STRUCTURAL EQUATION MODELING, *9*(1), 78–102

The Effects of Item Parceling on Goodness-of-Fit and Parameter Estimate Bias in Structural Equation Modeling

Deborah L. Bandalos
Department of Educational Psychology
University of Nebraska, Lincoln

Two simulation studies were conducted to investigate the effects of the practice of item parceling. In Study 1, unidimensional sets of normally and nonnormally distributed item-level data were categorized into 2-, 3-, and 4-item parcels. Analyses revealed that the use of item parcels resulted in better fitting solutions, as measured by the root mean squared error of approximation (RMSEA), comparative fit index (CFI), and chi-square test, when items had a unidimensional structure. Parceled solutions also resulted in less bias in estimates of structural parameters under these conditions than did solutions based on the individual items. In Study 2 the issue of whether the use of item parceling could mask a known multidimensional factor structure among a set of items was investigated. Results indicated that certain types of item parceling can obfuscate a multidimensional factor structure in such a way that acceptable values of fit indexes are found for a misspecified solution. In addition, parceling under these conditions was found to result in bias in the estimates of structural parameters. Although parceling can ameliorate the effects of coarsely categorized and nonnormally distributed item-level data when the items are unidimensional, the use of parceling with items that are multidimensional or for which the factor structure is unknown cannot be recommended.

The use of item parcels in structural equation modeling (SEM) has become quite common in recent years. Parceling involves summing or averaging item scores from two or more items and using these parcel scores in place of the item scores in a SEM analysis. A recent review of SEM applications (Bandalos & Finney, 2001)

Requests for reprints should be sent to Deborah L. Bandalos, University of Nebraska, Lincoln, Educational Psychology, 122 Bancroft Hall, P.O. Box 880345, Lincoln, NE 68588–0345. E-mail: dbandalos@unl.edu

examined issues of the publications *Journal of Educational Measurement, Journal of Educational Psychology, Applied Psychological Measurement, American Educational Research Journal, Educational and Psychological Measurement, Structural Equation Modeling,* and *Journal of Marketing Research* from 1989 to the present (issues of *Structural Equation Modeling* were examined from its inception in 1994 to the present) and found that, of 317 applied SEM or confirmatory factor analysis (CFA) studies, 62, or 19.6%, employed some type of parceling procedure. Of these, the majority (82.3%) were CFA applications. The technique of parceling or bundling items appears to have originated in the work of Cattell (1956; Cattell & Burdsall, 1975) and has been adopted by researchers in such areas as education (Cook, Dorans, & Eignor, 1988); psychology (Russell, Kahn, Spoth, & Altmaier, 1998; Schau, Stevens, Dauphinee, & Del Vecchio, 1995); marketing (Singh & Rhoads, 1991); and organizational research (Bagozzi & Edwards, 1998; Bagozzi & Heatherton, 1994).

Item parceling is often used for situations in which the data to be analyzed are nonnormally distributed, coarsely categorized, or both—conditions that violate the assumptions on which normal theory maximum likelihood and generalized least squares estimation techniques are based. Several studies have investigated the robustness of normal theory estimators to violations of normality (Finch, West, & MacKinnon, 1997; Harlow, 1985; Hutchinson & Olmos, 1998; Muthén & Kaplan, 1985, 1992) and coarse categorization (Babakus, Ferguson, & Jöreskog, 1987; Green, Akey, Fleming, Hershberger, & Marquis, 1997; Hutchinson & Olmos, 1998; Muthén & Kaplan, 1985, 1992). These studies found that, in the presence of nonnormality, parameter estimates are typically unbiased, but values of the chi-square test statistic and other fit indexes are adversely affected, and standard errors become attenuated. Under coarse categorization, chi-square values are typically found to be inflated when only two response categories are used, but this bias decreases with increasing numbers of categories. This bias is exacerbated for situations in which the distributions of the categorized variables are nonnormal, with opposite skew producing the worst results.

To mitigate these effects, item parceling has been adopted in many empirical studies as a means of obtaining item distributions that are more continuous and normally distributed. For example, parceling has been adopted in an attempt to circumvent problems with so called "difficulty factors" commonly found in factor analyses of dichotomously scored items, such as those typically found on achievement and aptitude tests. Additional advantages of item parceling have been suggested in other areas. In organizational research, Bagozzi and his colleagues (Bagozzi & Edwards, 1998; Bagozzi & Heatherton, 1994) have suggested that the use of parceling (referred to in their articles as a "partial disaggregation model") results in the estimation of fewer model parameters and will therefore result in a more optimal variable to sample size ratio and more stable parameter estimates, particularly with small samples. However, the assumption that smaller parameter to sample size ratios will necessarily result in greater stability of parameter estimates has been called into question by recent studies (e.g., MacCallum, Widaman, Zhang, & Hong, 1999; Marsh, Hau, Balla, & Grayson, 1998). Researchers have also cited such advantages as greater reliability and more definitive rotational results (Cattell & Burdsall, 1975; Kishton & Widaman, 1994) for parceled solutions. Finally, some authors argue for the use of parceling on the grounds that parceled solutions will typically result in better model fit than solutions at the item level (e.g., Thompson & Melancon, 1996).

Despite these advantages, the use of item parceling is not without controversy. Perhaps most important, the use of parceling depends on the unidimensionality of the items being combined, an assumption that is often not tested. In the Bandalos and Finney (2001) review, only 32.3% of published studies made any reference to the unidimensionality of the items being parceled. When this assumption is not met, the use of parcels can obscure rather than clarify the factor structure of the data (Hall, Snell, & Singer Foust, 1999; West, Finch, & Curran, 1995). It has also been found that the use of parceling can result in biased estimates of other model parameters (Hall et al., 1999). Finally, because the use of item parcels has the effect of reducing the number of data points that must be fit, solutions based on parcels will not yield as stringent a test of SEM models as would analyses based on the individual items.

PREVIOUS RESEARCH ON PARCELING

Despite the widespread use of parceling and the questions regarding its use, few empirical studies have addressed the issues discussed here. Of these, most have utilized actual data sets for which the population factor structure is unknown, making it difficult to determine the impact of parceling on the recovery of the true factor structure or on the accuracy of parameter estimates. In several of these studies, it was demonstrated that the use of item parcels as opposed to individual items resulted in better model–data fit (Bagozzi & Heatherton, 1994; Gribbons & Hocevar, 1998; Takahashi & Nasser, 1996; Thompson & Melancon, 1996). It was also found that solutions based on parcels containing more items resulted in a greater improvement in fit than those based on parcels with fewer items. However, this effect may have been due to the fact that solutions based on more items also included fewer parcels. Michael and Bachelor (1988) examined factor solutions based on exploratory factor analysis and found that, although the factor solutions produced by the item level and parcel level analyses were not completely consistent, the outcomes of these two analyses were likely to be similar "if the scales of the measures are quite homogenous" (p. 102). However, inspection of their results reveals some notable differences between the two solutions. Finally, Bagozzi and Edwards

(1998) found that the results of multiple-group comparisons based on item parcels differed from those based on the individual items.

