MULTIVARIATE BEHAVIORAL RESEARCH, *40*(2), 235–259

Using Parcels to Convert Path Analysis Models Into Latent Variable Models

Donna L. Coffman and Robert C. MacCallum
Department of Psychology and L.L. Thurstone Psychometric Laboratory
The University of North Carolina Chapel Hill

The biasing effects of measurement error in path analysis models can be overcome by the use of latent variable models. In cases where path analysis is used in practice, it is often possible to use parcels as indicators of a latent variable. The purpose of the current study was to compare latent variable models in which parcels were used as indicators of the latent variables, path analysis models of the aggregated variables, and models in which reliability estimates were used to correct for measurement error in path analysis models. Results showed that point estimates of path coefficients were smallest for the path analysis models and largest for the latent variable models. It is concluded that, whenever possible, it is better to use a latent variable model in which parcels are used as indicators than a path analysis model using total scale scores.

Structural equation modeling is a data analytic technique commonly used to examine patterns of relationships among constructs. Often, some or all of these constructs are measured by multi-item scales. The researcher then has several options for specifying the constructs in the structural equation model. These options have different implications for the parameter estimates and fit of the model and each of these options has advantages and disadvantages. In this article we suggest using parcels as indicators of latent variables, where parcels are aggregations (sums or averages) of several individual items. Advantages of using parcels as indicators of latent variables in structural equation models will be discussed.

The authors wish to thank Kristopher J. Preacher and Cheongtag Kim for valuable comments on an earlier draft of this manuscript.

Correspondence concerning this article should be addressed to Donna L. Coffman, Department of Psychology, UNC-CH, CB# 3270, Chapel Hill, NC 27599–3270. E-mail: dcoffman@email.unc.edu

LEVELS OF AGGREGATION

In a situation where a construct is measured by a scale with multiple items, several different aggregation levels of latent variable indicators can be considered: total disaggregation, partial disaggregation, and total aggregation. In a totally disaggregated model, each item serves as an indicator for a construct. In a partially disaggregated model, several items are summed or averaged resulting in parcels. These parcels are then used as indicators for constructs. Parcels are usually not defined in terms of content, and therefore they are usually not interpretable. Rather they represent the latent variable construct. In a totally aggregated model, all of the items for a scale are summed or averaged (Bagozzi & Edwards, 1998; Bagozzi & Heatherton, 1994; Gribbons & Hocevar, 1998). The result is that if only one scale is used to measure each construct, then there is only one indicator per construct and the model is a path analysis rather than a latent variable model. If more than one scale was used to measure a construct, then it is still possible to specify a latent variable, and each indicator is a total scale score. The present article will focus on comparing totally aggregated models to partially disaggregated models. Previous research has focused on totally disaggregated models and partially disaggregated models, particularly in the context of how parcels are constructed from items. These methods of constructing parcels and relevant research comparing items versus parcels as indicators of latent variables will be briefly reviewed.

Partially Disaggregated Models

Parcel construction. Parcels may be constructed in various ways depending on the nature of the model. When there is a broad construct that encompasses several first-order factors, all the items measure the broad construct (second-order factor) as well as a particular dimension (first-order factor). This type of model is usually referred to as a hierarchical model because the items load on various first-order factors that in turn load on a single second-order factor. Thus, the second-order factor explains the correlations among the first-order factors. Typically, items load on only one of the first-order factors. Given such a factor structure, there are two ways to construct parcels. First, homogenous parcels could be constructed such that each parcel is made up of items that load on the same first-order factor. Each parcel then represents a particular first-order factor, although different parcels will represent different first-order factors. Second, domain representative parcels (Kishton & Widaman, 1994) could be constructed such that each parcel is made up of items that load on different first-order factors. Although items within a given domain representative parcel represent different first-order factors, the parcels all represent the same second-order factor. Using either construction method, the parcels serve as indicators of the second-order factor.

For models in which all of the items measure a single construct, that is, a unidimensional scale, parcels may be constructed by randomly assigning items to parcels. Undimensionality may be determined by factor analysis. All items should load on the same factor. In this situation, there is not a hierarchical structure to the model. Parcels constructed in this manner are called unidimensional parcels (Kishton & Widaman, 1994).

In order to construct domain representative parcels, one needs to know, on the basis of theory or previous research, the broad construct assessed by the first-order factors and which items load on these first-order factors (Kishton & Widaman, 1994). For example, Widaman, Gibbs, and Geary (1987) examined the structure of adaptive behavior, in which six factors represented different domains of adaptive behavior. In this example, an item may measure a construct such as *independent living skills* as well as the broader construct of *adaptive behavior.*

Since there is some debate in the research literature about different methods of constructing parcels, we chose to construct parcels using different methods even though the primary goal of the current study is not to compare the different parcel construction methods but rather to compare each parcel model with a totally aggregated scale score model. Once the parcels have been constructed using either method, they may be used as indicators of a latent variable rather than using the individual items as indicators. Several advantages of using parcels (a partially disaggregated model) rather than items (a totally disaggregated model) as indicators of latent variables may be identified.

Advantages of parcels. One advantage of using parcels as indicators of constructs is that parcels generally have higher reliability than single items (Kishton & Widaman, 1994). This result is similar to that associated with test length, in which the reliability of a test increases as the number of items included in the test increases. Assuming parallel items, the reliability of a test of a particular length can be predicted using the Spearman-Brown prophecy formula when the reliability of the current test is known (Crocker & Algina, 1986; Nunnally, 1978). This same formula could be used to predict the reliability of a parcel constructed from a particular number of items assuming that the items are parallel. The formula in terms of items is

$$\rho_{xx} = \frac{k\rho_{ii}}{1+(k-1)\rho_{ii}}, \tag{1}$$

where k is the number of items in a composite, ρ_{xx} is the composite reliability, and ρ_{ii} is the reliability of a single item. As the number of items in a parcel increases, the reliability of the parcel should increase. At a certain point the reliability will level off and the addition of more items to a parcel will not result in a large increase in reliability. For example, if the reliability of a single item is .50, then as more items are added to the composite the reliability will increase until the composite consists of about six items, at which point the reliability begins to level off.

