 6 | Observed 95% interval of the sample alpha ($\hat{\alpha}$) for the simulated situations where the error score (e_{ij}) distribution is normal or non-normal.

n	k	dist1		dist2		dist3		dist4		dist5		dist6	
		LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
$\rho_{XX'} = 0.3$													
30	5	−0.351	0.580	−0.305	0.592	−0.404	0.587	−0.403	0.572	−0.392	0.574	−0.220	0.638
	10	−0.323	0.563	−0.203	0.595	−0.459	0.529	−0.421	0.536	−0.416	0.543	−0.104	0.647
	30	−0.303	0.550	−0.058	0.629	−0.522	0.478	−0.437	0.508	−0.462	0.501	0.070	0.688
50	5	−0.155	0.528	−0.114	0.544	−0.212	0.531	−0.193	0.516	−0.179	0.519	−0.057	0.585
	10	−0.136	0.512	−0.031	0.551	−0.256	0.475	−0.223	0.481	−0.199	0.493	0.050	0.604
	30	−0.116	0.505	0.094	0.591	−0.307	0.423	−0.229	0.458	−0.257	0.447	0.203	0.653
100	5	0.005	0.469	0.042	0.486	−0.045	0.464	−0.031	0.456	−0.016	0.461	0.077	0.525
	10	0.020	0.457	0.107	0.501	−0.087	0.411	−0.052	0.421	−0.039	0.431	0.172	0.554
	30	0.030	0.452	0.211	0.549	−0.136	0.361	−0.067	0.399	−0.089	0.387	0.310	0.616
1000	5	0.208	0.350	0.237	0.372	0.172	0.330	0.184	0.332	0.193	0.338	0.264	0.405
	10	0.213	0.346	0.280	0.400	0.128	0.282	0.155	0.300	0.170	0.313	0.336	0.454
	30	0.215	0.343	0.360	0.463	0.082	0.234	0.139	0.280	0.122	0.267	0.444	0.537
$\rho_{XX'} = 0.6$													
30	5	0.212	0.754	0.212	0.752	0.171	0.794	0.212	0.756	0.210	0.752	0.186	0.796
	10	0.231	0.743	0.253	0.745	0.167	0.768	0.197	0.744	0.210	0.743	0.267	0.778
	30	0.239	0.737	0.370	0.766	0.015	0.711	0.099	0.713	0.121	0.721	0.444	0.803
50	5	0.325	0.723	0.331	0.722	0.289	0.758	0.326	0.725	0.329	0.722	0.302	0.762
	10	0.338	0.716	0.360	0.717	0.286	0.736	0.319	0.715	0.328	0.714	0.363	0.749
	30	0.349	0.711	0.455	0.743	0.167	0.675	0.229	0.680	0.256	0.691	0.520	0.783
100	5	0.417	0.689	0.422	0.688	0.389	0.716	0.420	0.691	0.422	0.689	0.398	0.721
	10	0.426	0.682	0.444	0.687	0.392	0.697	0.415	0.682	0.420	0.682	0.448	0.714
	30	0.432	0.678	0.521	0.718	0.287	0.632	0.338	0.643	0.363	0.656	0.579	0.759
1000	5	0.537	0.619	0.539	0.620	0.528	0.631	0.539	0.621	0.539	0.620	0.534	0.636
	10	0.539	0.617	0.551	0.624	0.524	0.619	0.533	0.614	0.537	0.616	0.562	0.644
	30	0.541	0.616	0.607	0.667	0.441	0.546	0.473	0.566	0.494	0.582	0.657	0.712
$\rho_{XX'} = 0.8$													
30	5	0.596	0.875	0.602	0.872	0.543	0.903	0.598	0.876	0.602	0.874	0.549	0.904
	10	0.607	0.869	0.611	0.868	0.575	0.889	0.608	0.870	0.611	0.869	0.581	0.890
	30	0.612	0.866	0.646	0.868	0.550	0.875	0.577	0.867	0.589	0.867	0.661	0.882
50	5	0.656	0.860	0.661	0.858	0.613	0.885	0.657	0.861	0.661	0.858	0.617	0.885
	10	0.664	0.855	0.667	0.854	0.637	0.872	0.665	0.856	0.667	0.855	0.644	0.873
	30	0.667	0.853	0.692	0.855	0.624	0.858	0.643	0.853	0.653	0.853	0.705	0.869
100	5	0.704	0.842	0.707	0.840	0.672	0.863	0.705	0.842	0.707	0.841	0.675	0.863
	10	0.708	0.838	0.711	0.838	0.692	0.853	0.711	0.840	0.711	0.839	0.696	0.854
	30	0.711	0.836	0.729	0.841	0.683	0.840	0.695	0.836	0.702	0.836	0.741	0.854
1000	5	0.764	0.807	0.766	0.806	0.756	0.816	0.766	0.807	0.766	0.807	0.757	0.817
	10	0.766	0.805	0.767	0.806	0.762	0.812	0.767	0.807	0.767	0.806	0.765	0.814
	30	0.766	0.805	0.778	0.812	0.754	0.802	0.759	0.802	0.762	0.803	0.789	0.824