In contrast to the studies reviewed above, in which actual data sets were used to demonstrate various features of parceled solutions, three studies have investigated the effects of parceling either experimentally or analytically. Marsh et al. (1998), using simulated data, found that CFA solutions based on two (six-item), three (four-item), four (three-item), or six (two-item) parcels resulted in greater numbers of proper solutions than analyses based on two, three, four, or six individual items. However, solutions based on all 12 individual items resulted in proper solutions for all samples. The chi-square–degrees of freedom ratio increased with the number of parcels used and was highest for solutions based on the individual items. The results with regard to proper solutions are consistent with those of Yuan, Bentler, and Kano (1997), who showed analytically that CFA solutions based on parcels had greater power and smaller mean squared error than those based on individual items when the numbers of items used was equivalent to the number of parcels.

In a recent study by Hall et al. (1999), both simulated and empirical data sets were used to demonstrate the effects of parceling on parameter estimates and goodness of fit. Of particular interest in this study was the effect of misspecifying a model in such a way that items that double-loaded onto a secondary factor were put into parcels. These parcels were then forced to load only on the primary factor. In this way the secondary factor was essentially omitted from the model. The parceling was conducted in two ways. In the first, referred to as *isolated parceling,* the items with secondary loadings were combined into the same parcel, whereas in the second strategy, termed *distributed parceling,* the items with secondary loadings were put into parcels with items that did not share the influence of the secondary factor. Results showed that the influence of the type of parceling depended on whether the secondary factor was related to a third, endogenous factor in the model. If the secondary factor did not influence the third factor, the goodness of fit for the solutions based on the distributed strategy was not as good as that for the isolated strategy. However, if the secondary factor did influence the third factor, the fit for the model based on distributed parceling was superior to that for the model in which the isolated strategy was used. In the latter case, the estimate of the path from the primary to the third factor was biased for both parceling strategies, although slightly more bias was found for the distributed strategy.

Hall et al. (1999) explained these results in terms of the treatment of the variance resulting from the secondary factor. When items that were influenced by the secondary factor were put into separate parcels, the influence of the secondary factor became common to two of the parcels. This source of variation thus became shared variance and was reflected in higher loadings on the primary factor for those parcels. This strategy thus allowed the variance associated with the secondary factor to be absorbed into the primary factor. For the situation in which the secondary factor was related to the third factor, the path from the primary factor to the third factor increased under the distributed uniqueness strategy because it reflected the variance the third factor shared with both the primary and secondary factors. In this situation, the model fit well even though misspecified because the secondary factor was still able to affect the third factor, although somewhat indirectly. In contrast, under the isolated strategy, the variance associated with the secondary factor was isolated into one parcel and thus became not shared but error variance. Under this scenario, the effect of the secondary on the third factor had no outlet, resulting in a greater lack of fit.

THEORETICAL FRAMEWORK

A recent article by MacCallum et al. (1999) provided a theoretical framework for the results found in the Hall et al. (1999) study. Although the purpose of the study was to investigate the relation between the variable-to-factor ratio, level of communality, and sample size in the context of exploratory factor analysis, this study provides a general framework that underlies both exploratory and confirmatory factor analyses. Using a framework adapted from an earlier article by MacCallum and Tucker (1991), MacCallum et al. showed that there are two sources of sampling error that come into play when a sample covariance matrix is used as an estimator of the population parameters of a factor analysis model. The first of these effects relates to covariance among the unique factors and among the unique and common factors. The population factor analysis model can be represented by

$$\Sigma_{yy} = \Lambda\Phi\Lambda' + \Theta^2 \tag{1}$$

in which Σ_{yy} represents the population covariance matrix of the observed variables y, Λ is a matrix of common factor loadings, Φ is a matrix of correlations or covariances among the common factors, and Θ is a matrix of unique factor variances. In the population, unique factors are typically assumed to be uncorrelated with each other and with the common factors. However, such correlations can result from sampling error. Thus, the sample counterpart to Equation 1 would be

$$C_{yy} = \Lambda C_{cc}\Lambda' + \Lambda C_{cu}\Theta' + \Theta C_{uc}\Lambda' + \Theta C_{uu}\Theta' \tag{2}$$

where C_{yy} is a sample covariance matrix of the observed variables y, C_{cc} is a matrix of the sample correlations or covariances among the common factors, C_{cu} is a matrix of sample correlations or covariances among common and unique factors, and C_{uu} is a matrix of sample correlations or covariances among the unique factors. Although in the population C_{cu} and C_{uc} are typically assumed to be zero matrices and C_{uu} is assumed to be diagonal, this may not be the case in the sample. These nonzero elements of C_{cu}, C_{uc},, and C_{uu} will result in a lack of fit of the population model of Equation 1 to the sample data. In applications of CFA this lack of fit would typically be manifested in non-zero elements in the off-diagonal of C_{uu}.

Equation 2 illustrates the second effect of sampling error described by MacCallum et al. (1999). Note that the unique loadings in Θ serve as weights for the matrices C_{cu}, C_{uc}, and C_{uu}. When common factor loadings or communalities are high, these unique loadings will be low, and the elements in C_{cu}, C_{uc}, and C_{uu} will not have as great an impact on the solution for C_{yy}. Thus, with more reliable measures, the effects of the sampling error that results in nonzero elements in these matrices will be minimized and the solution will be more stable. Thus with more reliable measures a smaller sample size may be sufficient to obtain good fit. This finding is consistent with previous research (e.g., Bandalos, 1993; Velicer & Fava, 1998) and suggests that the efficacy of parceling items to reduce the variable-to-factor ratio will be mediated by the communalities of the items. For situations in which the communalities are low, parceling will have a greater effect on fit, whereas with high item communalities, parceling may not be necessary to obtain stable estimates, even when sample size is low.

RELEVANCE TO ITEM PARCELING

In the context of item parceling, the framework provided by MacCallum et al. (1999) is relevant in that it provides an explanation for the improvement in model fit associated with parceled solutions. With regard to the first sampling effect described by MacCallum et al., one result of parceling will be to reduce the size of the matrices C_{uu}, C_{cu}, and C_{uc}, with a consequent reduction in the contribution from error resulting from unmodeled associations in these matrices. In addition, as shown by Hall et al. (1999), parceling can be done in such a way that items with shared secondary influences, which would typically result in correlated uniquenesses or nonzero diagonal elements in C_{uu}, are parceled together. When these parcels are then treated as indicators of the same factor, the correlated uniqueness is reformulated as shared common variance and becomes part of the modeled variance in the diagonal of C_{cc} rather than of unmodeled associations in the off-diagonal elements of C_{uu}.

With regard to MacCallum et al.'s (1999) second sampling effect, the size of the elements of Θ will have an effect on the fit of the model through their use as weighting elements for the matrices C_{cu}, C_{uc}, and C_{uu}, as seen in Equation 2. Because parcels are based on more than one item, they will typically be more reliable than individual items, so that one effect of parceling is to reduce the size of the elements of Θ. Another reason that parceling improves model–data fit is therefore due to the reduction in the impact of the matrices C_{cu}, C_{uc}, and C_{uu} on the solution.