A second advantage of using parcels rather than items as indicators of latent variables involves the reduction in the number of measured variables in a model. Theoretically, the use of more indicators will lead to a better representation of a construct. Practically, the researcher must find an optimal number of indicators that sufficiently represents the construct but that is fairly small because as the number of indicators increases so does the order of the correlation or covariance matrix and the number of parameters to be estimated. The larger the order of the correlation matrix, the less likely the model is to fit well even if the model closely approximates the phenomenon under study. From this perspective, models with parcels as indicators are likely to fit better than models with items as indicators because the order of the parcel correlation matrix is much smaller than the order of the item correlation matrix. In a simulation study, Marsh, Hau, Balla, and Grayson (1998) found that as the number of indicators per factor increased from 2 through 12, the accuracy of the parameter estimates and the percentage of convergence to proper solutions improved although the goodness of fit declined even though the model was correctly specified. They also used the 12 indicators per factor to create parcels but the 12 indicators per factor model solutions were better behaved than the parceled solutions in terms of convergence to proper solutions. The estimated parameters for the unique variances decreased as the number of items included in each parcel increased. They noted that the variability of the factor loadings and unique variance estimates decrease as the number of items included in the parcel increases. Thus, the parameter estimates for the parceled solutions were more precise. However, there are fewer independent estimates of the factor loadings when using parcels. As noted above, there is essentially a trade-off between finding a sufficient number of indicators to represent a construct and indicators that are more reliable. Little, Lindenberger, and Nesselroade (1999) provide an excellent discussion of this issue and note that using parcels can achieve this balance by choosing an optimal number of indicators to represent a construct and reduce the diversity of the indicators thereby increasing their reliability.

Another advantage of parcels is that they can be used as an alternative to data transformations or alternative estimation techniques when working with nonnormally distributed variables. The most often used estimation method in structural equation modeling, maximum likelihood, assumes multivariate normality of the measured variables in the population. If the measured variables are not multivariate normal, then estimates of fit measures and estimates of standard errors of parameters may not be accurate (Hu & Bentler, 1998). Although there are alternative estimation methods that can be used with nonnormal variables, these methods have some disadvantages (Curran, West, & Finch, 1996) in that they require very large sample sizes or do not allow for assessment of model fit. Therefore, researchers have examined other ways in which one can correct for nonnormal data including parceling. If items are not normally distributed and they are combined to form parcels, the parcels may be more normally distributed than the original items (West, Finch, & Curran, 1995). Parceling could also be used with dichotomous items, which obviously cannot be normally distributed. Parcels created from such items may approximate a normal distribution, especially as the number of items comprising each parcel increases. If it is assumed that the dichotomous items measure a continuous underlying construct, then combining the items into parcels creates parcels with a more differentiated scale and the parcels would tend to have higher communalities than the items. One could also estimate a model using the dichotomous items as indicators of latent variables in Mx, Mplus, or LISREL; however, using parcels as indicators of latent variables still has all of the advantages mentioned above.

Previous research examining items versus parcels. Many studies have compared the effects on parameter estimates and overall model fit of fitting latent variable models with different types of indicators, such as items and parcels constructed in various ways. The results of these studies generally indicate that parcels have advantages over items as indicators of latent variables. In general, the use of parcels constructed using the unidimensional method results in less biased parameter estimates and better overall model fit than item level models. Overall goodness of fit has been found to be better for parceled models than for item level models using both empirical and simulated data (Bandalos, 2002; Bandalos & Finney, 2001; Gribbons & Hocevar, 1998). Parceling of nonnormal variables results in improved overall fit of the model compared to item level analysis of nonnormal variables (Plummer, 2000). In addition, the degree of nonnormality is reduced as a result of parceling (Bandalos, 2002). However, parceling may have some disadvantages when compared with analysis using individual items. Research has supported Mulaik's (2000) suggestion that parceling may result in misspecified models that fit better than or as well as correctly specified item level models (Plummer, 2000), and that parceling can obscure additional unmodeled factors (Hall, Snell, & Foust, 1999). In addition, as noted above, Marsh et al. (1998) found that the use of parcels may result in a decrease in convergence to proper solutions compared to individual indicators.

Totally Aggregated Models

While there has been considerable research on the use of items and parcels as indicators of latent variables, there has been little or no research that has examined the use of parcels rather than total scale scores. Bandolos (2002) did examine the use of parcels versus total scale scores, however, half of the simulated items were nonnormal. The use of parcels has several advantages over the use of totally aggregated models. As mentioned previously, if only one multi-item scale was used to assess a construct and all of the items for that scale are summed or averaged, the result is a measured variable. If possible, it is advantageous to specify latent variables rather than measured variables because measured variables are assumed to be measured without error (Kline, 1998). This assumption is not made when using latent variables. In latent variable models, estimates of the unique variance, of which error variance is a component, are obtained for each latent variable indicator, and thus, the error variance is estimated as part of the model. Using a model that assumes that the measured variables are error free when in fact they are not can bias the parameter estimates. Depending on the model and the correlations among the measured variables, it is difficult to predict the direction of the bias in the parameter estimates (Bollen, 1989). In most situations, however, the presence of measurement error will result in underestimation of effects among variables. The use of latent variables wherein such error is explicitly represented can correct for bias in estimates of such effects.

While it is advantageous to specify latent variables, it is not always possible to do so. For example, if there is only one indicator for a construct, as might occur when using a one-item scale, it is not possible to specify a latent variable for that construct because the model would not be identified and the unique variance for a single indicator cannot be estimated. When this is the case, Bollen (1989) suggested obtaining a reliability estimate $(\hat{\rho}_{xx})$ and using $(1-\hat{\rho}_{xx})(\hat{\sigma}_x^2)$ as an estimate of the unique variance of the indicator where $\hat{\sigma}_x^2$ is the variance of the measured variable indicator. Since the unique variance then becomes a fixed parameter, a latent variable can be specified with only one indicator. When analyzing a correlation matrix, the estimate of the unique variance simplifies to $(1-\hat{\rho}_{xx})$. We will be analyzing correlation matrices, so we will refer to the estimate of the unique variance as simply $(1-\hat{\rho}_{xx})$ but it should be kept in mind that when analyzing covariance matrices, the estimate of the unique variance is $(1-\hat{\rho}_{xx})(\hat{\sigma}_x^2)$. As with any structural equation model, either the variance of the latent variable or one of the paths emitted by the latent variable must be fixed to some chosen value so that the model will be identified. This technique does not estimate unique variance as part of the model, however it does not assume that the unique variance is zero as the path analysis model does. Therefore, the resulting parameter estimates should be less biased than they would be in a path analysis model. This effect has been supported by previous research (Stephenson & Holbert, 2003).

