An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data

David B. Flora
Arizona State University

Patrick J. Curran
University of North Carolina at Chapel Hill

Confirmatory factor analysis (CFA) is widely used for examining hypothesized relations among ordinal variables (e.g., Likert-type items). A theoretically appropriate method fits the CFA model to polychoric correlations using either weighted least squares (WLS) or robust WLS. Importantly, this approach assumes that a continuous, normal latent process determines each observed variable. The extent to which violations of this assumption undermine CFA estimation is not well-known. In this article, the authors empirically study this issue using a computer simulation study. The results suggest that estimation of polychoric correlations is robust to modest violations of underlying normality. Further, WLS performed adequately only at the largest sample size but led to substantial estimation difficulties with smaller samples. Finally, robust WLS performed well across all conditions.

Variables characterized by an ordinal level of measurement are common in many empirical investigations within the social and behavioral sciences. A typical situation involves the development or refinement of a psychometric test or survey in which a set of ordinally scaled items (e.g., 0 = *strongly disagree,* 1 = *neither agree nor disagree,* 2 = *strongly agree*) is used to assess one or more psychological constructs. Although the individual items are designed to measure a theoretically continuous construct, the observed responses are discrete realizations of a small number of categories. Statistical methods that assume continuous distributions are often applied to observed measures that are ordinally scaled. In circumstances such as these, there is the potential for a critical mismatch between the assumptions underlying the statistical model and the empirical characteristics of the data to be analyzed. This mismatch in turn undermines confidence in the validity of the conclusions that are drawn from empirical data with respect to a theoretical model of interest (e.g., Shadish, Cook, & Campbell, 2002).

This problem often arises in confirmatory factor analysis (CFA), a statistical modeling method commonly used in many social science disciplines. CFA is a member of the more general family of structural equation models (SEMs) and provides a powerful method for testing a variety of hypotheses about a set of measured variables. By far the most common method of estimation within CFA is maximum likelihood (ML), a technique which assumes that the observed variables are continuous and normally distributed (e.g., Bollen, 1989, pp. 131–134). These assumptions are not met when the observed data are discrete (as occurs when using ordinal scales), thus significant problems can result when fitting CFA models for ordinal scales using ML estimation (e.g., B. Muthén & Kaplan, 1985). Although several alternative methods of estimation have been available for some time, these have only recently become more accessible to applied researchers through the continued development of SEM software. Despite this increased availability, key unanswered questions remain about the accuracy, validity, and empirically informed guidelines for the optimal use of these methods, particularly with respect to conditions commonly encountered in behavioral research. The goal of our article was to systematically and empirically address several of these important questions.

Structural Equation Modeling (SEM)

SEM is a powerful and flexible analytic method that plays a critically important role in many empirical applications in social science research. Because the general linear model is

embedded within SEM, this modeling framework can be used in a wide variety of applications. The general goal of SEM is to test the hypothesis that the observed covariance matrix for a set of measured variables is equal to the covariance matrix implied by an hypothesized model. This relationship can be formally stated as

$$\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta}), \qquad (1)$$

where $\mathbf{\Sigma}$ represents the population covariance matrix of a set of observed variables and $\mathbf{\Sigma}(\boldsymbol{\theta})$ represents the population covariance matrix implied by $\boldsymbol{\theta}$, a vector of model parameters. The vector $\boldsymbol{\theta}$ thus defines the form of a particular SEM through the specification of means and intercepts, variances and covariances, regression parameters, and factor loadings.

A particular parameterization of $\boldsymbol{\theta}$ leads to the well-known CFA model (Jöreskog, 1969). In CFA, the covariance matrix implied by $\boldsymbol{\theta}$ is a function of $\mathbf{\Psi}$, a matrix of variances and covariances among latent factors; $\mathbf{\Lambda}$, a matrix of factor loadings; and $\mathbf{\Theta}_\varepsilon$, a matrix of measurement errors (i.e., uniqueness). The model-implied covariance structure is

$$\mathbf{\Sigma}(\boldsymbol{\theta}) = \mathbf{\Lambda}\mathbf{\Psi}\mathbf{\Lambda}' + \mathbf{\Theta}_\varepsilon. \qquad (2)$$

Usual assumptions for CFA are that the model is properly specified (i.e., the model hypothesized in Equation 2 corresponds directly to the model that exists in the population), that $\mathbf{\Theta}_\varepsilon$ is independent of the vector $\boldsymbol{\eta}$ of latent factors, and that the measurement errors themselves are uncorrelated (i.e., $\mathbf{\Theta}_\varepsilon$ is a diagonal matrix), although this latter condition is to some degree testable.

ML is the most commonly applied method for estimating the model parameters in $\boldsymbol{\theta}$. In addition to the usual CFA assumptions named above, ML assumes that the sample covariance matrix is computed on the basis of continuous, normally distributed variables.[1] Given adequate sample size, proper model specification, and multivariately normally distributed data (or more specifically, no multivariate kurtosis; Browne, 1984), ML provides consistent, efficient, and unbiased parameter estimates and asymptotic standard errors as well as an omnibus test of model fit (Bollen, 1989; Browne, 1984).

However, in many applications in the behavioral sciences, the observed variables are not continuously distributed.[2] Instead, variables are often observed on a dichotomous or ordinal scale of measurement. It is well-known from both statistical theory and prior simulation studies that ML based on the sample product–moment correlation or covariance matrix among ordinal observed variables does not perform well, especially when the number of observed categories is small (e.g., five or fewer). In particular, the chi-square model fit statistic is inflated (Babakus, Ferguson, & Jöreskog, 1987; Green, Akey, Fleming, Hershberger, & Marquis, 1997; Hutchinson & Olmos, 1998; B. Muthén & Kaplan, 1992), parameters are underestimated (Babakus et

al., 1987; B. Muthén & Kaplan, 1992), and standard error estimates tend to be downwardly biased (B. Muthén & Kaplan, 1985, 1992). An alternative approach to estimating CFA models for ordinal observed data involves the estimation and analysis of polychoric and polyserial correlations.

Polychoric Correlations

There is a long history of theory and research with polychoric and polyserial correlations dating back to Pearson (1901). The polychoric correlation estimates the linear relationship between two unobserved continuous variables given only observed ordinal data, whereas the polyserial correlation measures the linear relationship between two continuous variables when only one of the observed distributions is ordinal and the other is continuous. Thus, calculation of a polychoric correlation is based on the premise that the observed discrete values are due to an unobserved underlying continuous distribution. We adopt the terminology of B. Muthén (1983, 1984) to refer to an unobserved univariate continuous distribution that generates an observed ordinal distribution as a *latent response distribution.*

The relationship between a latent response distribution, *y**, and an observed ordinal distribution, *y,* is formalized as

$$y = c, \text{ if } \tau_c < y^* < \tau_{c+1}, \qquad (3)$$

with thresholds τ as parameters defining the categories $c = 0, 1, 2, \ldots, C - 1$, where $\tau_0 = -\infty$ and $\tau_C = \infty$. Hence, the observed ordinal value for *y* changes when a threshold τ is exceeded on the latent response variable *y**. The primary reason that ML based on sample product–moment relationships does not perform well with ordinal observed data is that the covariance structure hypothesis (see Equation 1) holds for the latent response variables but does not generally hold for the observed ordinal variables (Bollen, 1989, p. 434).

Polychoric correlations are typically calculated using the two-stage procedure described by Olsson (1979). In the first stage, the proportions of observations in each category of a univariate ordinal variable are used to estimate the threshold parameters for each univariate latent response variable separately. Formally, for an observed ordinal variable y_1, with thresholds denoted by a_i, $i = 0, \ldots, s,$ and y_2, with thresholds denoted by b_j, $j = 0, \ldots, r,$ the first step is to estimate

$$a_i = \Phi_1^{-1}(P_i) \qquad (4)$$

---

[1] This assumption specifically applies to *endogenous* variables, or *residuals* (see Bollen, 1989, pp. 126–127). Given our focus on the CFA model, all observed variables are endogenous here.

[2] Strictly speaking, all observed measurements are discrete, but here we are talking about coarse categorization of a theoretically continuous distribution resulting in a small number of discrete levels.

and

$$b_j = \Phi_1^{-1}(P_{\cdot j}), \quad (5)$$

where P_{ij} is the observed proportion in cell (i, j), $P_{i\cdot}$ and $P_{\cdot j}$ are observed cumulative marginal proportions of the contingency table of y_1 and y_2, and Φ_1^{-1} represents the inverse of the univariate standard normal cumulative distribution function. In the second stage, these estimated threshold parameters are used in combination with the observed bivariate contingency table to estimate, via maximum likelihood, the correlation that would have been obtained had the two latent response variables been directly observed. The log-likelihood of the bivariate sample is

$$\ell = \ln K + \sum_{i=1}^{s} \sum_{j=1}^{r} n_{i,j} \ln \pi_{i,j}, \quad (6)$$

where K is a constant, $n_{i,j}$ denotes the frequency of observations in cell (i, j), and $\pi_{i,j}$ denotes the probability that a given observation falls into cell (i, j) (Olsson, 1979).

Critically important in the calculation of a polychoric correlation is the assumption that the pair of latent response variables has a bivariate normal distribution. This assumption becomes evident in the reference to the univariate standard normal distribution in the calculation of the thresholds (see Equations 4 and 5). Under the assumption of bivariate normality for y_1^* and y_2^*, Olsson (1979) gave the probability $\pi_{i,j}$ (see Equation 6) that an observation falls into a given cell of the contingency table for y_1 and y_2 as

$$\pi_{i,j} = \Phi_2(a_i, b_j) - \Phi_2(a_{i-1}, b_j) - \Phi_2(a_i, b_{j-1})$$
$$+ \Phi_2(a_{i-1}, b_{j-1}), \quad (7)$$

where Φ_2 is the bivariate normal cumulative density function with correlation ρ. The ML estimate of ρ yields the polychoric correlation between the observed ordinal variables y_1 and y_2. Olsson reported limited simulation results showing that this two-stage method for calculating a polychoric correlation between two ordinal variables gives an unbiased estimate of the correlation between a pair of bivariate normal latent response variables.

Although the assumption of bivariate normality for latent response variables has been criticized as unrealistic in practice (e.g., Lord & Novick, 1968; Yule, 1912), other researchers have advocated the practical convenience of this assumption (e.g., B. Muthén & Hofacker, 1988; Pearson & Heron, 1913). Specifically, regardless of whether a pair of observed ordinal variables represents the realization of a categorized bivariate normal distribution or some other bivariate continuous distribution, if the continuous distributions are correlated, one would expect to see evidence for this correlation in the observed contingency table for the two ordinal variables. That is, for a positive correlation, inspection of the relative frequencies in the individual cells

of the contingency table for the two ordinal variables would be expected to reflect that lower values on one ordinal variable would be associated with lower values on the other ordinal variable, whereas higher values on one variable would be associated with higher values on the other. Given that the contingency table supplies the only observed data for estimation of the correlation between the latent continuous variables, it is necessary to specify some (albeit unknown) distribution for the underlying continuous variables to allow estimation of their correlation. Because of its well-known mathematical properties, the assumption of a bivariate normal distribution considerably facilitates estimation of the correlation, as shown in Equations 4–7. Although two correlated nonnormal y^* variables would be expected to generate a contingency table with similar patterns to that observed for two normal y^* variables with the same correlation, the extent to which calculation of the polychoric correlation is robust to this nonnormality remains a matter of empirical investigation. Our goal in this article was to pursue such an empirical examination.

To our knowledge, Quiroga (1992) represents the only simulation study that has empirically evaluated the accuracy of polychoric correlations under violations of the latent normality assumption. Quiroga manipulated the skewness and kurtosis of two continuous variables (i.e., y^* variables) to examine the effects of nonnormality on polychoric correlation estimates between two variables, each with four observed ordinal categories. The polychoric correlation values consistently overestimated the true correlation between the nonnormal latent response variables. However, the extent of the overestimation was small, with bias typically less than 2% of the true correlation. Although the findings of Quiroga suggest that polychoric correlations are typically robust to violation of the underlying y^* normality assumption, to our knowledge no prior studies have examined the effect of violating this assumption on fitting CFAs to polychoric correlations. That is, demonstrating lack of bias in the estimation of polychoric correlations is necessary but not sufficient for inferring the robustness of CFAs fitted to these correlations more generally. This is particularly salient when considering alternative methods for fitting these models in practice. Two important methods of interest to us in this article are fully weighted least squares (WLS) and robust WLS.

WLS Estimation

Both analytical and empirical work have demonstrated that simply substituting a matrix of polychoric correlations for the sample product–moment covariance matrix in the usual ML estimation function for SEM is inappropriate. Although this approach will generally yield consistent parameter estimates, it is known to produce incorrect test statistics and standard errors (Babakus et al., 1987; Dolan, 1994; Rigdon & Ferguson, 1991). Over the past several

decades, a WLS approach has been developed for estimating a weight matrix based on the asymptotic variances and covariances of polychoric correlations that can be used in conjunction with a matrix of polychoric correlations in the estimation of SEM models (e.g., Browne, 1982, 1984; Jöreskog, 1994; B. Muthén, 1984; B. O. Muthén & Satorra, 1995).

WLS applies the fitting function

$$F_{\text{WLS}} = [\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})]'\mathbf{W}^{-1}[\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})], \tag{8}$$

where $\mathbf{s}$ is a vector of sample statistics (i.e., polychoric correlations), $\boldsymbol{\sigma}(\boldsymbol{\theta})$ is the model-implied vector of population elements in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, and $\mathbf{W}$ is a positive-definite weight matrix. Browne (1982, 1984) showed that if a consistent estimator of the asymptotic covariance matrix of $\mathbf{s}$ is chosen for $\mathbf{W}$, then F_{WLS} leads to asymptotically efficient parameter estimates and correct standard errors as well as a chi-square-distributed model test statistic. Furthermore, Browne (1982, 1984) presented formulae for estimating the correct asymptotic covariance matrix in the context of continuously distributed observed data using observed fourth-order moments. Because these formulae hold without specifying a particular distribution for the observed variables, F_{WLS} is often called the asymptotically distribution free (ADF) estimator when used with a correct asymptotic covariance matrix.