dist1, Normal distribution for e_{ij}; dist2, distribution with negative kurtosis for e_{ij}; dist3, distribution with positive kurtosis for e_{ij}; dist4, skewed distribution for e_{ij}; dist5, skewed distribution with negative kurtosis for e_{ij}; dist6, skewed distribution with positive kurtosis for e_{ij}; LB, lower bound; UB, upper bound.

and Shultz (1993), coefficient alpha is not robust to the violation of the normal assumption (for either true or error scores). Non-normal data tend to result in additional error or bias in estimating internal consistency reliability. A larger error makes the sample coefficient less accurate, whereas more bias causes it to further under- or overestimate the actual reliability. We note that compared with normal data, leptokurtic true or error score distributions tend to result in additional negative bias, whereas platykurtic error score distributions tend to result in a positive

bias. Neither case is desired in a reliability study, as the sample coefficient would paint an incorrect picture of the test's internal consistency by either estimating it with a larger value or a much smaller value and hence is not a valid indicator. For example, for a test with reliability being 0.6, one may calculate the sample alpha to be 0.4 because the true score distribution has a positive kurtosis, and conclude that the test is not reliable at all. On the other hand, one may have a test with actual reliability being 0.4. But because the error score distribution has a negative kurtosis, the sample

coefficient is calculated to be 0.7 and hence the test is concluded to be reliable. In either scenario, the conclusion on the test reliability is completely the opposite of the true situation, which may lead to an overlook of a reliable measure or an adoption of an unreliable instrument. Consequently, coefficient alpha is not suggested for estimating internal consistency reliability with non-normal data. Given this, it is important to make sure that in addition to satisfying the assumptions of (essential) tau-equivalence and uncorrelated errors, the sample data conform to normal distributions before one uses alpha in a reliability study.

Further, it is generally said that increased data sizes help approximate non-normal distributions to be normal. This is the case with sample sizes, not necessarily test lengths, in helping improve the accuracy, bias and/or precision of using the sample coefficient in reliability studies with non-normal data. Given the results of the study, we suggest that in order for the sample coefficient alpha to be fairly accurate and in a reasonable range, a minimum of 1000 subjects is needed for a small reliability, and a minimum of 100 is needed for a moderate reliability when the sample data depart from normality. It has to be noted that for the four sample size conditions considered in the study, the sample coefficient alpha consistently underestimates the population reliability even when

normality is assumed (see **Table 2**). However, the degree of bias becomes negligible when sample size increases to 1000 or beyond.

In the study, we considered tests of 5, 10, or 30 items administered to 30, 50, 100, or 1000 persons with the actual reliability being 0.3, 0.6, or 0.8. These values were selected to reflect levels ranging from small to large in the sample size, test length, and population reliability considerations. When using the results, one should note that they pertain to these simulated conditions and may not generalize to other conditions. In addition, we evaluated the assumption of normality alone. That is, in the simulations, data were generated assuming the other assumptions, namely (essential) tau-equivalence and uncorrelated error terms, were satisfied. In practice, it is common for observed data to violate more than one assumption. Hence, it would also be interesting to see how non-normal data affect the sample coefficient when other violations are present. Further, this study looked at the sample coefficient alpha and its empirical sampling distribution without considering its sampling theory (e.g., Kristof, 1963; Feldt, 1965). One may focus on its theoretical SE (e.g., Bay, 1973; Barchard and Hakstian, 1997a,b; Duhachek and Iacobucci, 2004) and compare them with the empirical ones to evaluate the robustness of an interval estimation of the reliability for non-normal data.