Hence, there are two alternatives to a totally aggregated model that will be considered in the present study: A partially disaggregated model using parcels as indictors of latent variables and a model using a reliability correction. It is expected that it is more advantageous to use partially disaggregated models than to use the reliability correction technique because the unique variance estimated for latent variable indicators in the partially disaggregated model is not equal to error variance. According to the common factor model (Thurstone, 1947), unique variance includes both specific variance and error variance. Specific variance represents systematic variability due to factors that affect only a given measured variable. Error variance represents unsystematic variance due to random measurement error or unreliability. The reliability estimate includes all systematic variability and therefore, includes a combination of both common and specific variance, whereas $(1-\hat{\rho}_{xx})$ estimates unsystematic or error variance only. Thus, using $(1-\hat{\rho}_{xx})$ as an estimate of the unique variance will include only the error variance component of unique variance, and therefore, will more than likely underestimate the unique variance. A recent study found less biased parameter estimates for a latent variable model (with items as indicators) compared to a reliability corrected model (Stephenson & Holbert, 2003), and this underestimation of the unique variance in the reliability corrected model is a likely reason for the less biased parameter estimates in the latent variable model.

The Use of Parcels in Applied Research

We were interested in the different levels of aggregation utilized by applied researchers when they wish to examine relationships among constructs for which they have multi-item scales. A previous review of the applied structural equation modeling literature suggests that it is fairly common for researchers to use totally aggregated path analysis models when they could have constructed parcels and used the parcels as indicators of latent variables in partially disaggregated models (MacCallum & Austin, 2000).

We were interested in how researchers use parcels, the reasons they give for using them, and how they are constructing them. We were also interested in whether researchers use total scale scores (and thus path analysis) when they could have constructed parcels and used a partially disaggregated model, and if so, the reasons provided for using a totally aggregated model rather than a partially disaggregated model. To examine these issues, applied research articles from the 2001 issues of *Journal of Abnormal Psychology*, *Journal of Counseling Psychology*, *Journal of Educational Psychology*, and *Journal of Applied Psychology* were reviewed in addition to those reviewed by MacCallum and Austin (2000). Those articles (29 total) that used path analysis, confirmatory factor analysis, structural equation modeling, or regression analysis were examined more closely to determine whether or not the authors had used parcels. If parcels were not used and the authors used a measured variable model, then an attempt was made to determine whether the authors could have constructed parcels and used a latent variable model.

Many of the applications that used parcels as indicators of latent variables in confirmatory factor analysis studies (e.g., Lehrman-Waterman & Ladany, 2001; Paullay, Alliger, & Stone-Romero, 1994; Shore, Tetrick, Sinclair, & Newton, 1994) and in structural models (e.g., Fuller & Hester, 2001; Lopez & Little, 1996; Marsh, Kong, & Hau, 2001; Smith, Goldman, Greenbaum, & Christiansen, 1995), did not cite a reason for using parcels, but those that did most often cited enhanced reliability of indicators and reduction in the number of parameters to be estimated. Similarly, many did not state how the parcels were constructed or if they did, then the reason for constructing the parcels in a particular manner was not given or was unclear.

In many of the studies reviewed, the researchers used total scale scores in a path analysis model when they could have used parcels as indicators in a latent variable model (e.g., Cappella & Weinstein, 2001; Chan, Schmitt, DeShon, Clause, & Delbridge, 1997; Gong, Shenkar, Luo, & Nyaw, 2001; Hill & Fischer, 2001; Kahn, 2001; Lee & Liu, 2001; Lopez, 2001; Macan, 1994; Miller & Byrnes, 2001; Pajares & Miller, 1995; Perrone & Worthington, 2001; Portello & Long, 2001; Rioux & Penner, 2001; Rose & Feldman, 1997; Stice & Barrera, 1995; Weisz, Southam-Gerow, & McCarty, 2001; Whittaker & Robitschek, 2001). In all of these studies, no reason was provided for using a measured variable model rather than a latent variable model. In some cases, there were constructs for which the researchers had multiple indicators and could have specified a latent variable, as well as constructs for which the researcher did not have multiple indicators and thus could not have specified a latent variable. However, in these cases, a partially latent variable model could have been specified rather than a path analysis model.

In our review of the applied literature, we found that several researchers used the reliability correction technique described above rather than using a latent variable model with either items or parcels as indicators (e.g., Abbey, Andrews, & Halman, 1995; Allen & Griffeth, 2001; Hofmann & Morgeson, 1999; Masterson, 2001; VandeWalle, Cron, & Slocum, 2001). In several of the studies (Allen & Griffeth; Hofmann & Morgeson; Masterson), the authors stated that the reason they used the reliability correction technique rather than a latent variable model was due to their small sample size. In the remaining studies (Abbey et al.; VandeWalle et al.), the authors did not provide a reason for using the reliability correction technique rather than a latent variable model.

Rationale for Current Study

While there has been a considerable amount of research examining differences between parcel and item level analyses with respect to measures of overall fit and parameter estimates, there has been almost no research that has examined differences between parcel and total scale analyses. In addition, the previous research has mostly examined the factor loading estimates in confirmatory factor analysis models. To our knowledge, only two studies have examined structural coefficients. Bandolos (2002) examined structural coefficients, however, half of the simulated items were nonnormal. Hall et al. (1999) also examined the structural coefficient between two latent variables, however, they did not compare a parcel model with a total scale score model. Since they only examined two latent variables, the structural parameter in their models is equivalent to a correlation coefficient. The goal of the current study is to examine the path coefficients in structural models. For reasons discussed above it may be better to use parcels (a partially disaggregated model) rather than total scale scores (a totally aggregated model). As indicated in the review of published applied studies, it is fairly common for researchers to use total scale scores and thus a path analysis model when latent variables could have been specified if parcels had been constructed. The results of these path analyses of total scale scores are susceptible to the biasing effects of error, which could be avoided with the use of latent variables using parcels as indicators. The purpose of the present study is to compare partially disaggregated models in which parcels are used as indicators of the latent variables, path analysis models of the aggregated variables, and reliability corrected models in which the unique variance was fixed to $(1-\hat{\rho}_{xx})$. Based on the theoretical rationale discussed above, it is expected that the use of partially disaggregated models will result in less biased parameter estimates than the use of totally aggregated or reliability corrected models. Two approaches are taken in this study to examine this issue: (a) a demonstration using a sample drawn from an artificially simulated population, and (b) a demonstration using data from a large empirical study.