Browne (1982, 1984) primarily focused on WLS as applied to continuous but nonnormal distributions, whereas B. Muthén (1983, 1984) presented a "continuous/categorical variable methodology" (CVM) for estimating SEMs that allows any combination of dichotomous, ordered categorical, or continuous observed variables. With CVM, bivariate relationships among ordinal observed variables are estimated with polychoric correlations, and the SEM is fit using WLS estimation. The key contribution of CVM is that it essentially generalizes Browne's work with F_{WLS} beyond the case of continuous observed data, as Muthén described the estimation of the correct asymptotic covariance matrix among polychoric correlation estimates (B. Muthén, 1984; B. O. Muthén & Satorra, 1995). Thus, unlike normal-theory estimation, CVM provides asymptotically unbiased, consistent, and efficient parameter estimates as well as a correct chi-square test of fit with dichotomous or ordinal observed variables. Parallel but independent research by Jöreskog (Jöreskog, 1994; Jöreskog & Sörbom, 1988) similarly generalized Browne's work to the estimation of the correct asymptotic covariance matrix among polychoric correlations.[3]

Despite its asymptotic elegance, there are two potential limitations of full WLS estimation in research applications of CFA with ordinal data. First, although limited prior simulation evidence suggests that the computation of polychoric correlations is generally robust to violations of the latent normality assumption (Quiroga, 1992), to our knowledge the ramifications of these violations for the estimation of asymptotic covariances among polychoric correlations

have yet to be considered. Thus, although polychoric correlations may be generally unbiased, CFA model test statistics and standard errors might be adversely affected because of biases in the asymptotic covariance matrix introduced by nonnormality among latent response variables. Second, a frequent criticism against full WLS estimation is that the dimensions of the optimal weight matrix $\mathbf{W}$ are typically exceedingly large and increase rapidly as a function of the number of indicators in a model. By virtue of its size in the context of a large model (i.e., a model with many observed variables), $\mathbf{W}$ is often nonpositive definite and cannot be inverted when applying the WLS fitting function (e.g., Bentler, 1995; West, Finch, & Curran, 1995). Furthermore, calculation of these asymptotic values requires a large sample size to produce stable estimates. Specifically, Jöreskog and Sörbom (1996) suggested that a minimum sample size of $(k + 1)(k + 2)/2$, where k is the number of indicators in a model, should be available for estimation of $\mathbf{W}$. As the elements of $\mathbf{W}$ have substantial sampling variability when based on small sample sizes, this instability has an accumulating effect as the number of indicators in the model increases (Browne, 1984).

Thus, it is well-known that significant problems arise when using full WLS estimation in conditions commonly encountered in social science research. Simulation studies have shown that chi-square test statistics are consistently inflated when ADF estimation is applied to sample product–moment covariance or correlation matrices of continuous observed data (e.g., Chou & Bentler, 1995; Curran, West, & Finch, 1996; Hu, Bentler, & Kano, 1992). Similarly, simulation studies applying WLS estimation to the analysis of polychoric correlation matrices have also reported inflated chi-square test statistics (Dolan, 1994; Hutchinson & Olmos, 1998; Potthast, 1993) and negatively biased standard error estimates (Potthast, 1993). In particular, both Dolan and Potthast reported that these problems worsen as a function of increasing model size (i.e., number of indicators) and decreasing sample size. Thus, full WLS is often of limited usefulness in many applied research settings.

Robust WLS

To address the problems encountered when using full WLS with small to moderate sample sizes, B. Muthén, du Toit, and Spisic (1997; see also L. K. Muthén & Muthén, 1998, pp. 357–358) presented a robust WLS approach that

---

[3] Although the methods of B. Muthén and Jöreskog for estimating the asymptotic covariance matrix are highly similar, they differ in their treatment of the threshold parameters categorizing latent response distributions into observed ordinal distributions. The simulation study by Dolan (1994) suggested that the Jöreskog approach to estimating the asymptotic covariance matrix elicits virtually identical results as the Muthén approach with respect to the estimation of CFA models.

is based on the work of Satorra and colleagues (Chou, Bentler, & Satorra, 1991; Satorra, 1992; Satorra & Bentler, 1990).[4] With this method, parameter estimates are obtained by substituting a diagonal matrix, **V**, for **W** in Equation 8, the elements of which are the asymptotic variances of the thresholds and polychoric correlation estimates (i.e., the diagonal elements of the original weight matrix). Once a vector of parameter estimates is obtained, a robust asymptotic covariance matrix is used to obtain parameter standard errors. Calculation of this matrix involves the full weight matrix **W**; however, it need not be inverted. Next, Muthén et al. described a robust goodness-of-fit test via calculation of a mean- and variance-adjusted chi-square test statistic. Calculation of this test statistic also involves the full weight matrix **W** but similarly avoids inversion. An interesting aspect of this robust WLS method is that the value for the model degrees of freedom is estimated from the empirical data, in a manner inspired by Satterthwaite (1941; cited in Satorra, 1992), rather than being determined directly from the specification of the model. The robust goodness-of-fit test presented by Muthén et al. essentially involves the usual chi-square test statistic multiplied by an adjustment akin to the Satorra and Bentler (1986, 1988) robust chi-square test statistic, with model degrees of freedom estimated from the data.[5] Detailed formulae describing estimation of standard errors and model fit statistics with robust WLS were given by B. Muthén et al. as well as by L. K. Muthén and Muthén (1998, pp. 357–358).

The Current Study

Our motivating goal was to provide an empirical evaluation of a set of specific hypotheses derived from the statistical theory underlying the calculation of polychoric correlations and their use in estimating CFAs in applied research. A small number of simulation studies, most notably by Potthast (1993; see also Babakus et al., 1987; Dolan, 1994; Hutchinson & Olmos, 1998), have previously evaluated the use of polychoric correlations with full WLS estimation for CFA. In each of these studies, researchers generated $y*$ variables from continuous, multivariate normal data with known factor structures. Different sets of threshold values were then applied to these normal $y*$ variables to create ordinal y variables with varying observed distributions. These studies generally found that as the skewness or positive kurtosis of the observed (ordinal) variables departed from zero, the estimation of the known factor structure among the unobserved normal $y*$ variables deteriorated despite that the statistical theory underlying CFA with polychoric correlations makes no explicit assumption about the skewness and kurtosis of the observed ordinal variables. This finding likely occurred in part because highly kurtotic ordinal distributions produce contingency tables with low expected cell frequencies, which Olsson (1979) speculated would lead to poor polychoric correlation estimates and

which Brown and Bendetti (1977) showed would lead to biased tetrachoric correlations.[6]

In the present study, rather than manipulating the threshold values along normal $y*$ distributions to generate y variables of varying observed skewness and kurtosis, we instead manipulated the unobserved skewness and kurtosis of the $y*$ variables. In this way, we directly evaluated robustness of CFA with polychoric correlations under violation of the explicit, theoretical assumption that $y*$ variables underlying polychoric correlations be normally distributed. Furthermore, if there is little practical consequence to violation of the latent normality assumption, we will have partially demonstrated that this assumption merely provides a mathematical convenience facilitating the calculation of correlations among latent response variables, as implied by Pearson and Heron (1913) and B. Muthén and Hofacker (1988).[7] Thus, our study focused on the systematic variation of the unobserved distribution of $y*$ rather than on the observed distribution of y. This is an important distinction given that manipulation of the latter has been the sole focus of prior research (e.g., Potthast, 1993).

Statistical theory and prior empirical findings have highlighted three issues of critical importance when considering the estimation of CFA models (and SEMs in general) using polychoric correlations. First, although limited evidence suggests that polychoric correlations are robust to violations of bivariate normality for continuous latent response variables eliciting observed ordinal variables, we are aware of no prior research examining whether this robustness extends to the estimation of CFA model parameters, test statistics, and standard errors using WLS estimation with a correct asymptotic covariance matrix. Second, we are aware of no prior research examining the stability and accuracy of robust WLS under either bivariate normality or bivariate nonnormality for latent response variables. This is a promising method for use in social science research, and the finite sample performance of robust WLS is an issue of key

---

[4] Jöreskog and Sörbom (1996) described a similar approach, termed *diagonally weighted least squares.*

[5] A drawback of this robust WLS method is that the robust chi-square values cannot be used to compare nested models because the degrees of freedom may vary within a given model specification.

[6] The tetrachoric correlation is a special case of the polychoric correlation where both observed variables are dichotomous.

[7] It is possible to create any ordinal distribution from a normal distribution given the correct set of threshold values. Likewise, given an observed ordinal distribution in a practical setting (i.e., not an artificially simulated distribution), it is impossible to know whether it was generated from a normal or nonnormal continuous variable. However, applied researchers may be highly skeptical of whether it is realistic to assume that the latent response variable is in fact normal (e.g., it may represent propensity toward an abnormal behavior, such as illegal drug use).

interest. Finally, we are not aware of any existing systematic study that explores these questions with respect to the specific number of discrete levels observed in the ordinal scale; this has significant implications in the design of item response scales in anticipation of using these techniques in practice. We explore all three of these issues in detail in this article.

We drew on statistical theory and prior empirical research to generate several hypotheses that we empirically examined using a comprehensive computer simulation study. First, consistent with Quiroga (1992), we predicted that the estimates of polychoric correlations would be relatively unbiased as a function of minor to modest violations of normality of the latent response distributions. Next, we had several hypotheses pertaining to full WLS estimation of CFA models using polychoric correlations. Consistent with Dolan (1994) and Potthast (1993), we predicted that full WLS would lead to inflated test statistics and underestimated parameter standard errors and that these biases would worsen substantially as a function of increasing model size (i.e., number of indicators) and decreasing sample size. Although we focused on the chi-square test statistics as a primary outcome, this was essentially an examination of the discrepancy function upon which all commonly used fit indices for SEM are based. Thus, conditions that contribute to inflated chi-square values were also expected to contribute to bias in other fit indices, causing them also to suggest poorer fit.

A polychoric correlation matrix is a consistent estimator of the population correlation matrix for continuous latent response variables. Thus, like Rigdon and Ferguson (1991) and Potthast (1993), we expected to find unbiased parameter estimates independently of sample size and model complexity. However, because nonnormality in latent response distributions may introduce small positive biases in polychoric correlations (Quiroga, 1992), we predicted that nonnormality in latent response distributions would create modestly positively biased parameter estimates. Further, because calculation of the asymptotic covariance matrix among polychoric correlations takes into account fourth-moment information based on the distribution-free technique of Browne (1984), we did not predict that underlying nonnormality would have a substantial effect on full WLS chi-square test statistics or standard errors at larger sample sizes for less complex models.

Finally, like full WLS estimation, we expected that robust WLS would result in unbiased parameter estimates with normal latent response distributions and modestly overestimated parameters with nonnormal latent response distributions. Drawing from the results of Curran et al. (1996) relating to the Satorra–Bentler corrections applied to nonnormal but continuous measures, we predicted that robust WLS would generally produce unbiased parameter standard

errors and test statistics across a wider range of sample sizes and model complexity than would full WLS. However, it is not known to what degree robust WLS would provide accurate estimates at the smallest sample sizes for the most complex models. Taken together, we believe that our results make unique contributions both to the understanding of the finite sampling properties of these asymptotic estimators and to the provision of recommendations for applied researchers using these methods in practice.

Method

We used Monte Carlo computer simulation methodology to empirically study the effects of varying latent response distribution, sample size, and model size on the computation of chi-square model test statistics, parameter estimates, and associated standard errors pertaining to CFAs fitted to ordinal data. For each simulated sample of ordinal data, we calculated the corresponding polychoric correlation matrix and fit the relevant population model using both full and robust WLS estimation.

Continuous Latent Response Distributions

We generated random samples from five different continuous $y*$ distributions. The first was multivariate normal, whereas the remaining four represented increasing levels of nonnormality. We carefully chose the nonnormal distributions according to two criteria. First, the distributions were selected to allow examination of the separate effects of skewness and kurtosis on the outcomes of interest. Second, the distributions were selected to be representative of levels of nonnormality commonly encountered in applied psychological research as reported by Micceri (1989).[8] Table 1 presents specific population characteristics of each $y*$ distribution. Each sample of multivariate data was generated to be consistent with a given population correlation structure (see *Model Specifications* below).

Ordinal Observed Distributions

After sampling continuous data from the distributions described above, we transformed these samples into two-category and five-category ordinal data by applying a set of thresholds that remained constant across all $y*$ distributions.

---

[8] Our motivating goal for this study was not to determine what extreme level of nonnormality for $y*$ would guarantee a poorly estimated CFA model; that is, we had no interest in "breaking" the estimator. Instead, we sought to assess the performance of the methods used herein under conditions of nonnormality commonly encountered in practice. Although this decision potentially limits the generalizability of our findings, it also increases our ability to more closely examine realistic distributions that might commonly occur in practice.

Table 1

Skewness and Kurtosis of Univariate Latent Response (y*) Distributions and Five-Category Ordinal (y) Distributions

	y* distribution		y distribution	
Condition	Skewness	Kurtosis	Skewness	Kurtosis
Normal	0.00	0.00	0.02	−0.00
Low skewness vs. low kurtosis	0.75	1.75	0.30	0.22
Low skewness vs. moderate kurtosis	0.75	3.75	0.22	0.60
Moderate skewness vs. low kurtosis	1.25	1.75	0.49	−0.06
Moderate skewness vs. moderate kurtosis	1.25	3.75	0.47	0.39

Note. Skewness and kurtosis for y* distribution are nominal population values. Skewness and kurtosis for y distribution are approximate population values estimated from simulated samples of size $N = 50,000$.

For the two-category condition, a single threshold was used to transform the continuous response distributions into the observed dichotomy. For the five-category condition, a set of four thresholds was used to transform the data.[9] For the four nonnormal conditions, although these sets of thresholds resulted in observed ordinal distributions with nonzero levels of skewness and kurtosis, categorization resulted in y variables that were much less skewed and kurtotic than the y* population distributions used to generate them. See Table 1 for population characteristics of each y distribution.

Model Specifications

Each sample of multivariate data we generated had a population correlation matrix conforming to one of four factor model specifications that hold for the y* latent response variables. We selected these models to be representative of CFA model specifications that might be encountered in practice (e.g., models showing the relationships among a set of Likert-type variables). Model 1 (see Figure 1) consisted of a single factor measured by five ordinal indicators (i.e., observed y variables), each of which was determined by a latent response variable (i.e., an unobserved y* variable), as described above. Model 2 (see Figure 2) consisted of a single factor measured by 10 indicators. Model 3 (see Figure 3) consisted of two correlated factors each measured by five indicators. Finally, Model 4 (see Figure 4) was identical to Model 3, except each factor was measured by 10 indicators. As with Model 1, each of the ordinal indicators for Models 2, 3, and 4 was determined by its own y* variable.