The use of parceling can thus be seen to reduce the lack of fit of C_{yy} to the model implied by Equation 1 in three complementary ways: by reducing the size of the matrices C_{uu}, C_{cu}, and C_{uc}; by reducing the contribution of these elements to the sample matrix C_{yy}; and by reformulating variance due to correlations among unique factors, common factors, or both. The first two effects will occur to some degree for any type of parceling, regardless of how items are parceled together or whether secondary factors are involved. Note that these sources of error reduction function independently of each other. When all effects are strong, a substantial improvement in fit would be expected over the corresponding item level solution.

Although the studies cited previously provide some information with regard to the use of parceling, they do not address the issue of whether the use of parceling is effective in overcoming problems associated with nonnormally distributed, coarsely categorized data. Prior research suggests this will be the case, and this investigation includes a simulation study with three types of item distributions crossed with two, three, and four category data at five sample sizes to examine these effects in a more systematic manner than has been done previously. It was expected that parceling would have the greatest effect on fit for items with severely nonnormal distributions and a small number of categories and that this effect would be most evident at smaller sample sizes. A second simulation was designed to investigate whether, and to what extent, the use of parceling can obscure the factor structure of a set of multidimensional items. Given the theoretical framework provided by MacCallum et al. (1999) and the simulations of Hall et al. (1999), it would appear that this is possible. This study extends the work of Hall et al. by including a different model and by systematically varying the number of indicators and parcels, as well as the sample size and method of parceling the items.

DATA GENERATION AND ANALYSIS

Study 1

Study 1 was designed to investigate the efficacy of parceling techniques for overcoming problems associated with the use of nonnormally distributed, coarsely categorized data in SEM. The procedures used to generate the data for this study were designed to mimic the situation in which a continuous scale is thought to exist in theory, but due to limitations of the measurement process only categorical scale points are observed, which are often nonnormally distributed. The basic model (Model 1) for Study 1 was a structural model with two exogenous latent variables, each measured by 12 items, and one endogenous latent variable measured by 6 items. This model is shown in Figure 1. Values of the measurement parameters are not shown in the figure; all factor loadings, including those for the endogenous latent variable, were set at .7, whereas all measurement error variances were set at .3.

Data for Studies 1 and 2 were generated independently as continuous random normal variates using the SAS RANNORM procedure. A Cholesky factorization

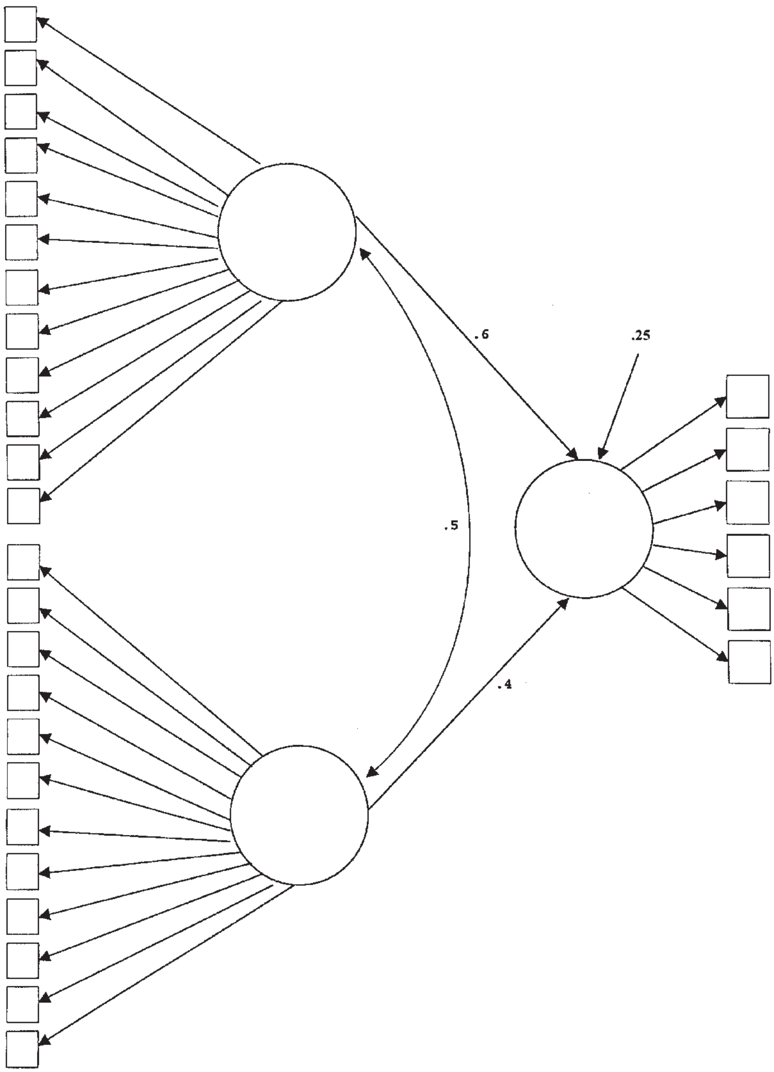

FIGURE 1 Model 1. Generating model for data in Study 1.

was used to impose the appropriate correlation structure on the data. The continuous data were then categorized into two, three, and four categories through the use of appropriate percentile cutpoints for Study 1. For each number of scale points, categorization was carried out in such a way that three types of distribution conditions (referred to in this article as "type") were formed. In Type 1 distributions, all categorized items were approximately normally distributed. (For the two-category condition, a uniform distribution was created, as a normal distribution was not possible.) In Type 2 and 3 distributions, item scores were categorized such that half were normally and half nonnormally distributed. For Distribution 2, values of skewness and kurtosis were approximately 3.5 and 11.5, respectively, whereas for Distribution 3, the corresponding values were approximately 5 and 25. These values were selected to represent levels of nonnormality extreme enough to prompt a researcher to employ parceling techniques.

Item-level data corresponding to the design factors described previously were generated using the SAS IML program for matrix procedures. Within each combination of number of categories and distribution condition, six 2-item, three 4-item, two 6-item parcels as well as one 12-item parcel (representing the use of a total scale score) were created from the 12 items for each of the exogenous factors. In addition, a condition in which the data were analyzed at the item level was included for the purposes of comparison. Items were assigned to parcels such that half of the items in each parcel were normally distributed and half were taken from one of the nonnormal distribution conditions. This procedure was designed to mimic that used by applied researchers in attempting to create parcel distributions that are more normally distributed than those of the individual items (e.g., Cook et al., 1988) by pairing skewed and normally distributed items within parcels. For Distribution 1, consisting of all normally distributed variables, the parceling procedure necessarily resulted in parcels in which all the variables were approximately normally distributed. This condition served as a baseline against which the effects of nonnormality could be evaluated. Finally, to study the effects of sample size and possible interactions of the design variables with sample size, data were generated at sample sizes of 100, 250, 400, 650, and 800 for each of the 45 (3 numbers of categories × 3 distribution conditions × 5 parcel types) cells described earlier.