DEMONSTRATION WITH ARTIFICIAL DATA

Since the use of empirical data precludes knowing the model that generated the data, it is useful to consider a demonstration using artificially simulated data. The procedure for producing simulated data began with construction of a model, which is presented in Figure 1. There are four constructs labeled C1, C2, C3, C4, each of which takes the form of a second-order factor. C1 and C2 are correlated and both influence C3, which then influences C4. For each construct there are 3 first-order factors, labeled C1a, C1b, C1c, and so forth. For each of the first-order factors there are three measured variable indicators, which can be viewed as items, resulting in a total of 36 measured variable indicators. The use of this measurement structure provides a basis for constructing several types of parcels as indicators of the four constructs.

Numerical values were assigned to all parameters as shown in Figure 1. The values are fairly small for the factor loadings of the items onto the first-order factors, as these values would be expected to be fairly small in practical applications. The values for the factor loadings of the first-order factors onto the second-order factors are somewhat larger than those for the items since these values would be expected to be somewhat larger in practical applications. The structural parameters are specified as having still larger values. Although the relative magnitudes of these various classes of coefficients as specified in our demonstration will not hold in all empirical examples, such variability among studies should not alter our find-

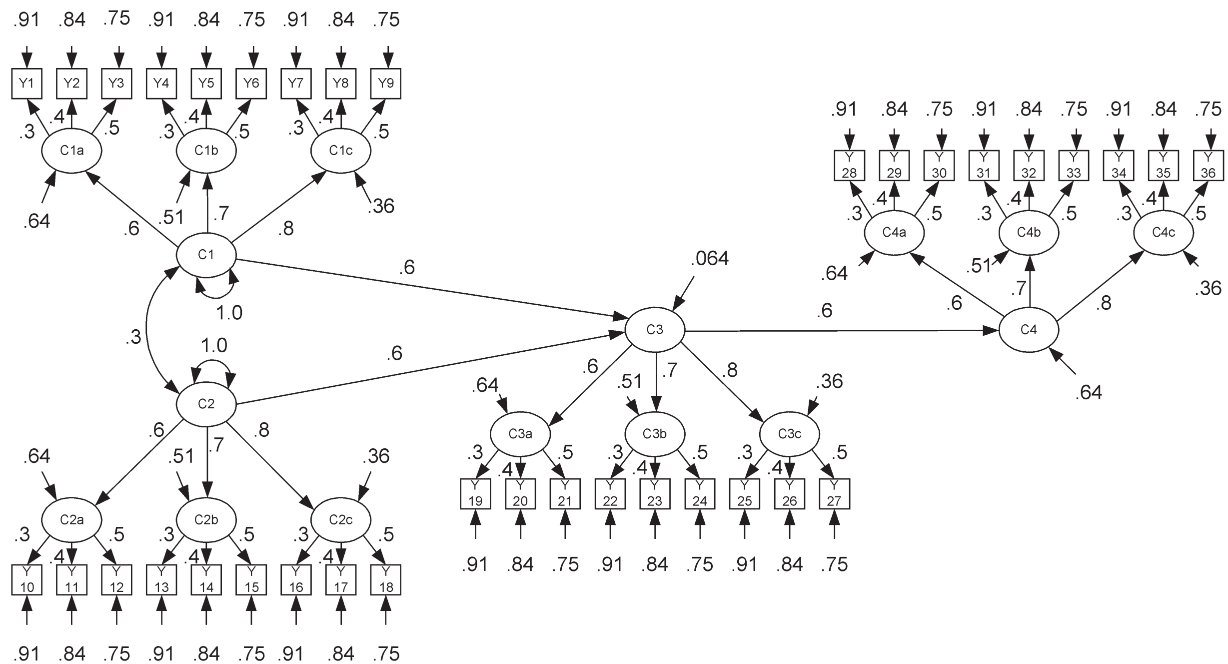

FIGURE 1 Population model for the artificial data.

ings and conclusions. The unique variance estimates were specified such that the model implied dispersion matrix would be a correlation matrix. The variances of the exogenous second-order factors were fixed to unity and the variances of the endogenous variables were constrained to unity by fixing the residual variances to the appropriate values. From the parameter values in Figure 1, the model implied correlation matrix for the 36 items was computed. This matrix was then treated as the population correlation matrix, **P**.[1]

Next, a sample of $N = 10{,}000$ was randomly drawn from a multivariate normal population with correlation matrix **P** and the items were summed in various ways to create indicators for the constructs C1-C4: total scale scores, homogeneous parcels, and domain representative parcels. Items Y1-Y9 were summed to create a total scale score for C1. Items Y10-Y18 were summed to create a total scale score for C2 and similarly items Y19-Y27 and items Y28-Y36 were summed to create total scale scores for C3 and C4, respectively. Domain representative parcels were created by summing one indicator from each of the first-order factors. For example, items Y1, Y5, and Y9 were summed to create a domain representative parcel to serve as an indicator of C1 and items Y2, Y6, and Y7 were summed to create another parcel for C1. Homogeneous parcels were constructed by summing the three indicators for each of the first-order factors. For example, items Y1, Y2, and Y3 were summed to create a homogeneous parcel to serve as an indicator of C1. This resulted in three parcels per construct. A 4 × 4 correlation matrix, $\mathbf{P}_S$, was computed for total scale model, a 12 × 12 correlation matrix, $\mathbf{P}_D$, was computed for the domain representative parcel model, and a 12 × 12 correlation matrix, $\mathbf{P}_H$, was computed for the homogenous parcel model. Finally, Cronbach's alpha was computed for the items for each of the second-order factors to use as a reliability estimate for the reliability corrected model.

A path analysis model of total scale scores was fit to $\mathbf{P}_S$, a reliability corrected model in which the unique variance was fixed to $(1-\hat{\rho}_{xx})$ was fit to $\mathbf{P}_S$, a latent variable model in which domain representative parcels were used as indicators was fit to $\mathbf{P}_D$, and a latent variable model in which homogeneous parcels were used as indicators was fit to $\mathbf{P}_H$. Given the very large sample size, sampling error should be minimized. Of critical interest is the comparison between the structural path coefficient estimates for the path analysis model of total scale scores and those from the models in which parcels were used as indicators. These estimates are presented in Table 1.