To facilitate interpretation, we defined the values of the population parameters to be homogeneous across all four model specifications. That is, within and across the models, all factor loadings were equal (λs = .70), and all uniquenesses were equal (Θ_{ε}s = .51). Importantly, these loadings represent the regression of the latent response variables (i.e., the y* variables), and not the ordinal indicators, on the factor. Models 3 and 4 included an interfactor correlation, which was set to $\Psi_{21} = .30$ for both. All y* distributions were standardized to have a mean of zero and standard deviation equal to one, such that $\lambda_i^2 + \Theta_{\varepsilon i} = 1$ for y_i^*. The variances of the factors and of the uniquenesses were all set to equal one. Thus, the proposed models for the latent response variables were scale invariant (Cudeck, 1989, p. 326), and it was valid to define the implied population covariance matrices for data simulation in terms of correlations.

Data Generation and Analysis

For each combination of y* distribution and model specification described above, we generated random samples of four different sizes: 100, 200, 500, and 1,000. We replicated the 5 (distributions) × 4 (model specifications) × 4 (sample sizes) × 2 (number of categories) = 160 independent cells of the study 500 times each, using EQS (Version 5.7b; Bentler, 1995) to generate all data for the study.[10] For each replication, we fit the relevant population factor model using both full and robust WLS estimation as implemented

---

[9] In the five-category condition, we used the Case 1 threshold set from B. Muthén and Kaplan (1985), such that that categorization of a normal y* population distribution would result in an ordinal y distribution with zero skewness and kurtosis. This same threshold set was also applied to each nonnormal y*distribution.

[10] We generated the raw data with the EQS implementation of the methods of Fleishman (1978), who presented formulae for simulating univariate data with nonzero skewness and kurtosis, and Vale and Maurelli (1983), who extended Fleishman's method to allow simulation of multivariate data with known levels of univariate skewness and kurtosis and a known correlational structure. To ensure that EQS properly created the data according to the desired levels of skewness and kurtosis for the y* variables, we generated a set of continuous data conforming to Model 1 with $N = 50,000$ for each level of our manipulation of the y* distributions. The values of univariate skewness and kurtosis calculated from these large-sample data closely reflected the nominal population values. Previous studies have also successfully implemented this method of data generation (e.g., Chou et al., 1991; Curran et al., 1996; Hu et al., 1992).
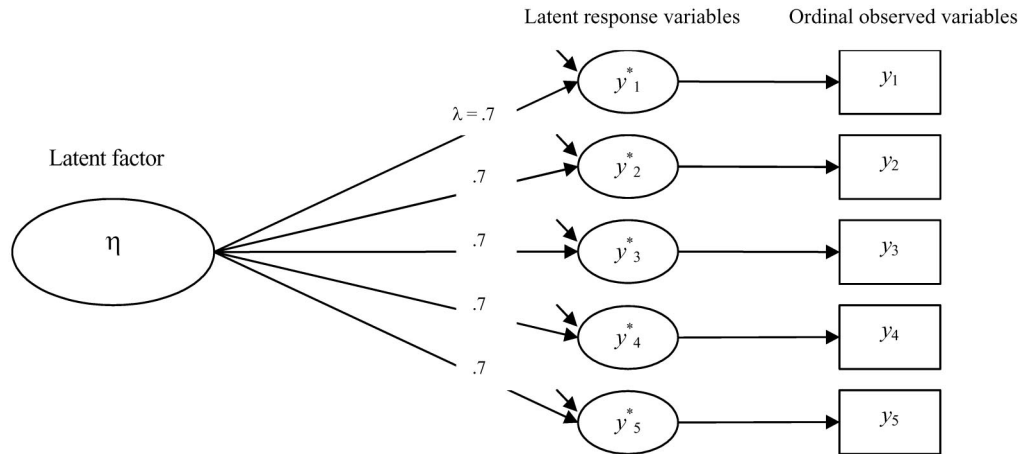
Figure 1. Model 1.

with Mplus (Version 2.01; L. K. Muthén & Muthén, 1998).[11]

Study Outcomes

We began by examining rates of improper solutions across all study conditions. An improper solution was defined as a nonconverged solution or a solution that converged but resulted in one or more out-of-bound parameters (e.g., Heywood cases). Because the underlying purpose of our study was to explicitly evaluate our proposed research hypotheses under conditions commonly encountered in applied research settings, we defined improper solutions to be invalid empirical observations. To maximize the external validity of our findings, we removed improper solutions from subsequent analyses (see Chen, Bollen, Paxton, Curran, & Kirby, 2001, for further discussion of this strategy).[12] We considered three major outcomes of interest: the chi-square test statistics, parameter estimates (i.e., factor loadings and factor correlations), and standard errors.

We examined the mean relative bias (RB; or percentage bias) of each outcome across all study conditions. The general form of this statistic is

$$RB = \left(\frac{\hat{\theta} - \theta}{\theta}\right) * 100, \tag{9}$$

where $\hat{\theta}$ is the estimated statistic (i.e., chi-square test statistic, CFA parameter estimate, or standard error) from a given replication, and θ is the relevant population parameter. For each outcome, we calculated the mean RB across replications within a given condition. Consistent with prior simulation studies (e.g., Curran et al., 1996; Kaplan, 1989), we considered RB values less than 5% indicative of trivial bias, values between 5% and 10% indicative of moderate bias, and values greater than 10% indicative of substantial bias.

For the chi-square test statistics from full WLS estima-

tion, we calculated relative bias with respect to the degrees of freedom for each model specification, given that this value reflects the expected value of a chi-square distribution. With robust WLS estimation, the degrees of freedom for the chi-square test statistic are estimated from the data and, therefore, the expected value of the chi-square statistic varies across samples within the same model specification. Thus, for each replication within a given condition, we used the estimated degrees of freedom to calculate the relative bias of the chi-square statistic elicited from the robust method. We then calculated the mean RB across samples within a given study condition.

Because of the large number of factor loadings estimated both within and across replications, we examined the mean factor loading values to summarize our results more efficiently. This strategy was appropriate given the homogeneous values of factor loadings within each model specification. Specifically, within a given cell of the study, we first calculated the mean across replications of each individual loading parameter separately (e.g., for Model 1, we obtained $\overline{\hat{\lambda}_{11}}$, $\overline{\hat{\lambda}_{21}}$, $\overline{\hat{\lambda}_{31}}$, $\overline{\hat{\lambda}_{41}}$, and $\overline{\hat{\lambda}_{51}}$). In addition, we calculated the

---

[11] Although we estimated our models using the commercial software package, Mplus (L. K. Muthén & Muthén, 1998), equivalent results would be expected using any software package with the capability for estimating polychoric correlations and their asymptotic variances and covariances (e.g., PRELIS/LISREL: Jöreskog & Sörbom, 1988, 1996; Mx: Neale, Boker, Xie, & Maes, 2002).

[12] Additional analyses reflected that the simultaneous inclusion of both proper and improper solutions resulted in modest changes in the summary statistics of the outcomes under study; however, the overall conclusions based only on the proper solutions remain substantively unchanged. Thus, there is no evidence suggesting that the omission of the improper solutions poses a threat to either the internal or the external validity of our study.
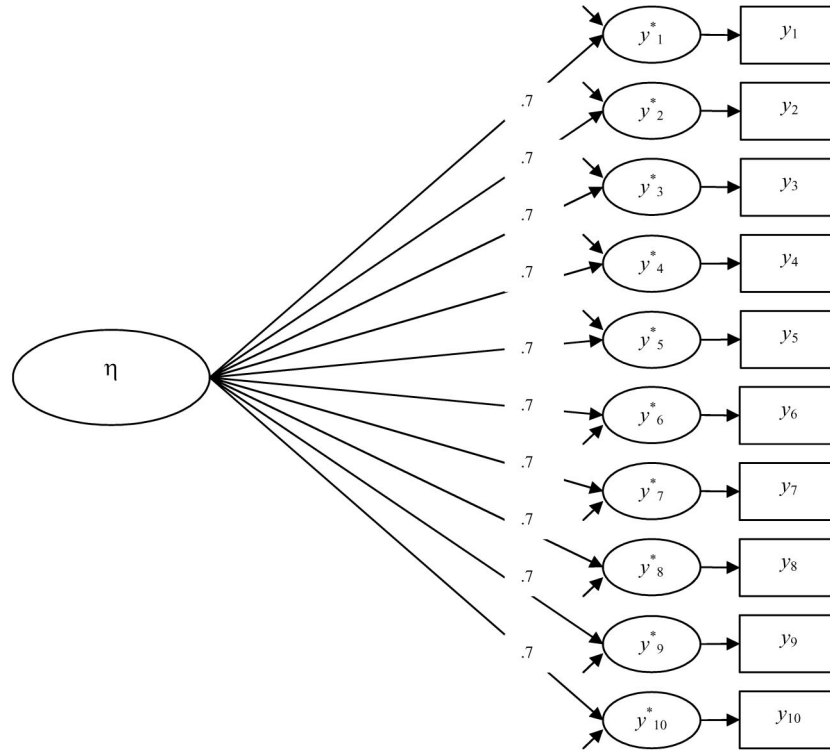
Figure 2. Model 2.

sample variance of each factor loading across replications. Because the population value of each loading parameter was homogeneous (i.e., .70) across and within all replications, we calculated the mean of these factor loadings within each cell of the study. Thus, for each cell of the experimental design we present

$$\text{Pooled Mean} = P^{-1} \sum_{i=1}^{P} \overline{\lambda}_i, \qquad (10)$$

where P is the number of indicators in the model. Also, to provide a statistic analogous to the standard deviation of the mean factor loading, we calculated the square root of the mean of the factor loading variances. Thus, for each cell, we present

$$\text{Pooled } SD = \sqrt{P^{-1} \sum_{i=1}^{P} \text{VAR}(\hat{\lambda}_i)}. \qquad (11)$$

Finally, we evaluated standard error estimation by comparing the mean value of the estimated standard error for the parameter of interest to the empirical estimate of the standard error as measured by the standard deviation of the parameter estimates across the replications of a given cell.

Results

Polychoric Correlation Estimates

Our first hypothesis related to the accuracy of polychoric correlation estimates resulting from the various latent response distribution (i.e., y^*) manipulations we used to generate the data conforming to population-level CFA models. Because there were only two possible correlation values implied by our CFA model specifications of interest, $\rho = .49$ (the correlation between two indicators loading on the same factor) and $\rho = .147$ (the correlation between two indicators loading on separate factors), we considered the effect of varying y^* distribution on these two population correlation values. Because polychoric correlations are calculated from bivariate distributions and because we wanted to focus on the robustness of polychoric correlations across varying y^* distributions prior to fitting CFA models, we began by generating 500 replications of bivariate data for each of two population correlation values (.49 and .147) within each combination of the y^* distribution, y^* categorization (two or five categories), and sample size conditions described above. Although this portion of our study analyzed only bivariate data, our results regarding the accuracy of polychoric correlations may be generalized across all fully multivariate distributions implied by all four of our CFA model specifications.
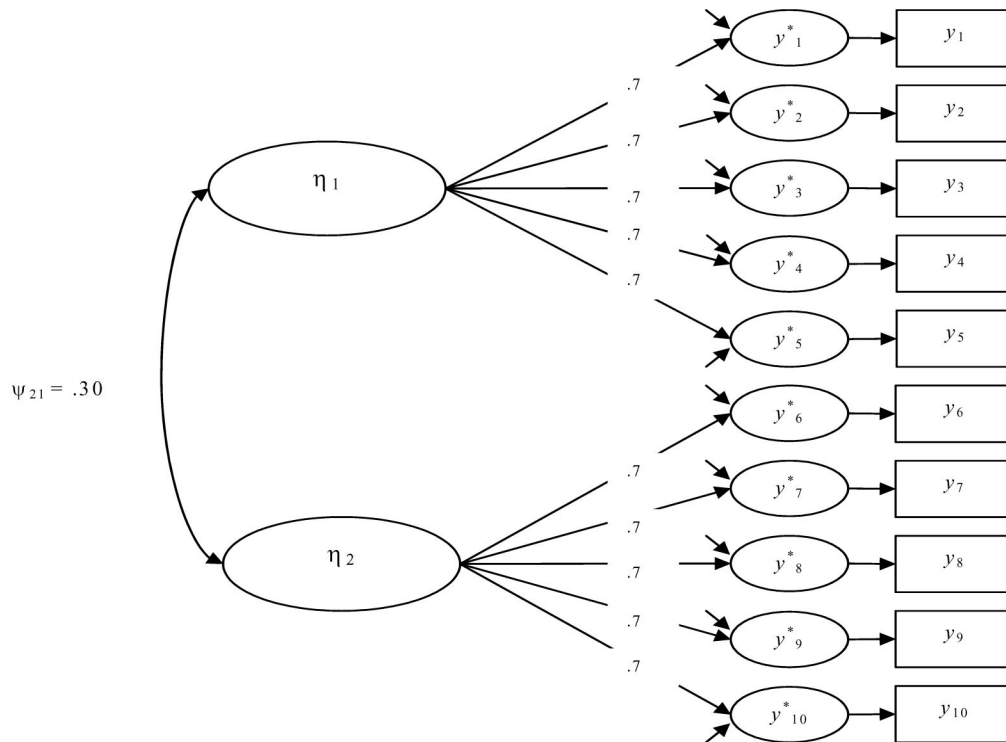
Figure 3. Model 3.

Table 2 summarizes results for the polychoric correlations in the different study conditions. When the data were generated from a bivariate normal y^* distribution, the mean polychoric correlation across replications was essentially equal to its population value (mean RB was less than 1% in nearly every cell with underlying skewness = 0 and underlying kurtosis = 0). The polychoric correlation estimates tended to become positively biased as a function of increasing nonnormality in the y^* distributions; however, mean RB remained under 10% in almost all cells. For example, Figure 5 presents box plots illustrating the distributions of polychoric correlation estimates of $\rho = .49$ obtained with $N = 1,000$ for the five-category condition. The figure reveals that although the correlation estimates were frequently positively biased, the center of these distributions did not depart substantially from the population correlation value, even with y^* nonnormality. Overall, correlation estimates of $\rho = .147$ rarely exceeded .16, and correlation estimates $\rho = .49$ rarely exceeded .52. Finally, sample size did not have any apparent effect on the accuracy of the polychoric correlations, although there was a tendency for correlations calculated from two-category data to be slightly more biased than those calculated from five-category data.