Study 2

Study 2 was designed to investigate the question of whether the use of parceling could obscure a known true factor structure. For this study, the model used in Study 1 was altered so that the first six items from each exogenous factor also had secondary loadings on a third factor. This model is shown in Figure 2. The values of these secondary loadings were set at .4. All other parameter values remained the same as in Study 1. This model was analyzed by fitting a (misspecified) two-factor structure to the exogenous variables to investigate whether the use of item parcels would result in greater ambiguity regarding the presence of the secondary factor than would use of the individual items. For this study, two types of parcels were created. In parcel type 1, double-loading and single-loading items were parceled together, whereas in parcel type 2, double-loading items were parceled only with

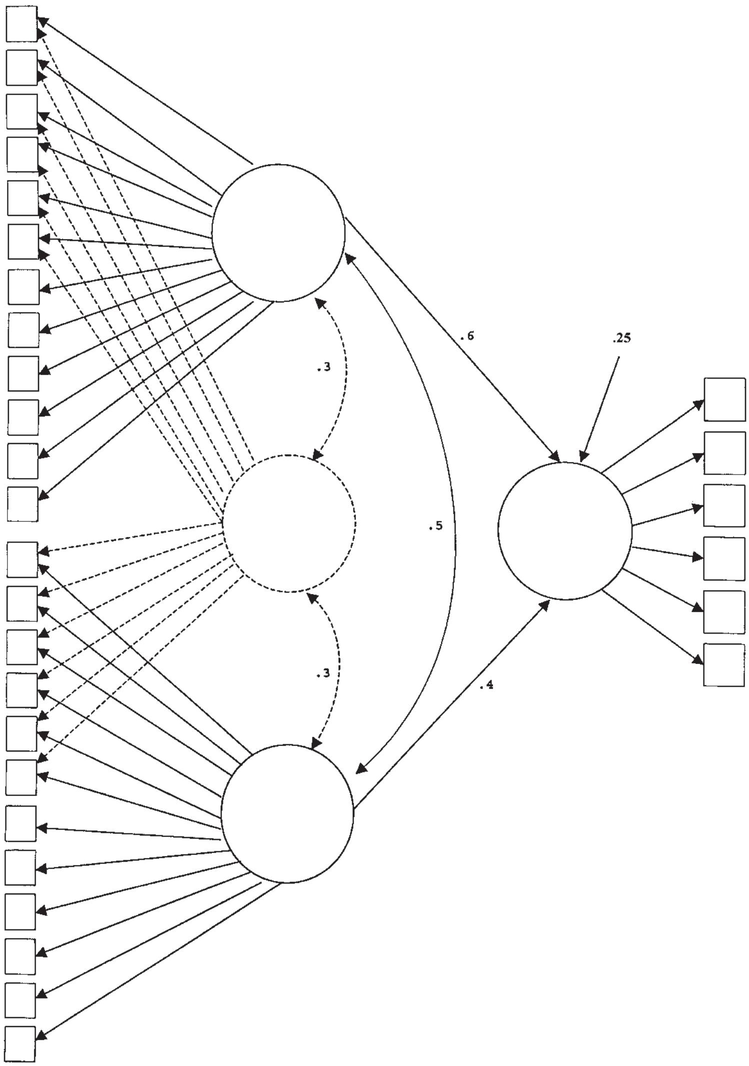

FIGURE 2 Model 2. Generating model for data in Study 2. Dashed lines represent parameters not estimated in the misspecified model.

other double-loading items and single-loading items were parceled only with single-loading items. These conditions correspond to Hall et al.'s (1999) distributed uniqueness and isolated uniqueness strategies, respectively, and are referred to as the "distributed" and "isolated" strategies. Regarding Figure 2, for two-item parcels, Items 1 and 7, 2 and 8, 13 and 19, 14 and 20, and so on would be parceled together in the distributed strategy; whereas Items 1 and 2, 3 and 4, 13 and 14, 15 and 16, and so on would be parceled together under the isolated strategy. Two- and three-item parcels were formed for each type of parcel under each of the five sample sizes used for Study 1. Models based on 12-item parcels (representing use of a total score) and on the individual items at each of the five sample sizes were also analyzed.

For each study, 150 replications for each cell in the design were generated using the SAS IML matrix subroutine. The data from each sample were then analyzed using the LISREL 8 program (Jöreskog & Sörbom, 1996). Parameter estimates and measures of overall fit (chi-square values, CFI, and RMSEA) for each sample were saved and analyzed.

RESULTS

Data within each study were analyzed using repeated measures analysis of variance techniques. Because of the large sample sizes, most effects were statistically significant even when the actual differences among conditions were extremely small. For this reason, measures of partial η^2 were used as measures of effect size. Only those effects that reached Cohen's (1988) large effect size of .14 were considered to be practically significant. Results for Studies 1 and 2 are presented separately in the following.

Study 1

Effects of parceling on parcel distributions. The effects of parceling on the resulting parcel distributions can be seen in Table 1, which shows the levels of skewness and kurtosis for the item level and parceled categorized data across 15,000 replications. Some variation from the target levels of skewness and kurtosis can be seen across the three types of categorization. The most notable feature of Table 1 is the degree to which skewness and kurtosis are reduced as a result of parceling items. In general, the reductions in skewness and kurtosis levels obtained by including more items in a parcel are fairly negligible.

Admissibility problems. Some problems with nonpositive definite input matrices resulted from categorized data. These problems were confined to the smallest sample size condition ($n = 100$) in which data were categorized in such a way as to result in moderately or severely nonnormal distributions. For the two-category data, 2% of cases in the moderately nonnormal condition and 49% of cases in the severely nonnormal condition resulted in input matrices that were nonpositive definite. For the three- and four-category data, 9% and 2%, respectively, of the samples in the se-

TABLE 1
Mean Skewness and Kurtosis Values: Study 1 (N = 15,000)

	CAT					
	2		*3*		*4*	
Distribution	*Skew*	*Kurtosis*	*Skew*	*Kurtosis*	*Skew*	*Kurtosis*
Unparceled						
1	.02	–2.00	–.01	.86	–.02	–.40
2	3.78	12.32	3.42	11.56	2.95	8.06
3	5.37	26.82	5.09	25.63	4.66	22.87
Two–item parcels						
1	.02	–1.59	.00	.82	.00	–.31
2	.54	–.62	1.31	3.27	.99	1.63
3	.41	–.88	1.48	5.03	.96	2.31
Four–item parcels						
1	.02	–1.38	–.01	.78	.00	–.29
2	.55	–.44	1.35	3.30	1.01	1.67
3	.42	–.62	1.48	4.85	.97	2.41
Twelve–item parcels						
1	.02	–1.38	–.01	.76	.00	–.31
2	.57	–.33	1.40	3.40	1.04	1.74
3	.44	–.49	1.52	4.90	1.02	2.56

Note. CAT = number of categories.

verely nonnormal condition resulted in nonpositive definite input matrices. These samples were therefore excluded from further analyses.