The results of the demonstration with artificial data show that the path coefficient estimates for the total scale model are severely attenuated when compared to the other models or to the population value. The estimates of the residual variances for the total scale model were much higher than those of the other models.

[1]The correlation matrix for the simulated data is provided on-line at http://www.unc.edu/~dcoffman/manuscripts or can be reproduced using the parameter values provided in Figure 1.

TABLE 1
Structural Parameter Estimates for Models Fit to Artificial Data Example

Parameter	*Population Value*	*Total Scale*	*Reliability Corrected*	*Domain Rep Parcels*	*Homogeneous Parcels*
ϕ_{21}	0.30	0.12	0.12	0.42	0.30
γ_1	0.60	0.26	0.46	0.34	0.60
γ_2	0.60	0.25	0.43	0.45	0.57
β_1	0.60	0.21	0.45	0.20	0.58
ζ_1	0.064	0.86	0.26	0.56	0.10
ζ_2	0.64	0.96	0.44	0.96	0.66
Goodness of fit					
χ^2		260.03	57.46	55.14	31.14
df		2	2	50	50
p		<.001	<.001	.29	.98
RMSEA		.11	.053	.0032	.00
(90% CI)		(.10, .12)	(.042, .065)	(.0, .0074)	(.0, .0)

The path coefficient estimates for the reliability corrected and domain representative parcel models were larger than those for the total scale model, but the path coefficient estimates for the homogeneous parcel model were much higher than those for the total scale model. The estimate of the correlation between the exogenous variables in the homogenous parcel model was also much larger than the corresponding estimate in the total scale model. The path coefficient estimates for the domain representative parcel model were slightly higher than those for the reliability corrected model and much higher than those for the total scale model. However, they were smaller than those for the homogeneous parcel model and the population values. The path coefficient estimates for the homogeneous parcel model were higher than those for any of the other models and the estimates were also the closest to the population values than any of the estimates for the other models. The estimates of the residual variances for the homogeneous parcel model were much smaller than those of the total scale model and were also closer to the population values. Clearly, the parameter estimates for the homogeneous parcel model were closer to the population values than the parameter estimates for any of the other models.

In terms of overall model fit, which is presented in Table 1, the homogeneous parcel model fit very well. The overall fit for the domain representative model was not as good as that of the homogeneous parcel model but was definitely better than that of the reliability corrected model or the total scale path analysis model. In fact, the null hypothesis of exact fit and the null hypothesis of close fit can both be rejected for the total scale path analysis model. The null hypothesis of exact fit can be rejected for the reliability corrected model. Thus, in terms of overall model fit, the homogeneous parcel model is the best fitting model.

DEMONSTRATION WITH EMPIRICAL DATA

The demonstration with empirical data involved applying a series of models with different types of indicators to empirical data and then evaluating the effects on parameter estimates. Empirical data were obtained from a large national survey that included behavioral, psychological, and social measures on a sample of children.

Variables

The variables used in the study included depression, mastery, and self-esteem. The correlations and standard deviations are provided in Tables 2, 3, and 4, respectively. Depression was measured by the 7-item Center for Epidemiological Studies–Depression (CES–D) scale. Mastery was measured using the Pearlin Mastery scale, a seven-item measure of the extent to which one perceives himself or herself as in control of forces that impact his or her life. Self-esteem was

TABLE 2
Correlations and Standard Deviations for the CES–Depression Items

CES–D	*VAR 1*	*VAR 2*	*VAR 3*	*VAR 4*	*VAR 5*	*VAR 6*	*VAR 7*
VAR 1	1						
VAR 2	0.251	1					
VAR 3	0.302	0.344	1				
VAR 4	0.073	0.104	0.042	1			
VAR 5	0.251	0.284	0.311	0.117	1		
VAR 6	0.271	0.313	0.556	0.04	0.328	1	
VAR 7	0.208	0.291	0.279	0.099	0.295	0.309	1
SD	.791	.858	.784	1.143	.964	.769	.801
M	.47	.67	.45	1.40	.75	.48	.57

Note. CES = Center for Epidemiological Studies.

TABLE 3
Correlations and Standard Deviations for the Mastery Items

Mastery	*VAR 1*	*VAR 2*	*VAR 3*	*VAR 4*	*VAR 5*	*VAR 6*	*VAR 7*
VAR 1	1						
VAR 2	0.363	1					
VAR 3	0.376	0.379	1				
VAR 4	0.170	0.149	0.274	1			
VAR 5	0.430	0.441	0.421	0.224	1		
VAR 6	0.218	0.156	0.257	0.374	0.226	1	
VAR 7	0.327	0.192	0.349	0.17	0.336	0.214	1
SD	.765	.772	.694	.632	.705	.623	.743
M	3.08	2.91	3.18	3.42	3.02	3.38	3.0

TABLE 4
Correlations and Standard Deviations for the Rosenberg Self-Esteem (RSE) Items

RSE	*VAR 1*	*VAR 2*	*VAR 3*	*VAR 4*	*VAR 5*	*VAR 6*	*VAR 7*	*VAR 8*	*VAR 9*	*VAR 10*
VAR 1	1									
VAR 2	0.600	1								
VAR 3	0.410	0.463	1							
VAR 4	0.440	0.523	0.408	1						
VAR 5	0.385	0.422	0.486	0.389	1					
VAR 6	0.404	0.485	0.453	0.453	0.405	1				
VAR 7	0.363	0.444	0.376	0.434	0.389	0.517	1			
VAR 8	0.280	0.285	0.335	0.232	0.282	0.341	0.298	1		
VAR 9	0.280	0.322	0.390	0.319	0.388	0.423	0.404	0.390	1	
VAR 10	0.348	0.393	0.467	0.359	0.431	0.468	0.426	0.397	0.628	1
SD	.538	.543	.571	.557	.645	.596	.571	.767	.704	.671
M	3.28	3.35	3.44	3.32	3.35	3.26	3.21	2.96	3.02	3.23

measured by the 10-item Rosenberg Self-Esteem (RSE) scale (Center for Human Resource Research, 2002).

Sample

Item level data collected in 1998 were obtained from the National Longitudinal Survey of Youth Child and Young Adult cohort (Center for Human Resource Research, 2002). There were 2,120 participants with complete self-esteem, mastery, and depression data.