Theory would not predict that varying the thresholds across observed variables would influence our findings described above. However, to examine this possibility empirically, we generated additional bivariate y^* data from the condition with skewness = 1.25 and kurtosis = 3.75. We applied the threshold values to y_1^* to create a five-category y_1 distribution with the same skewness and kurtosis values as given in the moderate versus moderate case in Table 1, but different threshold values were applied to y_2^* resulting in a five-category population distribution for y_2 with skewness = -2.36 and kurtosis = 5.15. When samples of size $N = 1,000$ (again with 500 replications) were generated from this bivariate population distribution with true correlation $\rho = .49$, the mean polychoric correlation estimate was 0.51 ($SD = 0.04$), with mean RB of 3.21%. Thus, this result shows that polychoric correlations remain only slightly biased when markedly different sets of threshold values are applied to the two correlated y^* variables generating the observed ordinal variables.

Finally, to evaluate the effect of extremely nonnormal distributions, we generated data from correlated y^* variables defined by drastically nonnormal distributions (skewness = 5, kurtosis = 50) and applied the same sets of thresholds as above to create five-category ordinal variables (resulting in skewness = 1.68 and kurtosis = 7.01 for y_1 and y_2 when the same set of thresholds was applied to both y_1^* and y_2^* and skewness = -7.86 and kurtosis = 64.63 for y_2 when different thresholds were applied to y_2^*). We again generated 500 replications with $N = 1,000$ from each of these bivariate population distributions. With a true correlation of $\rho = .49$, the mean polychoric correlation estimate was 0.65 ($SD = 0.04$), with mean RB of 31.43% when thresholds were equal across y_1^* and y_2^*. When the set of
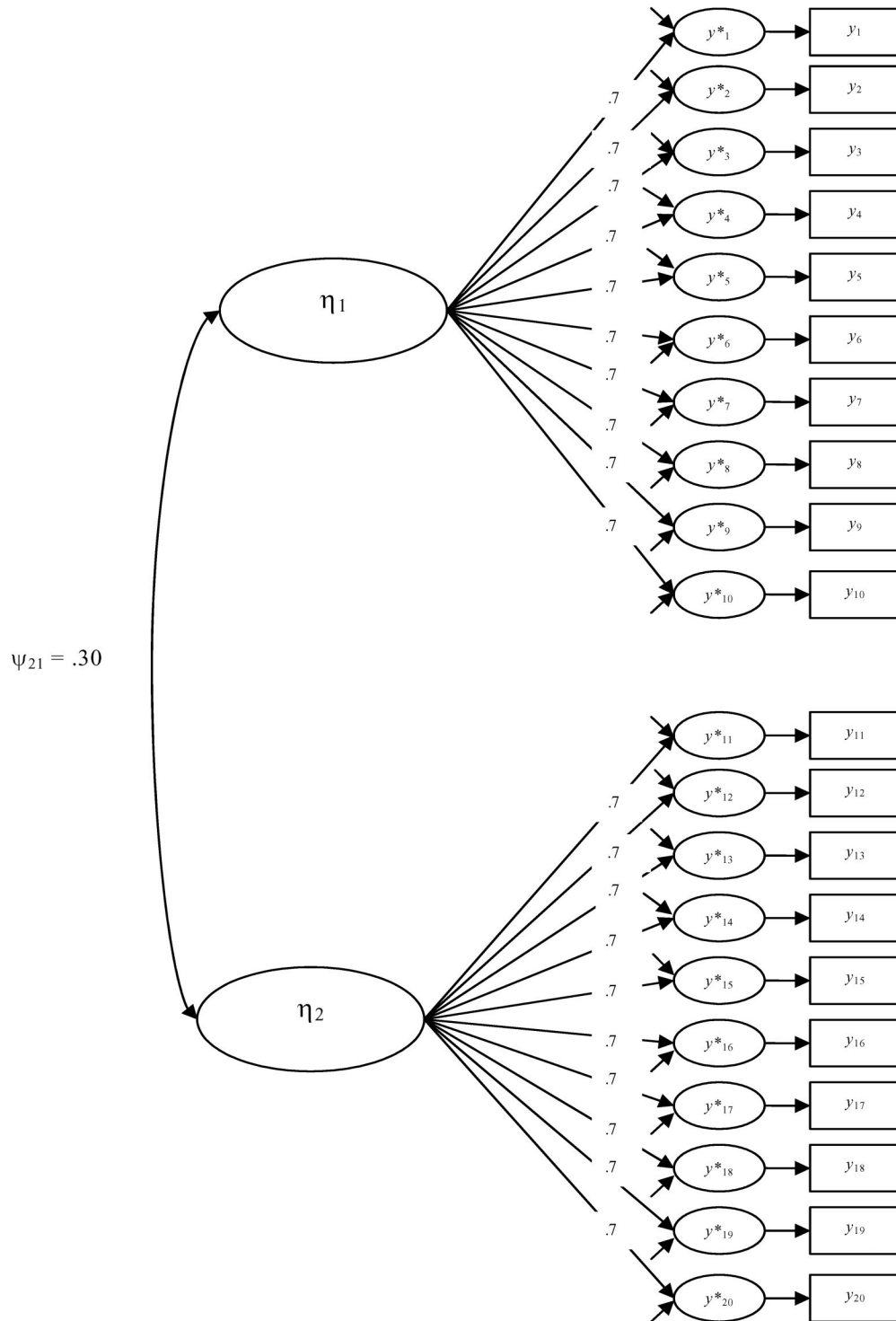
Figure 4. Model 4.

thresholds varied across y_1^* and y_2^*, the mean polychoric correlation was 0.64 ($SD = 0.09$), with mean RB of 32.21%. Thus, polychoric correlations do not appear to be robust to extreme violations of the latent normality assumption, and

this result does not seem to vary substantially when different sets of thresholds are applied to y^* variables.

In sum, consistent with Quiroga (1992), our empirical results suggest that polychoric correlations provide unbi-

Table 2
Means, Standard Deviations, and Relative Bias of Polychoric Correlation Estimates

	y* distribution		2 categories ρ = .147			5 categories ρ = .147			2 categories ρ = .49			5 categories ρ = .49		
N	s	k	M	SD	RB	M	SD	RB	M	SD	RB	M	SD	RB
100	0.00	0.00	.158	.15	7.46	.147	.11	0.26	.49	.12	−0.03	.49	.10	−0.55
	0.75	1.75	.144	.16	−1.95	.161	.11	9.62	.48	.13	−1.29	.51	.09	3.23
	0.75	3.75	.148	.15	0.43	.158	.12	7.43	.50	.13	1.27	.52	.09	5.53
	1.25	1.75	.161	.16	9.19	.153	.11	3.98	.52	.13	4.49	.50	.09	1.47
	1.25	3.75	.157	.15	6.89	.159	.11	8.18	.50	.14	2.03	.52	.09	4.44
200	0.00	0.00	.148	.10	0.47	.148	.08	0.76	.49	.10	−0.97	.49	.06	0.32
	0.75	1.75	.157	.11	6.66	.149	.08	1.34	.50	.09	1.33	.50	.06	1.55
	0.75	3.75	.153	.11	4.27	.156	.08	6.13	.50	.09	1.94	.51	.07	3.22
	1.25	1.75	.161	.12	9.31	.150	.08	2.03	.52	.09	6.99	.50	.06	1.29
	1.25	3.75	.163	.11	10.82	.150	.08	1.84	.51	.09	3.80	.51	.06	3.20
500	0.00	0.00	.151	.07	3.05	.147	.05	−0.13	.49	.06	0.78	.49	.04	0.35
	0.75	1.75	.152	.07	3.31	.150	.05	2.08	.50	.06	1.20	.50	.04	1.60
	0.75	3.75	.158	.07	7.61	.151	.05	2.48	.51	.06	3.17	.50	.04	2.99
	1.25	1.75	.160	.07	8.75	.155	.05	5.31	.52	.06	5.84	.50	.04	1.29
	1.25	3.75	.153	.07	4.20	.158	.05	7.35	.51	.06	3.51	.51	.04	4.55
1,000	0.00	0.00	.147	.05	−0.05	.147	.04	0.12	.49	.04	−0.81	.49	.03	0.61
	0.75	1.75	.153	.05	3.83	.151	.04	2.77	.50	.04	1.77	.50	.03	1.99
	0.75	3.75	.155	.05	5.77	.154	.04	4.67	.50	.04	2.73	.50	.03	2.73
	1.25	1.75	.162	.05	10.11	.153	.03	4.16	.52	.04	5.39	.49	.03	0.54
	1.25	3.75	.156	.05	5.82	.157	.04	7.18	.51	.04	3.58	.51	.03	3.94

Note. s = skewness; k = kurtosis; ρ = population correlation value; RB = mean relative bias across replications.

ased estimates of the correlations among normal and moderately nonnormal latent response variables. We did find evidence that severely nonnormal distributions led to substantial distortions in the estimation of the polychoric correlation structure and, given this distortion, we did not pursue the fitting of subsequent CFAs further. However, the robustness of polychoric correlations to minor-to-moderate violations of normality is a necessary but, quite importantly, not a sufficient condition to establish whether CFA models for latent response variables can be accurately fitted with polychoric correlation matrices. Thus, we next turn to the estimation of our CFA models using polychoric correlations estimating the relationships among latent response variables of varying distribution.

Estimation of CFA Models

Rates of improper solutions. The rates of improper and nonconverged solutions obtained with full WLS estimation are given in Table 3. With N = 100, full WLS did not produce any solutions for Model 4, the 20-indicator model (due to noninvertible weight matrices). In general, the rates of improper solutions were greater in the two-category versus five-category condition. For the two 10-indicator models (Models 2 and 3), two-category data produced high rates of improper solutions with N = 100, whereas the rates were near zero in the five-category condition. Also, nearly 100% of replications of Model 4 (the 20-indicator model)

were improper in the two-category condition where N = 200, whereas the corresponding rates in the five-category condition were only around 30%. Although the rates of improper solution obtained with full WLS varied somewhat across different y* distributions, this variation did not appear to be systematically associated with degree of nonnormality in y*. At the two largest sample sizes (N = 500 and N = 1,000), full WLS estimation converged to proper solutions of all four models across 100% of replications.

In sum, there were modest differences evident in the convergence rates for the two-category and five-category conditions. However, extensive analyses (not fully reported here) revealed that the empirical results for the chi-square test statistics, parameter estimates, and standard errors were nearly identical for these two conditions across all experimental conditions. Given this equivalence, we present the findings related to the five-category condition to retain scope and maximize focus.[13]

Chi-square test statistics. Tables 4–7 present findings for the chi-square test statistics from Models 1 through 4,

---

[13] A technical appendix containing all results from the two-category condition can be obtained from either David B. Flora or Patrick J. Curran or can be directly downloaded from www.unc.edu/~curran or from http://dx.doi.org/10.1037/1082-989X.9.4.466.supp. This appendix also includes example EQS code for data generation and Mplus code for model estimation.
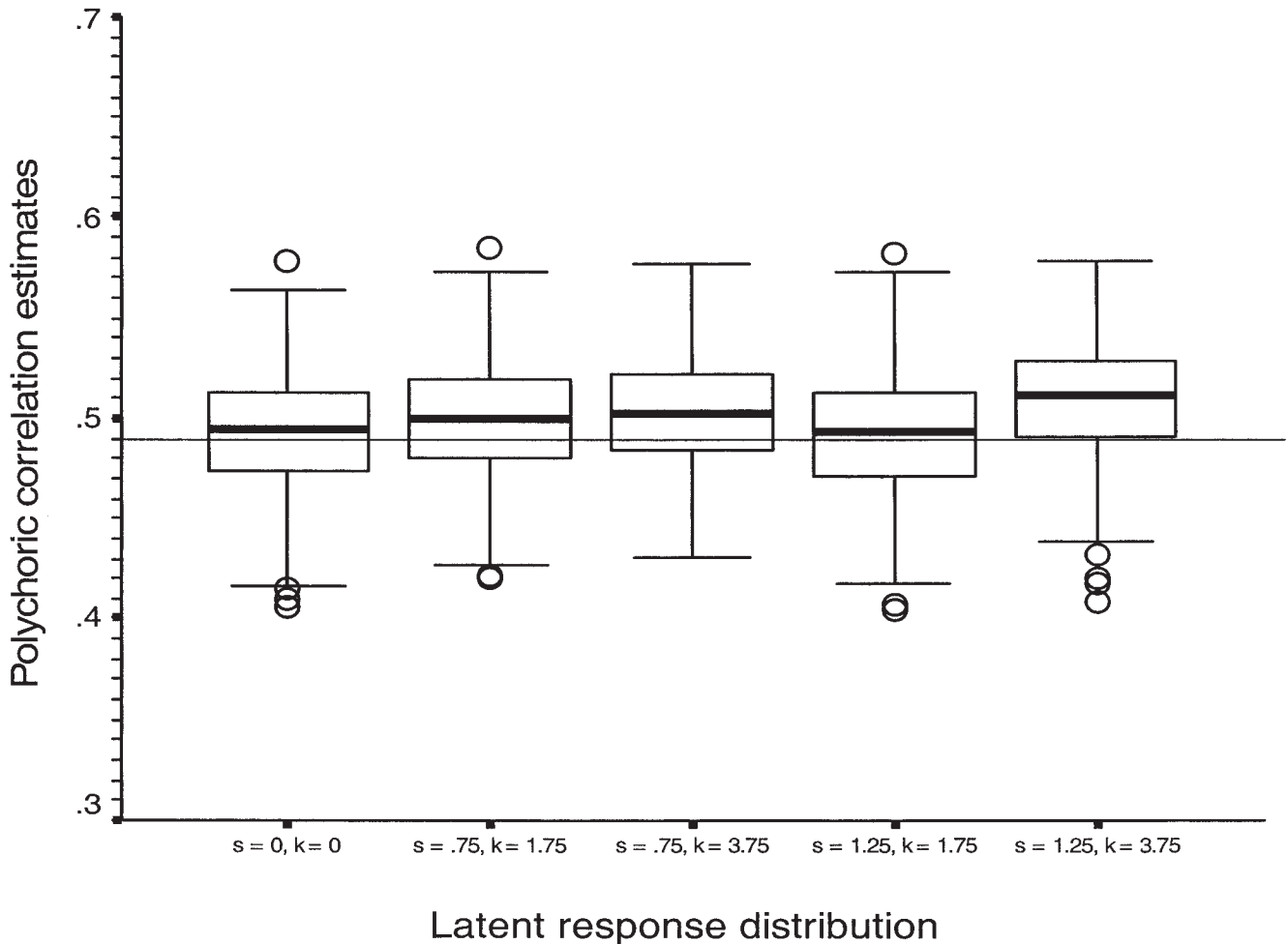
Figure 5. Distribution of polychoric correlation estimates obtained with $N = 1,000$ (horizontal line represents population correlation value). s = skewness; k = kurtosis.

respectively. Recall that with full WLS, the model degrees of freedom, and hence the expected value of the chi-square test statistic, is determined directly from model specification, whereas with robust WLS, the model degrees of freedom is estimated in part from characteristics of the data and varies across samples. Thus, although we present the across-replication mean and standard deviation of chi-square statistics obtained with full WLS, we do not present the across-replication mean and standard deviation of chi-square values obtained with robust WLS because the expected value of these statistics varies across replications within a given cell. For both methods of estimation, we present the mean RB of the chi-square test statistics and the observed Type I error rate (using $\alpha = .05$) for each cell. With robust WLS, we calculated chi-square percentage bias for each replication relative to its observed degrees of freedom and then computed the mean percentages across replications to estimate RB.