Rejection rates. Rejection rates for each sample were calculated based on a .05 level of significance. Table 2 shows the percentage of rejections by sample size, number of categories, and parcel type. Because the model is correct in the population, rejection rates in excess of 5% represent inflation of a Type I error rate of .05. A comparison of the rejection rates obtained from the continuous and normally distributed data with those obtained from coarsely categorized data, nonnormally distributed data, or both can be obtained by comparing the values in the first column of Table 2 with those on the right. These comparisons illustrate the expected negative impact of coarse categorization and nonnormality on Type I errors, which occurred more frequently for these types of data. As can be seen from the table, use of unparceled data with nonnormally distributed data (TYPE 2 or 3) resulted in rejection rates of 100% for all sample sizes and numbers of categories. All solutions in which items were parceled resulted in substantially lower rejection rates than unparceled solutions. This effect was most striking at the smaller sample sizes but was evident even with samples of 800. The number of items in a parcel had an inverse relationship with rejection rates: As the number of items increased

TABLE 2
Percentage of Model Rejections by *N*, TYPE, and CAT: Study 1

		CAT = 2			CAT = 3			CAT = 4		
	CAT = 0	TYPE 1	TYPE 2	TYPE 3	TYPE 1	TYPE 2	TYPE 3	TYPE 1	TYPE 2	TYPE 3
N = 100										
Unparceled	58	95	100[a]	100[b]	79	100	100[c]	73	100	100[d]
2-item parcels	20	37	33	34	24	51	51	23	33	35
4-item parcels	11	15	13	9	14	14	12	9	12	9
12-item parcels	8	6	7	9	9	9	11	7	10	8
N = 250										
Unparceled	13	71	100	100	36	100	100	22	100	100
2-item parcels	11	15	15	11	8	32	42	4	19	20
4-item parcels	9	11	10	11	5	9	10	7	9	81
2-item parcels	5	6	9	9	7	7	7	8	5	7
N = 400										
Unparceled	11	65	100	100	33	100	100	21	100	100
2-item parcels	9	18	11	11	13	37	49	11	20	22
4-item parcels	6	6	6	4	7	10	10	7	5	91
2-item parcels	1	5	5	4	4	5	3	3	3	4
N = 650										
Unparceled	6	52	100	100	27	100	100	16	100	100
2-item parcels	9	13	11	6	11	32	41	7	17	20
4-item parcels	9	4	9	7	7	10	13	5	8	11
12-item parcels	7	3	5	3	5	6	5	5	5	5
N = 800										
Unparceled	9	50	100	100	25	100	100	15	100	100
2-item parcels	4	10	8	7	9	35	46	7	20	17
4-item parcels	5	4	3	3	8	9	9	3	6	3
12-item parcels	3	3	5	5	5	4	4	3	6	6

Note. CAT = number of categories: 0 = continuous data. TYPE = type of parceling:1 = all normal distributions; 2 = half normally and half moderately nonnormally distributed items; 3 = half normally and half severely nonnormally distributed items.

[a]Percentage is based on a total of 147 cases. [b]Percentage is based on a total of 76 cases. [c]Percentage is based on a total of 136 cases. [d]Percentage is based on a total of 147 cases.

rejection rates decreased, with the lowest rejection rates resulting from the 12-item parcels. Although only solutions from the 12-item parcels resulted in rejection rates close to the nominal .05 error rate, the 4-item and, in some cases, the 2-item parcels also resulted in fairly acceptable rejection rates, especially given the levels of nonnormality and the coarseness of the scales involved. Finally, the effect of the number of categories on rejection rates was minimal, particularly for the 12-item parcels.

These results with regard to Type I error rates are illustrative of what would occur if a researcher were to base analyses on parceled rather than unparceled data. However, these solutions also differ in levels of power. Although power will generally be greater for solutions based on more "items," whether individual items or parcels, the greater reliability of parcels will also increase power to some extent. Because these two effects are operating simultaneously in this study, it is difficult to assess the degree to which the observed differences in rejection rates are affected by differences in power.

Overall goodness of fit. Values of the RMSEA and CFI were analyzed using repeated measures analysis of variance. These two indexes have recently been recommended for use (Hu & Bentler, 1998) because of their sensitivity to model misspecification. Factors or interactions with partial η^2 values of .14 or higher were considered to be practically significant. Mean values of these indexes are shown in Tables 3 and 4.

As would be expected, values of these indexes reached optimal values for solutions based on continuous and normally distributed data. However, this effect was mediated to some extent by sample size. With regard to the parceled solutions, the practice of parceling normally and nonnormally distributed items together appeared to have the expected effect of yielding better fitting solutions, regardless of the severity of the nonnormality. The number of items included in the parcels seems to have had little effect, particularly at the larger sample sizes, although values of partial η^2 reached the criterion level for this effect (.49 and .79, respectively, for the RMSEA and CFI, respectively). This effect appeared to be due mainly to the differences between the unparceled and the parceled data. For both the RMSEA and the CFI, the interaction of TYPE with the number of items in the parcel reached the partial η^2 squared criterion. Partial η^2 values were .35 and .71 for the RMSEA and CFI, respectively. This effect can be explained as follows: Goodness of fit for the unparceled data worsened with the introduction of more severely nonnormal data. For the parceled data, however, the introduction of nonnormal data had little effect on values of the fit indexes. Interactions of parcel size with sample size had a large effect on values of the CFI (partial η^2 = .46), reflecting the fact that values of this index increased with sample size for the unparceled data only. Finally, the interaction of parcel size with sample size and TYPE resulted in a partial η^2 value of .21 for the CFI. As can be

TABLE 3
Mean Values of the RMSEA by *N* and TYPE: Study 1

	Number of Items in Parcel			
Type of Data	*Unparceled*	*2*	*4*	*12*
N = 100				
Continuous[a]	.035	.028	.024	.027
TYPE[b] 1	.045	.033	.026	.027
2	.076	.038	.028	.028
3	.097	.040	.026	.028
N = 250				
Continuous	.012	.011	.014	.016
TYPE 1	.012	.013	.013	.015
2	.052	.019	.015	.015
3	.065	.019	.015	.015
N = 400				
Continuous	.008	.009	.009	.011
TYPE 1	.014	.011	.010	.011
2	.047	.014	.011	.011
3	.057	.015	.011	.011
N = 650				
Continuous	.005	.006	.008	.008
TYPE 1	.010	.008	.008	.009
2	.044	.011	.009	.008
3	.052	.011	.009	.008
N = 800				
Continuous	.005	.005	.006	.007
TYPE 1	.009	.006	.006	.007
2	.043	.009	.006	.007
3	.050	.010	.006	.007

Note. RMSEA = root mean squared error of approximation. TYPE = type of parceling: 1 = all normal distributions; 2 = half normally and half moderately nonnormally distributed items; 3 = half normally and half severely nonnormally distributed items.

[a]Continuous refers to uncategorized data.

seen from Table 4, values of the CFI for the unparceled solutions decreased as more severely nonnormal items were introduced. However, this effect dissipated with larger sample sizes.