Parcel Construction

Exploratory factor analyses were conducted on each of the scales to determine the dimensionality of the scales. All factor analyses were conducted using Comprehensive Exploratory Factor Analysis (CEFA) software (Tateneni, Mels, Cudeck, & Browne, 1998), maximum likelihood estimation, and oblique rotation (Crawford-Ferguson direct Quartimin). For each scale, models with different numbers of factors were examined and the decision as to the number of factors to retain was based on the χ^2 difference test, the RMSEA, and interpretability.

CES–D. For the CES–D scale a two-factor model was retained. Two items loaded on one factor and five items loaded on another factor. One factor could be considered *emotion or feeling* as the two items which loaded on this factor asked whether the respondent felt sad or depressed. The second factor could be considered *physical symptoms of depression*. This factor included items such as whether the respondent felt like eating. The correlation between the two factors was .684.

To construct homogeneous parcels, the two items loading on the first factor were assigned to one parcel. The remaining five items loading on the second factor were assigned to one of two parcels. Thus, there were a total of three parcels, two of which contained two items each and one that contained three items. To construct the domain representative parcels, the two items that loaded on the first factor were assigned to different parcels. The remaining five items representing the second factor were assigned to parcels such that there were again three parcels, two of which contained two items each and one that contained three items. One of the parcels was not domain representative since there were three parcels but only two items loading on the *emotion/feeling* factor.

Mastery. For the Pearlin mastery scale a two-factor model was retained. Two items loaded on one factor and five items loaded on a second factor. The two items loading on the first factor were worded positively whereas the five items loading on the second factor were worded negatively. For example, one of the items on the first factor asked whether the respondent felt they could do just about anything and one of the items on the second factor asked whether the respondent felt pushed around in life. The correlation between the two factors was –.507. Both homogeneous and domain representative parcels were constructed exactly as described above for the CES–D. Again, for the domain representative parcel model, one of the parcels was not domain representative. Since there were three parcels but only two items loading on one of the factors, there were not enough items from the first factor to distribute to each of the parcels.

Self-esteem. For the RSE scale a two-factor model was retained. Three items loaded on one factor and seven items loaded on another factor. These two factors were similar to the two mastery factors in that one consisted of positively worded items and the other consisted of negatively worded items. The correlation between the two factors was –.604. Homogeneous and domain representative parcels were constructed as described above except that the result was three parcels, two containing three items each, and one containing four items. All parcels were domain representative for the domain representative model.

In summary, using results of EFA for each of the three scales of interest, both homogeneous and domain representative parcels were constructed for each of the constructs. Self-esteem, mastery, and depression were measured by three parcels each. Following the exploratory factor analyses, some items were reverse scored so that all correlations among items and factors were positive.

Models

Two substantive models were created to represent hypothesized patterns of relationships among these three constructs. In the first substantive model, self-esteem and mastery predicted depression. The second substantive model was a mediation model in which mastery predicted self-esteem, and self-esteem predicted depression.

Design

For each of the two substantive models four different structural equation models were specified and fit to appropriate data: a total scale model, a reliability corrected model, a partially disaggregated domain representative parcel model, and a partially disaggregated homogeneous parcel model. The total scale model was a path analysis model in which the total score for each scale represented a measured variable. For the reliability corrected model the unique variance for each variable was fixed to $(1-\hat{\rho}_{xx})$ and the path from the latent variable to the measured variable indicator was fixed to $\sqrt{\hat{\rho}_{xx}}$.[2] Cronbach's alpha was obtained for each of the scales and used as an estimate of reliability. For the latent variable models, parcels constructed in one of the two ways described above were used as indicators of the latent variables. RAMONA (Browne, Mels, & Cowan, 1994) was used to fit all models to the sample correlation matrices using maximum likelihood estimation. Since the models are scale invariant, the analysis of correlation matrices is justified (Cudeck, 1989). The parameter estimates from the four versions of each model will be compared.

Results

Model 1. For the model in which mastery and self-esteem predicted depression, the point estimates and confidence intervals of the parameter estimates for the total scale model, the reliability corrected model, the domain representative parcel model, and the homogeneous parcel model are presented in Table 5. The point estimates of the path coefficients and the correlation between mastery and self-esteem were generally smallest (in absolute value) for the total scale model and were largest for the homogeneous parcel model. The estimates were similar for the reliability corrected model and the domain representative parcel model. The point estimate of the residual variance was highest for the total scale model and lowest for the homogeneous parcel model. The overall goodness of fit for the domain representative model was good whereas the goodness of fit for the homogeneous parcel model was unacceptable. The overall goodness of fit for the total scale model and the reliability corrected model could not be evaluated because the models are saturated.

[2]We could have fixed the variance of the latent variable to one or any other chosen value. Likewise, we could have also fixed the factor loading to another value but we chose $\sqrt{\hat{\rho}_{xx}}$ as suggested by Bollen (1989).

TABLE 5
Structural Parameter Estimates for the Regression of Self-Esteem and Mastery on Depression

Parameter	*Total Scale Model*	*Reliability Corrected Model*	*Domain Representative Parcel Model*	*Homogeneous Parcel Model*
γ_1	–.166	–.309	–.224	–.500
	(–.212, –.120)	(–.431, –.186)	(–.338, –.110)	(–.630, –.369)
γ_2	–.145	–.080	–.174	.041
	(–.191, –.099)	(–.198, .038)	(–.283, –.065)	(–.087, .169)
ϕ	.669	.835	.834	.840
	(.649, .688)	(.811, .860)	(.811, .854)	(.814, .863)
ζ	.919	.857	.854	.783
	(.901, .938)	(.824, .892)	(.821, .883)	(.738, .823)
Goodness of fit				
χ^2			140.106	797.413
df			24	24
p			.668	<.001
RMSEA			.048	.123
(90% CI)			(.04, .056)	(.116, .131)

Model 2. For the mediation model in which mastery predicted self-esteem and self-esteem predicted depression, the point estimates and confidence intervals of the parameter estimates for the total scale model, the reliability corrected model, the domain representative parcel model, and the homogeneous parcel model are presented in Table 6. The point estimates for the path coefficients were largest in absolute value for the model in which homogeneous parcels were used as indicators. The point estimates of the residual variances were highest for the total scale model and lowest for the homogeneous parcel model. The estimates for the reliability corrected model and the domain representative parcel model were similar. The overall goodness of fit for the domain representative parcel model was good, whereas the overall goodness of fit for the reliability corrected model was marginal to unacceptable and the overall goodness of fit for the homogeneous parcel model and total scale model was unacceptable.