Both the chi-square test statistics and their standard deviations tend to be positively biased across all cells of the

study, particularly with full WLS estimation. This bias increases as a function of increasing model complexity. For example, full WLS chi-square statistics from Model 1 have RB values ranging from 0.02% to 7.63% with $N = 500$, whereas full WLS chi-square values from Model 4 show much greater inflation, as RB is approximately 65% with $N = 500$. Comparison of the findings for Model 2 with those for Model 3 reveals that the bias in chi-square statistics is affected not only by the number of indicators for a model but also by model complexity. Model 2 and Model 3 both have 10 indicators, but in Model 3 the indicators measure two correlated factors, whereas the indicators for Model 2 measure a single dimension. Because of this added model complexity, the chi-square statistics for Model 3 are slightly more inflated than those for Model 2.

The effect of sample size on the inflation in chi-square test values varies substantially with model specification. Within each of the four models, the chi-square RB decreases as sample size increases, but this effect is more pronounced for larger models. In addition, there appears

Table 3
Rates of Improper Solutions Obtained With Full WLS Estimation

	y* distribution		2 categories Model				5 categories Model			
N	s	k	1	2	3	4	1	2	3	4
100	0.00	0.00	2.2	30.4	40.2	—	0.0	0.0	1.2	—
	0.75	1.75	0.6	31.6	40.6	—	0.0	0.4	1.4	—
	0.75	3.75	2.0	34.0	43.6	—	0.0	0.4	1.2	—
	1.25	1.75	1.6	36.4	52.0	—	0.0	0.0	1.2	—
	1.25	3.75	1.4	31.2	50.6 (0.2)	—	0.0	0.6	1.8	—
200	0.00	0.00	0.0	0.0	0.6	94.02 (5.77)	0.0	0.0	0.0	28.8
	0.75	1.75	0.0	0.0	0.2	86.00 (13.54)	0.0	0.0	0.0	27.8
	0.75	3.75	0.0	0.0	0.8	83.46 (16.54)	0.0	0.0	0.0	33.2
	1.25	1.75	0.0	0.2	0.4	66.51 (33.49)	0.0	0.0	0.0	21.6
	1.25	3.75	0.0	0.0	0.8	76.95 (23.05)	0.0	0.0	0.0	29.4

Note. Tabled values are proportions of improper solutions for each cell. Values in parentheses are proportions of nonconverged solutions. Dashes indicate that no solutions were obtained for any replications of these cells. For $N = 500$ and $N = 1,000$, all replications converged to proper solutions (rates of improper solutions equaled 0.0% for all cells with these sample sizes). s = skewness; k = kurtosis.

to be some indication that the chi-square statistics are affected by nonnormality in $y*$. Within any given level of model specification and sample size, chi-square statistics tend to be slightly more inflated in the conditions of nonnormal $y*$ distribution. Although there is often sub-

stantial variability in the distribution of chi-square statistics across the four nonnormal $y*$ distributions, the pattern of differences does not reflect any consistent pattern across the other study conditions. Importantly, the effect of nonnormality on chi-square statistics is much

Table 4
Chi-Square Test Statistics for Model 1 (Five Indicators, One Factor)

	y* distribution		Full WLS				Robust WLS	
N	s	k	M	SD	RB	% reject	RB	% reject
100	0.00	0.00	5.81	3.87	16.22	10.40	8.66	5.60
	0.75	1.75	5.78	4.13	15.67	8.80	8.05	6.40
	0.75	3.75	5.82	4.03	16.34	10.20	7.78	6.20
	1.25	1.75	6.03	4.27	20.56	10.00	10.65	7.20
	1.25	3.75	5.88	3.82	17.68	9.40	8.12	4.60
200	0.00	0.00	4.99	3.21	−0.28	5.40	−4.32	3.60
	0.75	1.75	5.27	3.51	5.30	7.20	−1.06	3.80
	0.75	3.75	5.55	3.91	10.99	8.40	4.23	6.00
	1.25	1.75	5.47	3.53	9.50	6.60	6.41	5.60
	1.25	3.75	5.56	3.55	11.13	8.00	7.60	7.60
500	0.00	0.00	5.06	3.32	1.28	5.80	−1.22	5.40
	0.75	1.75	5.00	3.00	0.02	5.20	−1.92	4.00
	0.75	3.75	5.38	3.45	7.63	7.60	4.98	6.00
	1.25	1.75	5.26	3.30	5.23	6.60	3.13	5.60
	1.25	3.75	5.02	2.95	0.43	4.40	−1.74	3.60
1,000	0.00	0.00	4.78	3.20	−4.44	5.00	−5.21	4.60
	0.75	1.75	4.89	2.99	−2.16	4.20	−3.55	4.40
	0.75	3.75	5.16	3.26	3.13	4.60	2.27	4.20
	1.25	1.75	5.50	3.53	10.00	6.80	9.47	7.00
	1.25	3.75	4.83	2.95	−3.49	4.00	−3.88	3.20

Note. For full weighted least squares (WLS), expected value of the chi-square statistic is 5; expected standard deviation is 3.16. For robust WLS, expected value of the chi-square statistic varies within each cell, thus it is not appropriate to calculate the mean and standard deviation of chi-square statistics across replications within a given cell. s = skewness; k = kurtosis; RB = mean relative bias across replications; % reject values are Type I error rates at $\alpha = .05$.

Table 5
Chi-Square Test Statistics for Model 2 (10 Indicators, One Factor)

N	$y*$ distribution		Full WLS				Robust WLS	
	s	k	M	SD	RB	% reject	RB	% reject
100	0.00	0.00	59.68	19.94	70.50	63.80	8.42	6.40
	0.75	1.75	61.33	20.39	75.21	66.06	8.65	7.40
	0.75	3.75	63.14	21.49	80.41	70.68	10.17	8.00
	1.25	1.75	64.51	19.28	84.32	77.60	16.10	12.40
	1.25	3.75	60.19	18.69	71.99	70.42	7.68	5.00
200	0.00	0.00	44.55	12.88	27.28	28.60	4.25	5.60
	0.75	1.75	45.14	12.73	28.98	32.00	4.05	3.80
	0.75	3.75	45.40	12.19	29.70	33.60	4.46	6.40
	1.25	1.75	47.88	13.86	36.81	38.80	9.15	10.20
	1.25	3.75	45.37	12.52	29.64	33.20	5.34	4.00
500	0.00	0.00	38.53	10.11	10.08	13.80	1.85	6.00
	0.75	1.75	38.13	9.68	8.95	12.80	0.11	3.60
	0.75	3.75	38.61	9.79	10.31	13.00	1.02	3.80
	1.25	1.75	41.01	10.23	17.23	18.60	6.19	7.80
	1.25	3.75	38.30	9.90	9.42	11.40	1.62	6.20
1,000	0.00	0.00	36.52	9.31	4.35	8.40	−0.15	5.00
	0.75	1.75	36.54	9.51	4.39	9.40	−0.22	6.00
	0.75	3.75	36.77	9.46	5.06	9.40	0.13	5.80
	1.25	1.75	39.00	9.78	11.42	14.00	6.44	9.20
	1.25	3.75	36.98	9.38	5.67	8.60	1.59	6.00

Note. For full weighted least squares (WLS), expected value of the chi-square statistic is 35; expected standard deviation is 8.37. For robust WLS, expected value of the chi-square statistic varies within each cell, thus it is not appropriate to calculate the mean and standard deviation of chi-square statistics across replications within a given cell. s = skewness; k = kurtosis; RB = mean relative bias across replications; % reject values are Type I error rates at α = .05.

less pronounced than are the effects of model specification and sample size.

In general, the robust WLS chi-square statistics are inflated across most cells of the study, although typically by less than 10% of their population values (except with N = 100, where RB values tend to be between 10% and 20%). This positive bias is drastically smaller with robust WLS estimation relative to full WLS estimation, especially for the larger models. Furthermore, the influence of sample size on the chi-square test statistics appears to be much greater with full WLS than it does with robust WLS. Figure 6 illustrates Type I error rates for the model chi-square test by sample size for Models 1 and 3, as estimated with both full WLS and robust WLS. The figure reflects that for small models (i.e., the five-indicator Model 1), both methods of estimation lead to Type I error rates close to the nominal alpha level of .05 across all four sample sizes. The figure also shows that for a more complex model (i.e., Model 3, where 10 indicators represent two factors), full WLS estimation produces vastly inflated Type I error rates with samples of size less than N = 1,000, whereas the robust WLS method continues to produce Type I error rates close to the nominal alpha level across all sample sizes. Finally, there does not appear to be any consistent tendency for

robust WLS chi-square values to be more biased under nonnormal $y*$.

Parameter estimates. Tables 8–11 present findings pertaining to the parameter estimates (i.e., factor loadings and interfactor correlations) obtained for Models 1, 2, 3, and 4, respectively. On average, the parameters are overestimated across all conditions, and this positive bias increases with increasing model size. However, the overall bias in parameter estimation is notably smaller with robust WLS than it is with full WLS, especially for larger models and smaller sample sizes. In general, the bias in parameter estimates obtained with robust WLS is consistently trivial (i.e., less than 5%, with the exception of estimates of factor correlations with N = 100 and 200, where RB is trivial to moderate, or less than 10%). Even with full WLS estimation, the parameter overestimation is typically small, rarely leading to parameter estimates greater than the corresponding population value by more than .1 in the correlation metric. Whereas the overestimation of parameters with full WLS increases as a function of increasing number of indicators for the model, there is no such effect of model size on parameter estimation with robust WLS. Furthermore, parameter overestimation with full WLS improves as a function of increasing sample size, particularly with larger mod-

Table 6
Chi-Square Test Statistics for Model 3 (10 Indicators, Two Correlated Factors)

N	$y*$ distribution		Full WLS				Robust WLS	
	s	k	M	SD	RB	% reject	RB	% reject
100	0.00	0.00	65.25	22.94	91.90	74.29	15.03	12.20
	0.75	1.75	65.95	23.71	93.96	73.63	16.85	12.40
	0.75	3.75	67.49	22.22	98.51	81.38	14.46	12.00
	1.25	1.75	65.78	22.92	93.48	75.51	16.76	12.20
	1.25	3.75	65.44	22.54	92.46	77.80	15.10	10.80
200	0.00	0.00	45.98	13.59	35.23	37.80	5.47	6.60
	0.75	1.75	46.34	15.03	36.29	39.40	6.53	7.60
	0.75	3.75	46.90	13.79	37.94	40.40	5.49	7.80
	1.25	1.75	47.73	14.19	40.38	42.40	10.20	8.60
	1.25	3.75	46.09	14.06	35.55	38.80	6.51	8.60
500	0.00	0.00	38.14	10.17	12.19	15.60	1.40	5.00
	0.75	1.75	38.97	10.06	14.62	16.60	3.33	6.60
	0.75	3.75	38.46	9.79	13.10	15.60	3.10	7.80
	1.25	1.75	40.16	11.12	18.13	19.00	6.93	8.20
	1.25	3.75	38.77	10.65	14.04	15.20	3.65	7.20
1,000	0.00	0.00	36.04	8.88	6.00	9.60	1.36	5.60
	0.75	1.75	36.29	9.28	6.74	10.20	1.65	5.80
	0.75	3.75	36.11	8.80	6.21	8.20	−0.19	3.80
	1.25	1.75	37.24	9.00	9.54	10.60	2.75	5.60
	1.25	3.75	35.69	9.38	4.97	8.60	−1.68	3.60

Note.   For full weighted least squares (WLS), expected value of the chi-square statistic is 34; expected standard deviation is 8.25. For robust WLS, expected value of the chi-square statistic varies within each cell, thus it is not appropriate to calculate the mean and standard deviation of chi-square statistics across replications within a given cell. s = skewness; k = kurtosis; RB = mean relative bias across replications; % reject values are Type I error rates at α = .05.

els. But with robust WLS, the effect of sample size on parameter estimation is quite small.

Parameter estimates (both factor loadings and factor correlations) seem to be affected by nonnormality in the $y*$ variables, although this effect is small. Across all cells, the parameter estimates are positively biased, even with normal $y*,$ and this parameter overestimation increases with greater nonnormality in $y*$. For example, using full WLS estimation, the mean RB of factor loadings when Model 1 is estimated from normally distributed $y*$ variables is 2.57% with N = 100 and 0.29% with N = 1,000, yet when the same model is estimated from $y*$ variables with univariate skewness = 1.25 and univariate kurtosis = 3.75, the mean RB of factor loadings is 4.86% with N = 100 and 2.29% with N = 1,000. Overall, this small effect of $y*$ nonnormality on parameter estimation was found with both full and robust WLS estimation.