Parameter estimate bias. Percentage of parameter estimate bias was calculated as (estimated value–true value) / true value × 100 for all parameters, where the true values were those from which the model was generated, as shown in Figure 1. Because the covariance matrix from which the population model was generated was based on continuous and normally distributed data, parameter estimate bias reflects the effects of coarse categorization and nonnormality. In a structural

TABLE 4
Mean Values of the CFI by N and TYPE: Study 1

	Number of Items in Parcel			
Type of Data	*Unparceled*	*2*	*4*	*12*
$N = 100$				
Continuous[a]	.975	.991	.994	.994
TYPE 1	.944	.983	.993	.994
2	.826	.977	.991	.993
3	.732	.973	.991	.993
$N = 250$				
Continuous	.996	.998	.998	.998
TYPE 1	.988	.997	.998	.998
2	.909	.993	.997	.998
3	.853	.990	.997	.998
$N = 400$				
Continuous	.998	.999	.999	.999
TYPE 1	.993	.998	.999	.999
2	.925	.996	.998	.999
3	.885	.995	.998	.999
$N = 650$				
Continuous	.999	.999	.999	.999
TYPE 1	.996	.999	.997	.999
2	.932	.998	.999	.999
3	.900	.997	.999	.999
$N = 800$				
Continuous	.999	.999	.999	.999
TYPE 1	.997	.997	.999	.999
2	.936	.998	.999	.999
3	.906	.998	.999	.999

Note. CFI = comparative fit index. TYPE = type of parceling: 1 = all normal distributions; 2 = half normal and half moderately nonnormally distributed items; 3 = half normal and half severely nonnormally distributed items.

[a]Continuous refers to uncategorized data.

model such as that used in this study, bias in the structural parameters would probably be of most interest to applied researchers. In Study 1, scaling differences prevented comparisons of parameter estimates from the solutions involving 12-item parcels with those based on continuous data and on 2- and 4-item parcels. Thus, only the results for the latter three types of solutions were compared.

Repeated measures analyses of variance were conducted for values of the two structural paths, the correlation between the two exogenous factors, and the disturbance term for the endogenous factor. The results for the structural paths and the factor correlation are easily summarized: No study factor reached the criterion level of partial eta^2 for any of these parameter estimates. Mean bias percentages for these parameters across all study conditions were approximately –2, –1, and –3.4, respectively, for the $F1 \Rightarrow F3$, $F2 \Rightarrow F3$, and $F1 \Rightarrow F2$ paths. Muthén, Kaplan, and Hollis (1987) expressed the opinion that bias of less than 10% to 15% could be considered negligible. If this criterion were used, the amount of bias in these parameter estimates can be considered minor.

Bias in the structural parameters appeared to be confined to the disturbance term. A partial η^2 value of .21 for the number of items in a parcel was obtained. These results can be summarized as follows: The use of unparceled data resulted in far greater bias (average bias across all conditions = 11.1%) than did the use of two-item or four-item parcels (average bias = .11% for each). Bias levels for the unparceled data reached levels as high as 29.5% for three-category data with severely nonnormal items included, at a sample size of 100. In contrast, for the two- and four-item parcels, the highest levels of bias were .28% and .27%, respectively, at this sample size.

Study 2

Rejection and convergence rates. Only one sample solution for the Study 2 data failed to converge. This sample was in the smallest sample size condition (n = 100), with three-item parcels based on the distributed parceling strategy. The percentage of samples within each condition that would have resulted in model rejection at the .05 alpha level was calculated and these results are shown in Table 5. Because the model in Study 2 was misspecified, a rejection of the model represents a correct decision, whereas a failure to reject would be a Type II error. For the unparceled data, rejection rates of 100% were obtained at every sample size. Rejection rates were also 100% for all cells in the isolated parceling condition, whereas the distributed condition yielded lower rejection rates. Parcels based on

TABLE 5
Percentage of Model Rejections by *N*, Parcel Type, and Number of Items in a Parcel: Study 2

Type of Data	*N = 100*	*N = 250*	*N = 400*	*N = 650*	*N = 800*
Unparceled	100	100	100	100	100
2-item parcels					
Distributed parceling	74	58	55	51	50
Isolated parceling	100	100	100	100	100
3-item parcels					
Distributed parceling	80[a]	93	99	100	100
Isolated parceling	100	100	100	100	100
12-item parcels[b]	2	7	4	5	9

[a]Out of 149; all other cells are out of 150. [b]Solutions for isolated and distributed parceling for the 12-item solutions were necessarily the same as they included all items in the parcel.

all 12 items resulted in extremely low rejection rates, reflecting, in part, the low levels of power associated with these solutions.

Overall goodness of fit. Repeated measures analyses of variance were conducted for values of the RMSEA and CFI. Because the cells of the study design were not fully crossed, separate repeated measures analyses were run in which the parceled solutions were compared to the solutions based on individual items.

Values of partial η^2 for the RMSEA reached Cohen's large effect size for the following main effects: parcel type (.98), number of items in the parcel (.92), and the interaction of parcel type and number of items (.98). Comparisons of solutions based on individual items to those based on parcels yielded a partial η^2 value of .98. RMSEA values are shown in Table 6 by sample size, parcel type, and number of items. As can be seen from the table, values were substantially smaller (indicating better fit) for the distributed than for the isolated parceling strategy. Overall, RMSEA values were smaller for two-item than for three-item parcels. This result is probably due to a design artifact that resulted in some three-item parcels that contained two nonnormally and one normally distributed items. Although these parcels were balanced with an equal number of three-item parcels containing one nonnormally and two normally distributed items, this strategy appears overall to have resulted in a worse fit for the three-item as compared to the two-item parcels. Values for the 12-item parcels were similar to those for the 2-item parcels. For solutions based on unparceled data, RMSEA values fell between those obtained from the distributed and isolated parceling strategies.

CFI values are shown in Table 7. For these measures, the effects of sample size (partial $\eta^2 = .14$), parcel type (partial $\eta^2 = .97$), number of items (partial $\eta^2 = .96$), and the parcel type by number of items interaction (partial $\eta^2 = .96$) reached Co-

TABLE 6
Mean RMSEA Values by *N*, Parcel Type, and Number of Items in a Parcel: Study 2

Type of Data	*N = 100*	*N = 250*	*N = 400*	*N = 650*	*N = 800*
Unparceled	.067	.055	.055	.054	.054
2-item parcels					
Distributed parceling	.027	.012	.008	.007	.005
Isolated parceling	.094	.089	.089	.090	.089
3-item parcels					
Distributed parceling	.033	.030	.031	.032	.031
Isolated parceling	.115	.111	.112	.112	.112
12-item parcels[a]	.025	.015	.011	.008	.008

Note. RMSEA = root mean squared error of approximation.

[a]Solutions for isolated and distributed parceling for the 12-item solutions were necessarily the same as they included all items in the parcel.

TABLE 7
Mean CFI Values by *N*, Parcel Type, and Number of Items in a Parcel: Study 2

Type of Data	*N = 100*	*N = 250*	*N = 400*	*N = 650*	*N = 800*
Unparceled	.927	.949	.951	.952	.952
2-item parcels					
Distributed parceling	.992	.998	.999	.999	.999
Isolated parceling	.931	.938	.938	.938	.938
3-item parcels					
Distributed parceling	.991	.994	.994	.994	.995
Isolated parceling	.924	.930	.929	.929	.930
12-item parcels[a]	.994	.998	.999	.999	.999

Note. CFI = confirmatory factor index.
[a]Solutions for isolated and distributed parceling for the 12-item solutions were necessarily the same as they included all items in the parcel.

hen's large effect size level. Overall, solutions based on parcels of all 12 items or on the distributed parceling strategy resulted in the highest CFI values. Comparisons of solutions based on parceled and unparceled data resulted in a partial η^2 value of .95, with the unparceled solutions yielding values that were lower than those based on the distributed parceling strategy but higher than those based on the isolated strategy. It can be seen from Table 7 that the sample size effect is fairly negligible and appears to reflect the fact that CFI values increase with sample size for the unparceled data only. Parcels based on all 12 items resulted in higher values than did 2- or 3-item parcels under the isolated parceling strategy. Aside from this, the effect of the number of items in a parcel was negligible. It is interesting to note that CFI values were above .9 for all conditions, even though the model was misspecified.