DISCUSSION

The results of both the empirical and artificial data demonstrations show that there is an increase (in absolute value) in the estimates of the structural parameters for the partially disaggregated parcel models when compared to the path analysis models of total scale scores. That is, the path coefficients increase in size and residual variances decrease. This effect is most dramatic when comparing the total scale

TABLE 6
Structural Parameter Estimates for the Mediation Model

Parameter	*Total Scale Model*	*Reliability Corrected Model*	*Domain Representative Parcel Model*	*Homogeneous Parcel Model*
γ	.669	.841	.836	.864
	(.649, .689)	(.817, .865)	(.815, .857)	(.841, .888)
β	–.256	–.355	–.371	–.410
	(–.289, –.223)	(–.398, –.312)	(–.411, –.331)	(–.451, –.369)
ζ_1	.552	.293	.301	.253
	(.527, .579)	(.255, .337)	(.267, .337)	(.214, .296)
ζ_2	.934	.874	.862	.832
	(.918, .952)	(.844, .905)	(.830, .890)	(.795, .863)
Goodness of fit				
χ^2	34.828	17.231	150.169	834.404
df	1	1	25	25
p	<.001	.032	.604	<.001
RMSEA	.126	.088	.049	.124
(90% CI)	(.092, .164)	(.054, .126)	(.041, .056)	(.116, .131)

models to the homogeneous parcel models. In general, the parameter estimates of any of the other models are larger (in absolute value) than those of the total scale score path analysis models. Previous research has suggested that measurement error attenuates parameter estimates and that using a latent variable model with items (totally disaggregated models) as indicators is advantageous because the biasing effects of measurement error are removed (Stephenson & Holbert, 2003). The results of the present study are consistent with previous research but also extend that research by demonstrating that the parameter estimates for latent variable models with parcels as indicators (partially disaggregated models) are similar to those for reliability corrected models and that the parameter estimates for the total scale models are severely attenuated compared to any of the other models due to the biasing effects of measurement error. This leads to the suggestion that applied researchers should utilize the information available to them in measurement scales. Rather than summing the items in a scale to create a total scale score, it is recommended that parcels be created to serve as indicators of latent variables.

Although there is debate in the literature over how to best construct parcels, it is not the primary goal of the present paper to compare different parcel construction methods. The results of the current study show that how the parcels are constructed is less important than the fact that they are used. That is, when comparing either the domain representative parcel model or the homogeneous parcel model to the total scale score path analysis model, the estimates of the path coefficients increase (in absolute value) and residual variances decrease. In addition, the demonstration with artificial data shows that the estimates for either parceling method are more similar to the population values than those for the total scale score path analysis model. The use of parcels to convert path analysis models into latent variable models is one more option for the applied researcher and one that has advantages as discussed here.

Implications for Applied SEM Literature

The results of the present study suggest that many published applications of path analysis could have been converted to latent variable models by using parcels as indicators of latent variables. Furthermore, the results of those studies may have indicated stronger relationships among the constructs if a latent variable model had been used rather than a path analysis model of total scale scores, and these effects may have influenced the substantive interpretation. In fact, it seems plausible that some studies using path analysis of total scale scores may have gone unpublished due to findings of small, nonsignificant, or uninteresting effects involving measured variables, whereas considerably stronger effects may have been found through use of a latent variable model. Given the results of this study, it is not recommended that researchers use path analysis models of total scale scores when testing relationships among constructs measured by multi-item scales.

Similarity of Reliability Corrected and Domain Representative Parcel Models

An interesting result of the present study is the similarity of parameter estimates between the reliability corrected model and the domain representative parcel model in the demonstration with empirical data. The question arises as to whether these parameter estimates will always be similar. It can be shown that if the estimate of $(1-\hat{\rho}_{xx})$ is a good estimate of the unique variance, then the parameter estimates will probably be similar. This will happen when the specific variance is zero or nearly zero because unique variance is composed of both specific variance and random measurement error. If it is the case that the specific variance is nearly zero, then one should be able to use the Spearman-Brown prophecy formula to predict the reliability of the total scale from the parcels, assuming that the parcels are parallel. If these parcels are treated as if they were items, then the Spearman-Brown prophecy formula can be used to predict the reliability of the total scale. This value can then be compared to the actual value for the reliability of the total scale. For the mediation model in which mastery predicts self-esteem and self-esteem predicts depression (Model 2), the point estimates of the unique variances for the mastery parcels are .624, .501, and .417 for the domain representative parcel model. Using the average unique variance across the three parcels (.514), and assuming specific variance to be zero, the average reliability of the parcels is 1 – .514 = .486 and using Equation 1, ρ_{xx} = 3(.486)/[1 + (3 – 1)(.486)] = .74. Since .74 is the predicted reliability of the mastery total scale, 1 – .74 = .26 is an estimate of the error variance for mastery using the Spearman-Brown prophecy formula. Using Cronbach's alpha for the total scale (.74), the $(1-\hat{\rho}_{xx})$ estimate of the error variance was also .26. Thus, the two values are essentially equal. If this were the case for all constructs in a model, then results from analyses using the reliability correction and domain representative parcels would be very similar.

If the specific variance was not zero or nearly zero, then the $(1-\hat{\rho}_{xx})$ estimate may be biased and the parameter estimates for the domain representative parcel model and the reliability corrected model may not be similar. For example, suppose that the $(1-\hat{\rho}_{xx})$ estimate of the error variance was .10 for the mastery measure in the reliability corrected model of Model 2, and thus ρ_{xx} = .90. In this case, suppose also that the true value of the unique variance was .26, and therefore the specific variance was .26 – .10 = .16. Thus, the .10 estimate of the error variance underestimates the unique variance. If .10 is then used as an estimate of the unique variance rather than the original .26 estimate, the point estimate of the path coefficient between mastery and self-esteem (γ) would decrease from .841 to .763 and the point estimate of the residual variance for self esteem (ζ_1) would increase from .293 to .418 (see Table 6 and Figure 2). The parameter estimates for the two models would then be substantially different.