Standard errors.   The conditions leading to greater bias in the standard errors were nearly identical to those that led to greater bias in chi-square test statistics, thus we do not present detailed results pertaining to standard error estimation.[14] Across all study conditions, the standard errors were consistently negatively biased relative to the standard deviation of the relevant parameter sampling distribution obtained empirically across replications. Specifically, standard

errors were more biased as a function of increasing model size and decreasing sample size but were not systematically affected by $y*$ nonnormality. Under full WLS estimation at N = 1,000, the mean factor loading standard error RB ranges from around −2.5% to −1% across replications of Model 1, from around −9% to −5% across replications of Models 2 and 3, and from around −25% to −22% across replications of Model 4. Although the robust WLS approach also frequently produced negatively biased standard errors, these biases were much smaller than the standard error bias produced with full WLS estimation. Using the robust WLS method with N = 1,000, the mean factor loading standard error RB ranges from around −1% to 1% across replications of Model 1, from around −3% to 1% across replications of Models 2 and 3, and from around −3% to −1% across replications of Model 4.

Table 7
Chi-Square Test Statistics for Model 4 (20 Indicators, Two Correlated Factors)

	y^* distribution		Full WLS				Robust WLS	
N	s	k	M	SD	RB	% reject	RB	% reject
100	0.00	0.00	—	—	—	—	17.13	15.00
	0.75	1.75	—	—	—	—	15.46	14.40
	0.75	3.75	—	—	—	—	16.19	12.80
	1.25	1.75	—	—	—	—	19.45	17.60
	1.25	3.75	—	—	—	—	17.03	13.80
200	0.00	0.00	1235.67	317.21	631.16	100.00	8.21	8.20
	0.75	1.75	1271.38	327.99	652.29	100.00	7.93	7.80
	0.75	3.75	1388.40	404.18	721.54	100.00	7.73	7.60
	1.25	1.75	1279.96	322.21	657.37	100.00	10.97	11.00
	1.25	3.75	1292.71	369.64	664.92	100.00	9.07	10.00
500	0.00	0.00	272.76	37.90	61.39	97.80	3.94	7.80
	0.75	1.75	275.02	40.44	62.73	97.40	3.58	6.80
	0.75	3.75	279.43	42.22	65.34	97.80	2.92	6.20
	1.25	1.75	285.95	42.51	69.20	99.20	5.91	7.80
	1.25	3.75	273.99	38.48	62.13	99.00	3.58	6.80
1,000	0.00	0.00	210.55	26.49	24.59	63.80	1.53	6.40
	0.75	1.75	211.90	25.21	25.38	66.60	2.30	7.00
	0.75	3.75	211.82	25.63	25.34	65.00	1.98	5.80
	1.25	1.75	219.15	26.87	29.68	74.00	3.92	6.80
	1.25	3.75	211.43	24.62	25.11	67.40	2.96	7.20

Note. For full weighted least squares (WLS), expected value of the chi-square statistic is 169; expected standard deviation is 18.38. For robust WLS, expected value of the chi-square statistic varies within each cell, thus it is not appropriate to calculate the mean and standard deviation of chi-square statistics across replications within a given cell. Dashes indicate values that are not available because of noninvertible weight matrices. s = skewness; k = kurtosis; RB = mean relative bias across replications; % reject values are Type I error rates at $\alpha = .05$.

Discussion

We used a comprehensive simulation study to empirically test our set of theoretically generated research hypotheses pertaining to the performance of CFA for ordinal data with polychoric correlations using both full WLS (as per Browne, 1984, and B. Muthén, 1983, 1984) and robust WLS (as per Muthén et al., 1997) estimation. The results of our study provided support for each of our proposed hypotheses.

First, as predicted, we replicated Quiroga's (1992) findings that polychoric correlations among ordinal variables accurately estimated the bivariate relations among normally distributed latent response variables and that modest violation of normality for latent response variables of a degree that might be expected in applied research leads to only slightly biased estimates of polychoric correlations. Both our results and Quiroga's suggest that this finding occurs regardless of whether the same set of threshold values is applied to both latent response variables leading to the observed ordinal variables. A thorough examination of the effects of differing thresholds and level of skewness across the two variables on polychoric correlation estimates is

beyond the scope of our study. However, we note that Quiroga found that polychoric correlation estimates remained accurate when the ordinal variables had differing thresholds or skewness of opposite sign. Specifically, although increasing levels of nonnormality tended to produce increasingly positively biased correlation estimates, these biases remained quite small and were typically less than .03 in the correlation metric. Furthermore, we found that variability in polychoric correlations was not affected by nonnormality and that sample size had no effect on the mean polychoric correlation estimates across cells.

These results likely occurred because nonnormality in the latent response variables, in combination with the threshold values used to categorize the data, did not produce contingency tables with low expected cell frequencies (see Brown & Bendetti, 1977; Olsson, 1979). Because our results indicated that the polychoric correlations accurately recovered the correlations among the unobserved y^* variables even under modest violation of the normality assumption, we proceeded to examine the adequacy of fitting CFA models to these correlation structures.

As we described earlier, it has been analytically demonstrated that when CFA models are fitted using observed
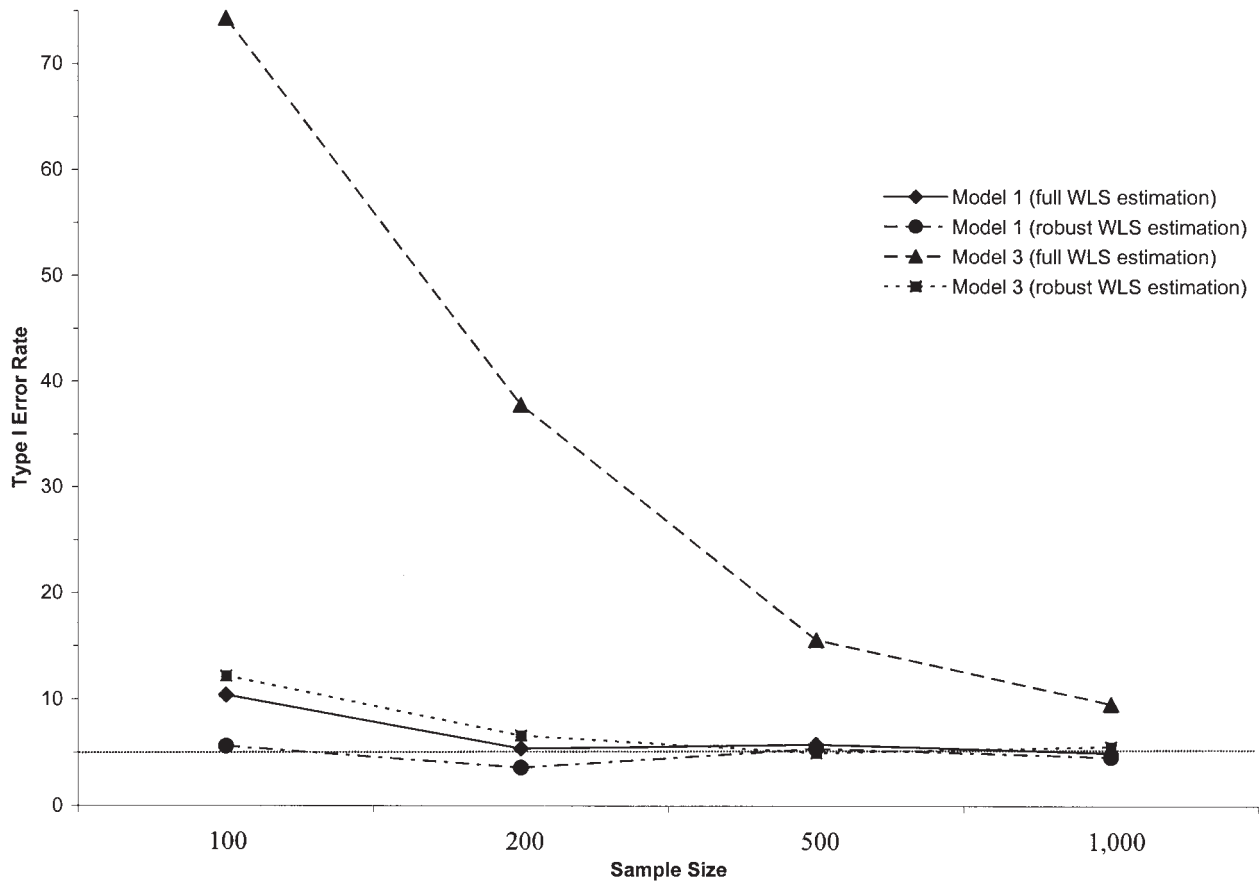
Figure 6. Type I error rates with nominal alpha level = .05 for chi-square test of model fit by sample size for Model 1 (5 indicators of one factor) and Model 3 (10 indicators of two factors). WLS = weighted least squares.

polychoric correlation matrices, full WLS estimation produces asymptotically correct chi-square tests of model fit and parameter standard errors (e.g., B. Muthén, 1983, 1984; B. O. Muthén & Satorra, 1995). However, in practice, the use of full WLS is often problematic, especially when models with a large number of indicators are estimated with sample sizes typically encountered in social science research. In our study, we found that when the 20-indicator model (Model 4) was estimated using $N = 100$, the estimated asymptotic covariance matrix **W** was consistently nonpositive definite and could not be inverted for any of the replications. Thus, not a single solution across all replications was obtained in this situation using full WLS estimation. This finding is consistent with Browne's (1984) observation that this method of estimation "will tend to become infeasible" as the number of variables approaches $p = 20$ (p. 73) and Jöreskog & Sörbom's (1996) recommendation that a minimum sample size be attained for estimation of **W**. When Model 4 was estimated with samples of size $N = 200$, **W** was usually invertible, but the full WLS fitting function frequently failed to converge to a proper CFA solution.

However, the overall rates of nonconvergence and im-

proper solutions obtained with full WLS in the present study are less than those found in simulation studies applying full WLS to continuously distributed data (e.g., Curran et al., 1996). Nonetheless, other studies applying full WLS estimation to the analysis of polychoric correlations have found rates of nonconvergence and improper solutions similar to those reported here: Neither Dolan (1994) nor Potthast (1993) obtained nonpositive definite weight matrices or improper solutions for any replications of their respective studies, which analyzed samples of size 200 and greater, whereas Babakus et al. (1987) obtained high rates of nonconvergence and improper solutions only with samples of 100.

One advantage of the robust WLS method relative to full WLS is that sample solutions for the CFA model can still be obtained with robust WLS estimation even when the weight matrix is nonpositive definite. Thus, robust WLS estimation successfully obtained solutions to Model 4 with $N = 100$, whereas full WLS did not. Furthermore, our findings show that the likelihood of obtaining an improper solution or encountering convergence difficulty is near zero with robust

Table 8
Mean Factor Loadings for Model 1 (Five Indicators, One Factor)

	y^* distribution		Full WLS			Robust WLS		
			Pooled	Pooled		Pooled	Pooled	
N	s	k	M	SD	RB	M	SD	RB
100	0.00	0.00	.718	.080	2.57	.702	.078	0.29
	0.75	1.75	.726	.081	3.71	.711	.078	1.57
	0.75	3.75	.731	.082	4.43	.717	.078	2.43
	1.25	1.75	.717	.081	2.43	.703	.078	0.43
	1.25	3.75	.734	.079	4.86	.719	.076	2.71
200	0.00	0.00	.710	.056	1.43	.701	.055	0.14
	0.75	1.75	.716	.055	2.29	.707	.054	1.00
	0.75	3.75	.722	.056	3.14	.713	.054	1.86
	1.25	1.75	.713	.055	1.86	.705	.054	0.71
	1.25	3.75	.724	.056	3.43	.715	.055	2.14
500	0.00	0.00	.705	.035	0.71	.701	.034	0.14
	0.75	1.75	.711	.034	1.57	.707	.034	1.00
	0.75	3.75	.717	.035	2.43	.713	.034	1.86
	1.25	1.75	.707	.033	1.00	.703	.033	0.43
	1.25	3.75	.717	.033	2.43	.713	.033	1.86
1,000	0.00	0.00	.702	.024	0.29	.700	.024	0.00
	0.75	1.75	.709	.024	1.29	.707	.023	1.00
	0.75	3.75	.713	.024	1.86	.711	.024	1.57
	1.25	1.75	.706	.023	0.86	.704	.023	0.57
	1.25	3.75	.716	.024	2.29	.714	.024	2.00

Note. Factor loading population value = .7. s = skewness; k = kurtosis; WLS = weighted least squares; Pooled M is the mean within-cell factor loadings; Pooled SD is the square root of within-cell factor loading variances; RB = mean relative bias across replications.

WLS estimation, even when the model is large and the sample is small.

Next, we hypothesized that full WLS estimation of CFA models for ordinal data using polychoric correlations would lead to inflated test statistics and underestimated standard errors. Our findings supported this prediction in that, on average, the chi-square test statistics were positively biased and parameter standard errors were negatively biased in nearly every cell of the study. The results also supported our more specific predictions that these biases would increase as a function of decreasing sample size and increasing model complexity. Specifically, the effect of sample size was larger for more complex models. With the five-indicator model, chi-square statistics were substantially biased at $N =$ 100, whereas sample sizes of 200 and even 500 led to substantially biased chi-square statistics for the two 10-indicator models. Sample sizes as large as $N = 1,000$ still led to heavily inflated chi-square values for the 20-indicator model. However, with the two 10-indicator models, chi-square values showed substantial inflation with sample sizes below 1,000. For the 20-indicator model, full WLS estimation produced vastly inflated chi-square values across all sample sizes, virtually guaranteeing rejection of a properly specified CFA model with $N = 200$ or $N = 500$. Thus, our findings are similar to those of Dolan (1994), who con-

cluded that a sample size of 200 is not sufficient to estimate an eight-indicator model with full WLS using polychoric correlations, and to those of Potthast (1993), who reported significant problems when nine-parameter and larger models are estimated with a sample size as large as 1,000.

The model test statistics produced with robust WLS also tended to be positively biased relative to their expected values; however, these biases were substantially less than those observed with the test statistics produced by full WLS. For the 10- and 20-indicator models, the mean RB of the robust WLS test statistics was typically less than 10% for all cells with sample size 200 or greater. Whereas full WLS estimation led to high Type I error rates for the 10- and 20-indicator models, robust WLS model rejection rates were much closer to the nominal .05 level. Thus, our prediction that robust WLS estimation might produce test statistics accurately distributed as chi-square, with degrees of freedom estimated from the data as per B. Muthén et al. (1997), was supported across all four model specifications estimated with samples of size 200 or greater.