REFERENCES

Barchard, K. A., and Hakstian, R. (1997a). The effects of sampling model on inference with coefficient alpha. *Educ. Psychol. Meas.* 57, 893–905.

Barchard, K. A., and Hakstian, R. (1997b). The robustness of confidence intervals for coefficient alpha under violation of the assumption of essential parallelism. *Multivariate Behav. Res.* 32, 169–191.

Bay, K. S. (1973). The effect of non-normality on the sampling distribution and standard error of reliability coefficient estimates under an analysis of variance model. *Br. J. Math. Stat. Psychol.* 26, 45–57.

Caplan, R. D., Naidu, R. K., and Tripathi, R. C. (1984). Coping and defense: constellations vs. components. *J. Health Soc. Behav.* 25, 303–320.

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

DeVellis, R. F. (1991). *Scale Development.* Newbury Park, NJ: Sage Publications.

Duhachek, A., and Iacobucci, D. (2004). Alpha's standard error (ASE): an accurate and precise confidence interval estimate. *J. Appl. Psychol.* 89, 792–808.

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

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika* 43, 521–532.

Fleiss, J. L. (1986). *The Design and Analysis of Clinical Experiment.* New York: Wiley.

Graham, J. M. (2006). Congeneric and (essential) tau-equivalent estimates of score reliability. *Educ. Psychol. Meas.* 66, 930–944.

Green, S. B., and Hershberger, S. L. (2000). Correlated errors in true score models and their effect on coefficient alpha. *Struct. Equation Model.* 7, 251–270.

Green, S. B., and Yang, Y. (2009). Commentary on coefficient alpha: a cautionary tale. *Psychometrika* 74, 121–135.

Guttman, L. A. (1945). A basis for analyzing test-retest reliability. *Psychometrika* 10, 255–282.

Headrick, T. C. (2002). Fast fifth-order polynomial transforms for generating univariate and multivariate non-normal distributions. *Comput. Stat. Data Anal.* 40, 685–711.

Headrick, T. C. (2010). *Statistical Simulation: Power Method Polynomials and Other Transformations.* Boca Raton, FL: Chapman & Hall.

Johanson, G. A., and Brooks, G. (2010). Initial scale development: sample size for pilot studies. *Educ. Psychol. Meas.* 70, 394–400.

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

Lord, F. M., and Novick, M. R. (1968). *Statistical Theories of Mental Test Scores.* Reading: Addison-Wesley.

MathWorks. (2010). *MATLAB* (Version 7.11) [Computer software]. Natick, MA: MathWorks.

Miller, M. B. (1995). Coefficient alpha: a basic introduction from the perspectives of classical test theory and structural equation modeling. *Struct. Equation Model.* 2, 255–273.

Novick, M. R., and Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika* 32, 1–13.

Nunnally, J. C. (1967). *Psychometric Theory.* New York: McGraw-Hill.

Nunnally, J. C., and Bernstein, I. H. (1994). *Psychometric Theory*, 3rd Edn. New York: McGraw-Hill.

Raykov, T. (1998). Coefficient alpha and composite reliability with interrelated nonhomogeneous items. *Appl. Psychol. Meas.* 22, 69–76.

Shultz, G. S. (1993). *A Monte Carlo study of the robustness of coefficient alpha.* Masters thesis, University of Ottawa, Ottawa.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika* 74, 107–120.

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

Zimmerman, D. W., Zumbo, B. D., and Lalonde, C. (1993). Coefficient alpha as an estimate of test reliability under violation of two assumptions. *Educ. Psychol. Meas.* 53, 33–49.