Parameter estimate bias. As in Study 1, percentage of parameter estimate bias was calculated as (estimated value–true value) / true value × 100 for all parameters. Because bias in the structural parameters would probably be of most interest to applied researchers, repeated measures analyses were conducted to determine the effect of the design factors on levels of bias for the two structural paths, the disturbance term, and the correlation between the two exogenous factors.

Table 8 presents the percentage of bias for the two structural paths as well as for the correlation between the two exogenous factors across all study conditions. Bias in the estimates of both structural paths was affected by the number of items in the parcel (partial η^2 values for the $F1 \Rightarrow F3$ and $F2 \Rightarrow F3$ paths were .90 and .86, respectively). As can be seen from Table 8, this effect is due to the larger amounts of bias resulting from the 12-item parcels. The effect of the type of parceling reached the partial η^2 criterion (partial $\eta^2 = .16$) for the $F2 \Rightarrow F3$ path, with distributed par-

TABLE 8
Percentage of Parameter Estimate Bias for Structural Paths and Factor Correlation by *N*, Parcel Type, and Number of Items in a Parcel: Study 2

Type of Data	*F1* → *F3*	*F2* → *F3*	*F1* ↔ *F2*[a]
N = 100			
Unparceled	–5.9	–19.0	–8.0
2-item parcels			
Distributed parceling	–4.8	–11.7	–11.8
Isolated parceling	–5.9	–15.6	–4.8
3-item parcels			
Distributed parceling	–4.2	–15.0	–13.9
Isolated parceling	–5.9	–16.7	–3.1
12-item parcels[b]	17.4	10.9	
N = 250			
Unparceled	–6.1	–16.5	–5.8
2-item parcels			
Distributed parceling	–5.2	–13.5	–11.0
Isolated parceling	–6.1	–17.3	–4.3
3-item parcels			
Distributed parceling	–5.6	–13.4	–12.9
Isolated parceling	–6.1	–18.7	–2.3
12-item parcels	16.1	8.9	
N = 400			
Unparceled	–6.0	–16.8	–7.1
2-item parcels			
Distributed parceling	–5.0	–13.7	–12.5
Isolated parceling	–6.1	–17.7	–5.5
3-item parcels			
Distributed parceling	–5.6	–14.3	–12.3
Isolated parceling	–6.1	–17.9	–3.6
12-item parcels	16.5	8.3	
N = 650			
Unparceled	–5.7	–15.9	–5.9
2-item parcels			
Distributed parceling	–4.8	–12.8	–11.2
Isolated parceling	–5.8	–16.8	–4.3
3-item parcels			
Distributed parceling	–5.5	–13.3	–10.3
Isolated parceling	–5.8	–17.9	–2.5
12-item parcels	17.2	9.5	
N = 800			
Unparceled	–5.7	–16.5	–6.1
2-item parcels			
Distributed parceling	–4.8	–13.3	–11.6
Isolated parceling	–5.8	–17.4	–4.5
3-item parcels			
Distributed parceling	–5.0	–13.8	–11.3
Isolated parceling	–5.8	–18.7	–2.6
12-item parcels	17.3	8.8	

[a]Factor correlations in the 12-item solution were taken as fixed and thus were not estimated. [b]Solutions for distributed and isolated parceling for the 12-item solutions were necessarily the same as they included all items in the parcel.

celing resulting in less bias than the isolated parceling strategy for this path. However, the magnitude of the effects of parcel type on bias in this path was fairly minor. In fact, based on Muthén et al.'s (1987) criterion, the overall levels of bias for both the structural paths can be considered negligible under most of the study conditions. It is interesting to note, however, that the use of a parcel composed of all 12 items resulted in positive levels of bias, whereas the use of any other type of parceling or of the individual items resulted in negative bias.

Bias in values of the factor correlation was also small in magnitude, as can be seen from Table 8. Estimates of this parameter were affected at the criterion level only by the type of parceling (partial $\eta^2 = .52$). Use of the distributed parceling strategy resulted in larger amounts of bias than did use of isolated parceling.

Levels of bias for the disturbance parameter were impacted at the criterion level only by type of parcel (partial $\eta^2 = .20$). Bias percentages for these parameter estimates ranged from 19% to 28.5%, with more severe bias resulting from the isolated parceling strategy than from distributed parceling.

DISCUSSION

This article investigates some ramifications of the widespread practice of using item parcels in SEM. Applied researchers working in areas in which variables tend to be measured on coarsely categorized scales or have unacceptable levels of nonnormality have defended this practice on the grounds that it yields data with distributions that are more continuous and normally distributed, resulting in a better model–data fit. The results of this study support the argument that the use of item parcels will result in lower levels of nonnormality and in better fitting solutions for situations in which items are unidimensional. Solutions based on item parcels, rather than on individual items, were found to result in lower rejection rates overall for a known true model. With small sample sizes (100 or 250), rejection rates were greatly inflated for solutions in which parceling was not used. In contrast, rejection rates for the parceled data were much closer to the nominal .05 level, especially for parcels based on 4 or 12 items.

The practice of parceling had similar effects on values of commonly used fit indexes such as the RMSEA and CFI. Although values of these indexes worsened with the inclusion of nonnormal and coarsely categorized data for unparceled solutions, the use of item parceling appeared to ameliorate these effects. In addition, although the use of individual items resulted in some parameter estimate bias at smaller sample sizes, no parameter estimate bias was observed for the solutions based on parceled items. The results of this study suggest that the practice of parceling items does result in improved fit as well as less biased solutions in the presence of coarsely categorized items, nonnormally distributed items, or both, if the items have a unidimensional structure. Although this improvement in fit is presumably due, at least in part, to the superiority of the distributional characteristics of the parcels over those of the items, it should be pointed out that the use of parcels also results in fewer data points that must be fit. Thus, improvement in fit is due in part to the reduction in the number of variances and covariances that must be accounted for by the model. This is reflected in the results of parsimony-adjusted Normed Fit Index values (not reported in this article), which take model parsimony into account. Although other fit indexes favored the 12-item parcels, this index yielded the highest values for the 2-item parcels, which had the highest number of degrees of freedom among the parceled solutions.