Because of its reliance on the complete asymptotic variance–covariance matrix, it is not surprising that full WLS estimation produced increasingly biased test statistics as a function of decreasing sample size combined with increasing model size. In contrast, robust WLS estimation is pri-

Table 9
Mean Factor Loadings for Model 2 (10 Indicators, One Factor)

N	y* distribution s	k	Full WLS Pooled M	Pooled SD	RB	Robust WLS Pooled M	Pooled SD	RB
100	0.00	0.00	.775	.080	10.71	.703	.066	0.43
	0.75	1.75	.782	.078	11.71	.712	.066	1.71
	0.75	3.75	.783	.082	11.86	.715	.067	2.14
	1.25	1.75	.780	.077	11.43	.712	.064	1.71
	1.25	3.75	.786	.081	12.29	.716	.067	2.29
200	0.00	0.00	.746	.051	6.57	.702	.047	0.29
	0.75	1.75	.751	.051	7.29	.707	.047	1.00
	0.75	3.75	.753	.052	7.57	.711	.047	1.57
	1.25	1.75	.745	.051	6.43	.707	.046	1.00
	1.25	3.75	.760	.051	8.57	.717	.047	2.43
500	0.00	0.00	.721	.030	3.00	.701	.029	0.14
	0.75	1.75	.726	.031	3.71	.707	.029	1.00
	0.75	3.75	.731	.031	4.43	.712	.030	1.71
	1.25	1.75	.722	.030	3.14	.704	.029	0.57
	1.25	3.75	.733	.031	4.71	.714	.030	2.00
1,000	0.00	0.00	.710	.021	1.43	.701	.020	0.14
	0.75	1.75	.716	.021	2.29	.706	.021	0.86
	0.75	3.75	.722	.021	3.14	.712	.021	1.71
	1.25	1.75	.713	.021	1.86	.704	.020	0.57
	1.25	3.75	.723	.022	3.29	.713	.021	1.86

Note. Factor loading population value = .7. s = skewness; k = kurtosis; WLS = weighted least squares; Pooled M is the mean within-cell factor loadings; Pooled SD is the square root of within-cell factor loading variances; RB = mean relative bias across replications.

marily a function of only the asymptotic variances, and not the covariances, among the sample correlation estimates. Therefore, solutions obtained with robust WLS estimation are not affected by inaccuracies in the full weight matrix nearly to the same extent that full WLS solutions are affected.

Because polychoric correlations provide consistent estimates of the relationships among latent response variables, we predicted that CFA parameter estimates would be unbiased. Our findings suggest that, for normally distributed latent response variables, parameter estimates obtained with full WLS estimation tended to be somewhat positively biased with overestimation increasing as a function of increasing model size and decreasing sample size. However, these biases were relatively small across all cells of the simulation: Even when the 20-indicator model was estimated with $N = 200$, estimates of the population factor loading .70 were consistently less than .80, and estimates of the population factor correlation .30 were typically less than .40. Dolan (1994) found that parameters tended to be slightly overestimated with $N = 400$ and less, whereas Potthast (1993) concluded that parameter estimate bias was trivial across all cells of her simulation study, which only had two conditions of sample size, $N = 500$ and $N = 1,000$. Our results essentially replicated these findings. With robust

WLS estimation, parameter estimates were mostly unbiased with normally distributed latent response variables.

As predicted, for both full WLS estimation and robust WLS estimation, we found that increasing levels of nonnormality in latent response variables was associated with greater positive bias in parameter estimates, echoing the tendency of polychoric correlations to be positively biased when observed ordinal data derives from nonnormal latent response variables. However, in that nonnormality in latent response distributions produces only slightly biased polychoric correlations, this nonnormality introduces only slight bias in CFA parameter estimates. As we noted earlier, because the polychoric correlations resulting from extremely nonnormal continuous latent response distributions were substantially distorted, we did not fit CFA models to these correlation structures. Theory would strongly predict that the CFA parameter estimates and standard errors would be biased as a function of the distorted correlation structure.

Finally, we found that nonnormality in latent response variables contributed to only a slight increase in the positive bias of chi-square values obtained with full WLS estimation but not with robust WLS estimation. Similarly, Babakus et al. (1987), Potthast (1993), and Hutchinson and Olmos (1998) found that chi-square test statistics become more biased with increasing nonnormality in the observed ordinal

Table 10
Mean Parameter Estimates for Model 3 (10 Indicators, Two Correlated Constructs)

	$y*$ distribution		Full WLS						Robust WLS					
			Loadings (λ)			Correlation (ψ_{21})			Loadings (λ)			Correlation (ψ_{21})		
N	s	k	M	SD	RB	M	SD	RB	M	SD	RB	M	SD	RB
100	0.00	0.00	.744	.097	6.29	.363	.165	21.16	.703	.078	0.43	.313	.115	4.49
	0.75	1.75	.746	.101	6.57	.340	.175	13.31	.707	.080	1.00	.301	.122	0.39
	0.75	3.75	.753	.099	7.57	.354	.178	17.87	.713	.078	1.86	.319	.130	6.44
	1.25	1.75	.749	.095	7.00	.363	.174	21.00	.709	.078	1.29	.320	.118	6.55
	1.25	3.75	.762	.097	8.86	.371	.173	23.69	.721	.078	3.00	.327	.120	8.86
200	0.00	0.00	.727	.060	3.86	.339	.098	13.04	.701	.054	0.14	.306	.079	1.86
	0.75	1.75	.733	.061	4.71	.337	.101	12.46	.708	.055	1.14	.311	.087	3.53
	0.75	3.75	.737	.061	5.29	.340	.100	13.29	.713	.055	1.86	.311	.084	3.56
	1.25	1.75	.729	.061	4.14	.348	.100	15.92	.705	.055	0.71	.317	.083	5.81
	1.25	3.75	.740	.060	5.71	.332	.106	10.54	.716	.054	2.29	.311	.087	3.57
500	0.00	0.00	.712	.035	1.71	.316	.059	5.22	.701	.034	0.14	.302	.054	0.82
	0.75	1.75	.718	.036	2.57	.318	.055	6.10	.707	.034	1.00	.305	.051	1.55
	0.75	3.75	.722	.036	3.14	.317	.057	5.70	.711	.035	1.57	.304	.052	1.35
	1.25	1.75	.716	.035	2.29	.322	.057	7.48	.705	.034	0.71	.310	.052	3.40
	1.25	3.75	.725	.036	3.57	.323	.058	7.53	.714	.034	2.00	.310	.053	3.28
1,000	0.00	0.00	.706	.024	0.86	.308	.038	2.71	.701	.024	0.14	.302	.036	0.52
	0.75	1.75	.712	.025	1.71	.313	.038	4.39	.706	.025	0.86	.306	.037	2.14
	0.75	3.75	.716	.025	2.29	.309	.041	2.87	.710	.025	1.43	.302	.039	0.81
	1.25	1.75	.709	.024	1.29	.321	.037	6.85	.703	.024	0.43	.314	.036	4.72
	1.25	3.75	.719	.024	2.71	.315	.041	5.03	.713	.024	1.86	.309	.039	2.91

Note. Factor loading population value = .7; factor correlation population value = .3. s = skewness; k = kurtosis; WLS = weighted least squares; RB = mean relative bias across replications. For loadings, *M* is the pooled mean within-cell factor loadings; *SD* is the square root of within-cell factor loading variances.

Table 11
Mean Parameter Estimates for Model 4 (20 Indicators, Two Correlated Constructs)

	$y*$ distribution		Full WLS						Robust WLS					
			Loadings (λ)			Correlation (ψ_{21})			Loadings (λ)			Correlation (ψ_{21})		
N	s	k	M	SD	RB	M	SD	RB	M	SD	RB	M	SD	RB
100	0.00	—	—	—	—	—	—	—	.703	.069	0.43	.307	.111	2.24
	0.75	1.75	—	—	—	—	—	—	.711	.068	1.57	.305	.107	1.62
	0.75	3.75	—	—	—	—	—	—	.716	.070	2.29	.315	.106	5.09
	1.25	1.75	—	—	—	—	—	—	.708	.068	1.14	.318	.098	5.92
	1.25	3.75	—	—	—	—	—	—	.718	.069	2.57	.310	.102	3.33
200	0.00	0.00	.786	.104	12.29	.423	.185	41.02	.703	.048	0.43	.302	.078	0.54
	0.75	1.75	.788	.102	12.57	.404	.184	34.80	.709	.048	1.29	.307	.076	2.22
	0.75	3.75	.786	.106	12.29	.413	.179	37.52	.712	.048	1.71	.303	.075	0.96
	1.25	1.75	.784	.102	12.00	.419	.183	39.74	.705	.047	0.71	.315	.071	5.12
	1.25	3.75	.790	.104	12.86	.400	.180	33.35	.715	.048	2.14	.307	.074	2.19
500	0.00	0.00	.745	.036	6.43	.359	.061	19.59	.701	.030	0.14	.302	.047	0.66
	0.75	1.75	.748	.036	6.86	.362	.062	20.61	.706	.030	0.86	.308	.045	2.71
	0.75	3.75	.753	.037	7.57	.361	.063	20.26	.712	.030	1.71	.306	.045	1.87
	1.25	1.75	.745	.035	6.43	.367	.059	22.18	.704	.029	0.57	.312	.046	4.06
	1.25	3.75	.756	.035	8.00	.366	.063	22.07	.714	.030	2.00	.311	.048	3.57
1,000	0.00	0.00	.723	.023	3.29	.328	.039	9.40	.701	.021	0.14	.301	.034	0.26
	0.75	1.75	.728	.023	4.00	.331	.036	10.18	.706	.021	0.86	.302	.030	0.63
	0.75	3.75	.733	.023	4.71	.333	.039	11.07	.711	.022	1.57	.304	.033	1.47
	1.25	1.75	.725	.022	3.57	.339	.039	13.02	.704	.021	0.57	.311	.035	3.68
	1.25	3.75	.736	.023	5.14	.337	.037	12.48	.713	.021	1.86	.309	.031	2.87

Note. Factor loading population value = .7, factor correlation population value = .3. Dashes indicate values that are not available because of noninvertible weight matrices. s = skewness; k = kurtosis; WLS = weighted least squares; RB = mean relative bias across replications. For loadings, *M* is the pooled mean within-cell factor loadings; *SD* is the square root of within-cell factor loading variances.

variables, although these researchers found a greater effect of nonnormality than we did here.

However, it is crucial to keep in mind that in the current study, we created nonnormal ordinal observed data by categorizing nonnormal continuous latent response variables. This manipulation is fundamentally different from that implemented in these other studies in which nonnormal ordinal observed data were created by varying the thresholds used to categorize normal continuous latent response variables. As illustrated in Table 1, the nonnormal $y*$ variables from which we generated our sample data had more extreme levels of skewness and kurtosis than the observed, ordinal variables from which polychoric correlations and CFAs were estimated. Indeed, the levels of skewness and kurtosis in our ordinal observed data were quite close to those of the normal distribution, despite that our continuous latent response variables were much more nonnormal.

As such, observed ordinal distributions obtained in practice may often have more extreme levels of skewness and kurtosis than those used here. Because prior studies (e.g., Babakus et al., 1987; Hutchinson & Olmos, 1998; Potthast, 1993; Rigdon & Ferguson, 1991) have effectively demonstrated the effects of increased skewness and kurtosis in observed ordinal variables on the estimation of CFAs, we deemed a more thorough manipulation of y variables to be beyond the scope of the current study. Rather, as stated above, our intent was to evaluate the effect of violation of a crucial theoretical assumption for estimation of CFAs using polychoric correlations, namely the latent normality assumption for $y*$, and the manipulations we chose for our simulations were explicitly targeted to do so. Thus, combining our findings with those from previous studies, we were able to reach three general conclusions about how distribution of $y*$ and y affects the estimation of CFA models from observed ordinal data.

First, estimation of CFA models is robust to moderate violation of the latent normality assumption for $y*$ variables, an assumption implicit in the statistical theory underpinning the polychoric correlation, at least under conditions to those studied here. Because polychoric correlations provide robust estimates of the true correlation even when different sets of thresholds are applied to $y*$ variables, it follows that estimation of CFA models is not substantially affected according to whether or not threshold sets are constant across indicators.

Second, to the extent that the observed ordinal variables have nonzero skewness and kurtosis (e.g., as a result of threshold sets that lead to a dramatically different distribution shape for y relative to a normal or moderately nonnormal $y*$), full WLS estimation is known to produce biased chi-square test statistics and parameter standard error estimates. This latter finding likely occurs because of an increased tendency for low expected cell frequencies in observed contingency tables, especially in the context of

small-to-moderate sample sizes (e.g., fewer than 1,000; Potthast, 1993).[15]

Third, when the population $y*$ variables are of extreme nonnormality (e.g., skewness = 5, kurtosis = 50), the likely result is that the observed ordinal variables themselves will also have exaggerated levels of skewness and kurtosis, thus again leading to low expected frequencies in observed contingency tables. With regard to consideration of the joint effects of underlying nonnormality and varying thresholds across indicators, to the extent that these factors jointly produce observed contingency tables with low (or zero) expected cell frequencies, they are likely to lead to inaccurate polychoric correlations (as shown by Brown & Bendetti, 1977), which in turn adversely affect estimation of CFA models.

Implications for Applied Research

There are several specific implications of our findings with respect to using these analytic methods in practice. First, our findings suggest that the estimation of CFA models using polychoric correlations is robust to the moderate levels of nonnormality in the latent response variables that we considered here. Consistent with Quiroga (1992), our results showed that polychoric correlations become only slightly inflated when the latent response variables are moderately nonnormal. In turn, this bias in the correlation estimates contributes to a modest overestimation of CFA model parameters and has little effect on chi-square test statistics or parameter standard error estimates. Our findings support Pearson and Heron's (1913) argument that the latent normality assumption is merely a mathematical convenience that has little practical importance when the latent response variables are moderately (but not extremely) nonnormal.