Zumbo, B. D. (1999). *A Glance at Coefficient Alpha With an Eye Towards Robustness Studies: Some Mathematical Notes and a Simulation Model (Paper No. ESQBS-99-1).* Prince George, BC: Edgeworth Laboratory for Quantitative Behavioral Science, University of Northern British Columbia.

Zumbo, B. D., and Rupp, A. A. (2004). "Responsible modeling of measurement data for appropriate inferences: important advances in reliability and validity theory," in *The SAGE Handbook of Quantitative Methodology for the Social Sciences*, ed. D. Kaplan (Thousand Oaks: Sage), 73–92.

APPENDIX
CODE IN MATLAB

```
function result=mcalpha(n,k,evar,rho,rep)
%
% mcalpha - obtain summary statistics for sample alphas
%
%   result=mcalpha(n,k,evar,rho,rep)
%
% returns the observed mean, standard deviation, and 95% interval (qtalpha)
% for sample alphas as well as the root mean square error (rmse) and bias for
% estimating the population alpha.
%
% The INPUT arguments:
%     n - sample size
%     k - test length
%     evar - error variance
%     rho - population reliability
%     rep - number of replications
%
        alphav=zeros(rep,1);
        tbcd=[0,1,0,0,0,0];
        ebcd=[0,1,0,0,0,0];
%
% note: tbcd and ebcd are vectors containing the six coefficients, c₀,…,c₅,
% used in equation (8) for true scores and error scores, respectively. Each
% of them can be set as:
% 1. [0,1,0,0,0,0] (normal)
% 2. [0,1.643377,0,-.319988,0,.011344] (platykurtic)
% 3. [0,0.262543,0,.201036,0,.000162] (leptokurtic)
% 4. [-0.446924 1.242521 0.500764 -0.184710 -0.017947,0.003159] (skewed)
% 5. [-.276330,1.506715,.311114,-.274078,-.011595,.007683] (skewed
% platykurtic)
% 6. [-.304852,.381063,.356941,.132688,-.017363,.003570] (skewed leptokurtic)
%
        for i=1:rep
            alphav(i)=alpha(n,k,evar,rho,tbcd,ebcd);
        end
        rmse=sqrt(mean((alphav-rho).^2));
        bias=mean(alphav-rho);
        qtalpha=quantile(alphav,[.025,.975]);
        result=[mean(alphav),std(alphav),qtalpha,rmse,bias];

function A=alpha(n,k,evar,rho,tbcd,ebcd)
%
% alpha - calculate sample alpha
%
%   alp=alpha(n,k,evar,rho,tbcd,ebcd)
%
% returns the sample alpha.
%
% The INPUT arguments:
%       n - sample size
%       k - test length
%       evar - error variance
%       rho - population reliability
```

```
%       rep - number of replications
%       tbcd - coefficients for generating normal/nonnormal true score
%              distributions using power method polynomials
%       ebcd - coefficients for generating normal/nonnormal error score
%              distributions using power method polynomials
%
        tvar=evar*rho/((1-rho)*k);
        t=rfsimu(tbcd,n,1,5,tvar);
        e=rfsimu(ebcd,n,k,0,evar);
        xn=t*ones(1,k)+e;
        x=round(xn);
        alp=k/(k-1)*(1-sum(var(x,1))/var(sum(x,2),1));

function X=rfsimu(bcd,n,k,mean,var)
%
% rfsimu - generate normal/nonnormal distributions using 5-th order power
%  method polynomials
%
%  X=rfsimu(bcd,n,k,mean,var)
%
% returns samples of size n by k drawn from a distribution with the desired
% moments.
%
% The INPUT arguments:
%       bcd - coefficients for generating normal/nonnormal distributions using
%              the 5-th order polynomials
%       k - test length
%       evar - error variance
%       rho - population reliability
%       rep - number of replications
%       tbcd - coefficients for generating normal/nonnormal true score
%              distributions using power method polynomials
%       ebcd - coefficients for generating normal/nonnormal error score
%              distributions using power method polynomials
%
        Z=randn(n,k);
        Y=bcd(1)+bcd(2)*Z+bcd(3)*Z.^2+bcd(4)*Z.^3+bcd(5)*Z.^4+bcd(6)*Z.^5;
        X=mean+sqrt(var)*Y;
```