Study 2 addressed the issues of parameter estimate bias and fit in the context of items that are not unidimensional. Both this study and the earlier study by Hall et al. (1999) illustrate ways in which the use of item parcels can result in the absorption of the effects of an unmodeled secondary factor into other model parameters. Although the absorption of these effects can result in improved fit for the misspecified model, both studies have shown that this process can also result in biased estimates for other model parameters, as well as an inaccurate assessment of model fit. Although in this study the amount of parameter estimate bias found for the parceled solutions was, on average, no greater than for unparceled solutions, the fit of solutions based on the distributed parceling strategy was deceptively good. This would make it difficult to detect misspecifications of the type modeled in this study if the distributed parceling strategy were used. In practice, of course, researchers will generally not be in a position to know whether their parceling strategy was an isolated or distributed one, which further complicates the interpretation of results. It therefore seems imperative that applied researchers keep in mind the possible tradeoffs between obtaining a model that fits well and a model that accurately represents the relations among the variables. A well-fitting model is of little use if it is an inaccurate representation of these relations. Examination of the RMSEA and CFI values support the conclusion that the use of item parceling can obfuscate the true factor structure of a set of items. These indexes all yielded more optimal values for the solutions based on the distributed parceling strategy, even though the model was misspecified. Based on recent guidelines offered by Hu and Bentler (1999) for the CFI and RMSEA, most, if not all, of the solutions based on this type parceling would be incorrectly judged to be well specified.

An inspection of the model modification indexes obtained for the two parceling strategies sheds some light on the reasons for the differences in model fit of the isolated and distributed strategies. Under the isolated parceling strategy many large modification indexes were obtained for measurement error covariances, indicating that the covariation among the item parcels was not well explained by the factor under this strategy. This is reasonable, because in the isolated strategy not all of the parcels contained items influenced by the secondary construct and were therefore not all measuring the same thing. Because of this, the measurement errors for parcels containing items influenced by the secondary construct were highly correlated with each other but were not correlated with those for the parcels that did not contain such items. In the distributed strategy, however, all parcels contained items influenced by the secondary construct. This did not result in correlated measurement errors, as the variance due to the secondary factor was shared among all the parcels and was thus not unique to any of them. However, it did result in an upward bias of the factor loadings for this parceling strategy. The influence of the secondary construct was thus effectively absorbed into existing model parameters and was not revealed by modification indexes under the distributed but not under the isolated parceling strategy. These results have important implications for model modifications conducted on parceled data. Although use of the distributed parceling strategy tended to result in a failure to reject a misspecified model, some model rejections will still occur under this strategy. When this happens, researchers may attempt to modify the model using modification indexes or some other post hoc strategy. However, in this study modification indexes for data based on the distributed parceling strategy failed to identify the source of the problem. Thus the use of modification indexes is unlikely to help researchers identify problems such as incorrect factor structures in the type of situation studied here. The use of this parceling strategy with incorrectly specified factor structures can therefore result in either in a failure to reject a misspecified model or in an inability to identify the sources of misfit in a poorly fitting model.

As with any investigation, this study has several limitations that bear discussion. This study was designed to be primarily exploratory in nature. Only one model was investigated with a view to discerning whether the type of item parceling that is typically seen in practice could have the effect of obscuring a true factor structure. Although this result was demonstrated in this study, it is possible that the structure investigated here is unusual in this regard. However, several additional population studies conducted by the author with models based on other multidimensional structures and other parameter values yielded results that were entirely consistent with those of this study. Even so, this study represents only a starting point; further studies utilizing a variety of models should be conducted. In addition, the effects of the number of items in a parcel as well as the total number of parcels should be investigated more thoroughly in future studies.

Should item parceling be used? At the present time the following tentative advice can be offered to applied researchers. For situations in which items have a well-known factor structure, parceling together items that are known a priori to be unidimensional appears to result in less bias in structural parameters than use of the individual items when items are coarsely categorized, nonnormally distributed, or both. Parceled solutions also appear to ameliorate the effects of coarsely categorized and nonnormally distributed item-level data on model fit. However, use of item parceling in situations in which the factor structure has not been well studied or is not unidimensional cannot be recommended as this practice can yield misleading results with regard to the actual fit of one's model, as well as biased estimates of other model parameters.

REFERENCES

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

Bagozzi, R. P., & Edwards, J. R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods, 1,* 45–87.

Bagozzi, R. P., & Heatherton, T. F. (1994). A general approach to representing multifaceted personality constructs: Application to state self-esteem. *Structural Equation Modeling, 1,* 35–67.

Bandalos, D. L. (1993). Factors affecting cross-validation of confirmatory factor analysis models. *Multivariate Behavioral Research, 2,* 351–374.

Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: New developments and techniques.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cattell, R. B. (1956). Validation and intensification of the sixteen personality factor questionnaire. *Journal of Clinical Psychology, 12,* 205–214.

Cattell, R. B., & Burdsall, C. A. (1975). The radial parcel double factor design: A solution to the item-versus-parcel controversy. *Multivariate Behavioral Research, 10*, 165–179.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cook, L. L., Dorans, N. J., & Eignor, D. R. (1988). An assessment of the dimensionality of three SAT-Verbal test editions. *Journal of Educational Statistics, 13,* 19–43.

Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling, 4,* 87–107.

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.

Gribbons, B. C., & Hocevar, D. (1998). Levels of aggregation in higher level confirmatory factor analysis: Application for self-concept. *Structural Equation Modeling, 5,* 377–390.

Hall, R. J., Snell, A. F., & Singer Foust, M. (1999). Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods, 2,* 233–256.

Harlow, L. L. (1985). *Behavior of some elliptical theory estimators with nonnormal data in a covariance structures framework: A Monte Carlo study.* Unpublished doctoral dissertation, University of California, Los Angeles.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

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

Kishton, J. M., & Widaman, K. F. (1994). Unidimensional versus domain representative parceling of questionnaire items: An empirical example. *Educational and Psychological Measurement, 54,* 757–765.

MacCallum, R. C., & Tucker, L. R. (1991). Representing sources of error in the common factor model: Implications for theory and practice. *Psychological Bulletin, 109,* 502–511.

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

Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33,* 181–220.

Michael, W. B., & Bachelor, P. (1988). A comparison of the orthogonal and the oblique factor structures of correlation matrixes of individual items and of composites of items (subtests) derived from a standardized affective measure. *Educational and Psychological Measurement, 48,* 93–103.

Muthén, B. O., & 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,* 171–189.

Muthén, B. O., & 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., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 52,* 431–462.

Russell, D. W., Kahn, J. H., Spoth, R., & Altmaier, E. M. (1998). Analyzing data from experimental studies: A latent variable structural equation modeling approach. *Journal of Counseling Psychology, 45,* 18–29.

Schau, C., Stevens, J., Dauphinee, T. L., & Del Vecchio, A. (1995). The development and validation of the Survey of Attitudes toward Statistics. *Educational and Psychological Measurement, 55,* 868–875.

Singh, J., & Rhoads, G. K. (1991). Boundary role ambiguity in marketing-oriented positions: A multidimensional, multi-faceted operationalization. *Journal of Marketing Research, 28,* 328–338.

Takahashi, T., & Nasser, F. (1996, April). *The impact of using item parcels on ad hoc goodness of fit indices in confirmatory factor analysis: An empirical example.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Thompson, B., & Melancon, J. G. (1996). *Using item "Testlest"/"Parcels" in confirmatory factor analysis: An example using the PPSDQ-78.* (ERIC Document No. ED 404 349)

Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods, 3*, 231–251.

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

Yuan, K.-H., Bentler, P. M., & Kano, Y. (1997). On averaging variables in a CFA model. *Behaviormetrika, 24,* 71–83.