Although our study demonstrates robustness to the latent normality assumption, it is important to stress that our study does not offer a thorough assessment of the effects of skewness and kurtosis of the observed ordinal variables. In

---

[15] An anonymous reviewer made the astute observation that for many cells of the study, chi-square values were more biased in the condition of $y*$ skewness = 1.75 and kurtosis = 1.75 than in the condition of skewness = 1.75 and kurtosis = 3.75. This pattern was likely the result of the particular combination of skewness and kurtosis in the observed ordinal variables rather than in the latent $y*$ variables. Specifically, Table 1 reflects that the former distributional condition had observed skewness of 0.486 in the context of slightly negative kurtosis, whereas the latter distribution was slightly less skewed in the context of positive kurtosis. In the present study, our aim was to illustrate robustness to the latent normality assumption rather than to investigate the effect of particular combinations of skewness and kurtosis in the observed variables. This finding offers partial evidence that nonnormality in the latent $y*$ variables per se does not affect estimation to the same extent as the level of nonzero skewness and kurtosis in the observed ordinal variables.

practice, researchers may likely observe ordinal distributions that are more skewed and kurtotic than those examined here. We refer applied researchers to prior studies (e.g., Babakus et al., 1987; Hutchinson & Olmos, 1998; Potthast, 1993; Rigdon & Ferguson, 1991) to further understand the effects of high skewness and kurtosis among observed ordinal variables on estimation of CFAs using polychoric correlations.

Consistent with previous studies for both continuous and ordinal data, our results demonstrate that for CFA models of realistic size (e.g., with 10 or more indicators), the desirable asymptotic properties of full WLS estimation are not observed with the types of sample sizes typically encountered in applied psychological research, even with $N = 1,000$. In the situation where a researcher wishes to fit a large model with $N = 1,000$ or fewer, our findings imply that when proper CFA solutions are obtained with full WLS (which can be quite rare at small-to-moderate sample sizes), these usually have inflated chi-square test statistics and parameter estimates and negatively biased standard errors. In contrast, robust WLS estimation nearly always produces a proper solution with test statistics, parameter estimates, and standard errors that are much less vulnerable to the effects of increasing model size and decreasing sample size. Furthermore, even for the situations in which full WLS estimation performs well (i.e., with small models and large sample size), robust WLS still produces less biased parameter estimates, standard errors, and test statistics. These results support the recommendation that applied researchers closely consider robust WLS estimation for CFAs with ordinal scales (and for SEMs more generally), particularly when testing medium-to-large models with a moderate-to-small sample size.

Given this recommendation, a word of caution is warranted. Despite its apparent finite-sample superiority to full WLS estimation, robust WLS estimation still leads to slightly biased test statistics and standard errors when large models are estimated with small samples. This inflation of the test statistic increases Type I error rates for the chi-square goodness-of-fit test, thereby causing researchers to reject correctly specified models more often than expected. It may be that researchers can supplement the chi-square goodness-of-fit test with other fit indices often computed in applications of SEM to retain a model that would otherwise be rejected on the basis of the chi-square goodness-of-fit test alone, although we did not include such measures as formal outcomes of this study. Examples of such fit indices include the comparative fit index (Bentler, 1990) and the root-mean-square error of approximation (Steiger, 1990). However, because both of these indices are based in part on the sample value of the fit function (i.e., Equation 8), these are likely to be biased to some degree as well. There has been little research explicitly evaluating the performance of these statistics in the context of CFA with polychoric correlations, although Hutchinson and Olmos (1998) reported promising initial findings for the comparative fit index.

Study Limitations and Directions for Future Research

We believe that the design of our study allows for empirically informed conclusions about the practical usefulness of CFA using polychoric correlations across several model specifications of increasing complexity, a broad range of sample sizes, several distributions for the latent response variables, and two types of WLS model-fitting procedures. However, this study shares the basic limitation of all Monte Carlo simulations; namely, it is possible that the findings cannot be generalized beyond the specific conditions studied here. Nonetheless, we believe that our experimental design and subsequent findings enable stronger predictions to be made about the finite-sample performance of these methods across a broad range of realistic conditions than have been previously available.

We defined our nonnormal underlying distributions by selecting levels of skewness and kurtosis to represent what we propose to be minor-to-moderate departures from normality that might be commonly encountered in behavioral research. It is important to reiterate that we were not interested in exploring how severely nonnormal the underlying distributions would need to be made for these estimators to fail. Instead, we feel that our results provide strong empirical evidence that robust WLS performs exceptionally well across a variety of commonly encountered conditions in applied research, and full WLS works almost as well for larger sample sizes and for simpler models. In sum, we are not concluding that robust or full WLS is unaffected by any degree of nonnormality of the latent response, but we do conclude that these methods are well-behaved for a variety of nonnormal distributions that might be expected in practice. Future research is needed to determine more accurately at what point full or robust WLS estimation begins to break down under violations of latent normality. Furthermore, we again remind readers that our results pertain most directly to violation of the normality assumption for latent response variables, a crucial theoretical assumption for the estimation of polychoric correlations. This is in distinct contrast to prior research that has addressed the effects of varying skewness and kurtosis among observed ordinal data in the context of normal latent response variables. We believe our findings augment this prior work in important ways.

Given the scarcity of studies that have explicitly assessed the application of CFA using polychoric correlations to realistic CFA models for ordinal variables, we felt that it was important to focus exclusively on the estimation of properly specified models. However, in practice, a given model specification is rarely exactly correct (see, e.g., Cudeck & Henly, 2003; MacCallum, 1995). Therefore, it is important that future research assess the ability of CFA with polychoric correlations to test misspecified models. In particular, our study provides promising results in support of the robust WLS method of B. Muthén et al. (1997) when

applied to correctly specified models. However, the performance of this method for the estimation of misspecified models is still unknown. Because this method calculates test statistics and standard errors in a manner similar to the Satorra and Bentler (1986, 1988) method for continuous observed data, one might predict from the results of Curran et al. (1996) that robust WLS would produce significantly underestimated chi-square statistics under model misspecification, thus reducing power to reject a misspecified model. Again, future research is needed to determine whether robust WLS maintains appropriate statistical power to detect a misspecification under violations of distributional assumptions.

Finally, we considered homogeneous values for the factor loadings both within and across models. This is of course a simplification, and it would be interesting to examine the more realistic situation of heterogeneous factor loading values in future research. Although we did not empirically examine the effects of varying the values of factor loadings here, we can draw on statistical theory to make some predictions about this issue. Namely, higher factor loading values are reflective of greater factor determinacy, a condition that has many benefits in model estimation and testing (see MacCallum, Widaman, Zhang, & Hong, 1999, for an excellent review). Given our focus of interest, we would predict that larger factor loadings would likely serve to convey information more saliently about violation of the underlying normality distributions within the CFA. Similarly, weaker loadings would likely dampen these effects. Taken together, we would expect that a heterogeneous set of factor loading values would simultaneously exacerbate and weaken the effect of the violation of assumptions, the specific effect of which would depend on many other experimental design features. Thus, although future research is needed to better understand the influences of heterogeneous loading structures, we do not believe that our use of homogenous loadings is a limitation here.

Conclusion

In sum, we make the following conclusions based on our experimental design and associated empirical findings. First, with the exception of a small number of modest differences in accuracy of polychoric correlation estimation and model convergence under full WLS, there were few to no differences found in any empirical results as a function of two-category versus five-category ordinal distributions. Second, moderate nonnormality of latent response distributions did not significantly effect the accuracy of estimation of polychoric correlations, although severe nonnormality did. Third, full WLS rarely resulted in accurate test statistics, parameter estimates, and standard errors under either normal or nonnormal latent response distributions, and this accuracy only occurred at large sample sizes and for less complex models. Fourth, robust WLS resulted in accurate test statistics, parameter estimates and standard errors under

both normal and nonnormal latent response distributions across all sample sizes and model complexities studied here (although there was modest bias found at the smallest sample size). Fifth, as we studied four variations of a CFA model here, we would anticipate that these findings would generalize to SEMs of comparable complexity. Finally, all of our conclusions are based on proper model specifications, and future research must address the role of statistical power and the ability of full and robust WLS to detect misspecifications when such misspecifications truly exist.

References

Babakus, E., Ferguson, C. E., & Jöreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research, 37,* 72–141.

Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin, 107,* 238–246.

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

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Brown, M. B., & Bendetti, J. K. (1977). On the mean and variance of the tetrachoric correlation coefficient. *Psychometrika, 42,* 347–355.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141). Cambridge, England: Cambridge University Press.

Browne, M. W. (1984). Asymptotic distribution free methods in the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37,* 127–141.

Chen, F., Bollen, K., Paxton, P., Curran, P. J., & Kirby, J. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods and Research, 29,* 468–508.

Chou, C.-P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 37–54). Thousand Oaks, CA: Sage.

Chou, C.-P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology, 44,* 347–357.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105,* 317–327.

Cudeck, R., & Henly, S. J. (2003). A realistic perspective on pattern representation in growth data: Comment on Bauer and Curran (2003). *Psychological Methods, 8,* 378–383.

Curran, P. J., West, S. G., & Finch, G. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1,* 16–29.

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable

estimators using simulated data. *British Journal of Mathematical and Statistical Psychology, 47,* 309–326.

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

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

Hu, L.-T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112,* 351–362.

Hutchinson, S. R., & Olmos, A. (1998). Behavior of descriptive fit indexes in confirmatory factor analysis using ordered categorical data. *Structural Equation Modeling, 5,* 344–364.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34,* 183–202.

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika, 59,* 381–389.

Jöreskog, K. G., & Sörbom, D. (1988). *PRELIS: A program for multivariate data screening and data summarization. A preprocessor for LISREL* (2nd ed.). Mooresville, IN: Scientific Software.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's reference guide.* Chicago: Scientific Software.

Kaplan, D. (1989). A study of the sampling variability of the *z*-values of parameter estimates from misspecified structural equation models. *Multivariate Behavioral Research, 24,* 41–57.

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

MacCallum, R. C. (1995). Model specification: Procedures, strategies, and related issues. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 16–36). Thousand Oaks, CA: Sage.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4,* 84–99.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105,* 156–166.

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics, 22,* 48–65.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49,* 115–132.

Muthén, B., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes.* Unpublished manuscript.

Muthén, B., & Hofacker, C. (1988). Testing the assumptions underlying tetrachoric correlations. *Psychometrika, 53,* 563–578.

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor-analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38,* 171–180.

Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor-analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology, 45,* 19–30.

Muthén, B. O., & Satorra, A. (1995). Technical aspects of Muthén's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika, 60,* 489–503.

Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide.* Los Angeles: Author.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2002). *Mx: Statistical modeling* (6th ed.). Richmond, VA: Department of Psychiatry, Virginia Commonwealth University.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44,* 443–460.

Pearson, K. (1901). Mathematical contributions to the theory of evolution, VII: On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London, Series A, 195,* 1–47.

Pearson, K., & Heron, D. (1913). On theories of association. *Biometrika, 9,* 159–315.

Potthast, M. J. (1993). Confirmatory factor analysis of ordered categorical variables with large models. *British Journal of Mathematical and Statistical Psychology, 46,* 273–286.

Quiroga, A. M. (1992). *Studies of the polychoric correlation and other correlation measures for ordinal variables.* Unpublished doctoral dissertation, Acta Universitatis Upsaliensis.

Rigdon, E. E., & Ferguson, C. E. (1991). The performance of the polychoric correlation coefficient and selected fitting functions in confirmatory factor analysis with ordinal data. *Journal of Marketing Research, 28,* 491–497.

Satorra, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. In P. Marsden (Ed.), *Sociological methodology 1992.* Washington, DC: American Sociological Association.

Satorra, A., & Bentler, P. M. (1986). Some robustness properties of goodness of fit statistics in covariance structure analysis. *1986 ASA Proceedings of the Business and Economic Section,* 549–554.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *1988 ASA Proceedings of the Business and Economic Section,* 308–313.

Satorra, A., & Bentler, P. M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Computational Statistics and Data Analysis, 10,* 235–249.

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika, 6,* 309–316.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton-Mifflin.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25,* 173–180.

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika, 48,* 465–471.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with non-normal variables: Problems and remedies. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56–75). Thousand Oaks, CA: Sage.

Yule, G. U. (1912). On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society, 75,* 579–652.

New Editors Appointed, 2006–2011

The Publications and Communications Board of the American Psychological Association announces the appointment of seven new editors for 6-year terms beginning in 2006. As of January 1, 2005, manuscripts should be directed as follows:

- *Experimental and Clinical Psychopharmacology* (www.apa.org/journals/pha.html), **Nancy K. Mello, PhD,** McLean Hospital, Massachusetts General Hospital, Harvard Medical School, 115 Mill Street, Belmont, MA 02478-9106.

- *Journal of Abnormal Psychology* (www.apa.org/journals/abn.html), **David Watson, PhD,** Department of Psychology, University of Iowa, Iowa City, IA 52242-1407.

- *Journal of Comparative Psychology* (www.apa.org/journals/com.html), **Gordon M. Burghardt, PhD,** Department of Psychology or Department of Ecology & Evolutionary Biology, University of Tennessee, Knoxville, TN 37996.

- *Journal of Counseling Psychology* (www.apa.org/journals/cou.html), **Brent S. Mallinckrodt, PhD,** Department of Educational, School, and Counseling Psychology, 16 Hill Hall, University of Missouri, Columbia, MO 65211.

- *Journal of Experimental Psychology: Human Perception and Performance* (www.apa.org/journals/xhp.html), **Glyn W. Humphreys, PhD,** Behavioural Brain Sciences Centre, School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom.

- *Journal of Personality and Social Psychology: Attitudes and Social Cognition* section (www.apa.org/journals/psp.html), **Charles M. Judd, PhD,** Department of Psychology, University of Colorado, Boulder, CO 80309-0345.

- *Rehabilitation Psychology* (www.apa.org/journals/rep.html), **Timothy R. Elliott, PhD,** Department of Psychology, 415 Campbell Hall, 1300 University Boulevard, University of Alabama, Birmingham, AL 35294-1170.

Electronic submission: As of January 1, 2005, authors are expected to submit manuscripts electronically through the journal's Manuscript Submission Portal (see the Web site listed above with each journal title).

Manuscript submission patterns make the precise date of completion of the 2005 volumes uncertain. Current editors, Warren K. Bickel, PhD, Timothy B. Baker, PhD, Meredith J. West, PhD, Jo-Ida C. Hansen, PhD, David A. Rosenbaum, PhD, Patricia G. Devine, PhD, and Bruce Caplan, PhD, respectively, will receive and consider manuscripts through December 31, 2004. Should 2005 volumes be completed before that date, manuscripts will be redirected to the new editors for consideration in 2006 volumes.