In conclusion, the reliability corrected model and the domain representative parcel model may or may not yield similar parameter estimates depending on whether the specific variance for each measured variable is near zero. It is therefore advantageous to use the partially disaggregated model with parcels as indicators so that one does not have to assume that the specific variances are zero. For the models examined here, the similarity of results held more closely between the reli-

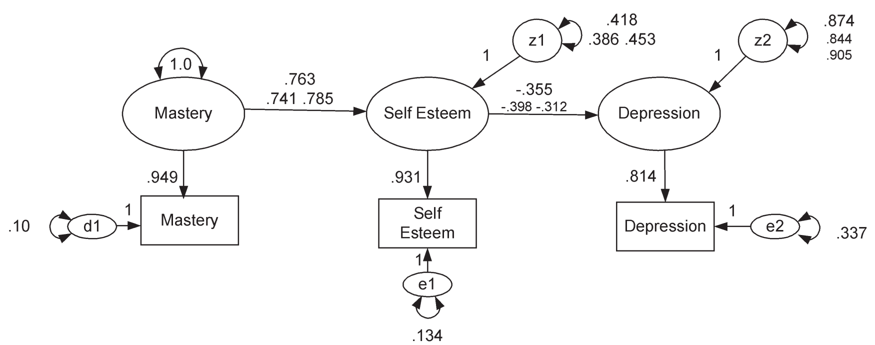

FIGURE 2 Parameter estimates and confidence intervals for the mediation total scale model with the unique variance of mastery fixed to .10.

ability correction model and the models with domain representative parcels than between the reliability correction model and the models with homogeneous parcels. This finding might be attributable to smaller specific variances of the domain representative parcels than the homogeneous parcels. As discussed below, there is some reason to believe that such a relationship might often hold in practice.

The Choice of Reliability Estimate

Throughout the study, Cronbach's alpha has been used as a reliability estimate. One could choose another type of reliability estimate such as coefficient omega (McDonald, 1999), test-retest reliability, alternative forms, or the generalizability coefficient (see DeShon, 1998, for an example of using the generalizability coefficient as an estimate of error variance). The choice of the reliability estimate is irrelevant to the point made here. That is, unreliability is an estimate of error variance rather than unique variance. Furthermore, using reliability estimates from the same sample in which the model is to be fit may introduce sample dependencies, and the generalizability of the results to other samples may be limited. If the reliability estimates are obtained from a different sample, then the issue of the stability of the estimates across samples arises (Raykov & Widaman, 1995). Given these limitations, it is suggested that the partially disaggregated model be used whenever possible rather than the reliability corrected model.

Implications for Overall Model Fit

Another consistent finding of the present study was that for the demonstration with empirical data the models with homogeneous parcels as indicators had consistently poor overall fit, whereas the models with domain representative parcels as indicators had consistently good overall fit (as measured by the RMSEA). Although it may seem counterintuitive at first, one would expect the domain representative parcels to be more similar to one another and the homogeneous parcels to be less similar to one another. In other words, the specific variance for each domain representative parcel should be smaller than the specific variance for each homogeneous parcel. The domain representative parcels contain items from each of the first-order factors and are therefore representative of each of them. When these parcels are then used as indicators of the second-order factor, they have more variability in common and less variability due to a particular parcel. When the parcels are created by selecting items from only one of the first-order factors and these parcels are then used as indicators of the second-order factor, the parcels would tend to have less variability in common and more variability that is due to a particular parcel.

The overall goodness of fit for the reliability corrected model and the total scale model could be evaluated only for the mediation model (Model 2) in the empirical example. The fit for the reliability corrected model was marginal to poor and the fit for the total scale model was poor (as measured by the RMSEA). Both of these models would be rejected based on the χ^2 test with one degree of freedom. In both of these particular cases, it is unlikely that low power is an issue due to the large sample size ($N = 2{,}120$), however low power associated with tests of path analysis models that have low degrees of freedom may be an issue in many instances. Failure to reject a model may mean either that the model fits the data well or that there was not sufficient power to reject the model. Models such as the total scale path analysis models and reliability corrected models typically have few degrees of freedom, and therefore, tests of model fit have low power. MacCallum, Browne, and Sugawara (1996) found that tests of fit of models with low degrees of freedom have low power even when N is reasonably large, and that a very large N is needed to achieve adequate power. Using latent variable models with parcels as indicators increases the degrees of freedom, and therefore, the power of tests of model fit. MacCallum et al. found that for models with moderate to large degrees of freedom, adequate power is achieved with moderate sample sizes.

For the demonstration with artificial data, both the total scale path analysis model and the reliability corrected model would be rejected based on the χ^2 test with two degrees of freedom. The RMSEA value for the total scale path analysis model also indicates poor model fit. In this case, low power should not be an issue given the sample size of $N = 10{,}000$.

Conclusions

When researchers want to evaluate models of relationships among several constructs and each construct is measured by a multi-item scale, then there are several alternatives: total aggregation models of scale scores, models in which the unique variance is fixed to a reliability corrected estimate, and latent variable models in which parcels or items are used as indicators of the latent variables. The different alternatives define the constructs differently. In the path analysis model, the measured variable includes measurement error whereas the reliability corrected model defines the construct as the reliable portion of the measured variable. The latent variable model defines the constructs as common factors that account for the correlations among the indicators. The reliability corrected model is an improvement over the total scale model in that it does not assume that the variables are measured without error. However, it does assume that the specific variance for each measured variable is zero. We recommend that researchers use latent variable models whenever possible because latent variable models do not assume that the variables are measured without error or that the specific variance is zero. In particular, we recommend that researchers use latent variable models in which parcels are used as indicators of the latent variables. Use of latent variable models will reduce the biasing effects of measurement error and will provide more valid, and often substantially higher, estimates of effects among constructs of interest.

REFERENCES

Abbey, A., Andrews, F. M., & Halman, J. L. (1995). Provision and receipt of social support and disregard: What is their impact on the marital quality of infertile and fertile couples. *Journal of Personality and Social Psychology*, *68*(3), 455–469.

Allen, D. G., & Griffeth, R. W. (2001). Test of a mediated performance-turnover relationship highlighting the moderating roles of visibility and reward contingency. *Journal of Applied Psychology*, *86*(5), 1014–1021.

Bagozzi, R. P., & Edwards, J. R. (1998). A general approach for representing constructs in organization research. *Organizational Research Methods*, *1*(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*(1), 35–67.

Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling*, *9*(1), 78–102.

Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.) *New developments and techniques in structural equation modeling* (pp. 269–296). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Browne, M. W., Mels, G., & Cowan, M. (1994). *Path analysis: RAMONA: SYSTAT for DOS advanced applications* (Version 6, pp. 167–224). Evanston, IL: SYSTAT.

Cappella, E., & Weinstein, R. S. (2001). Turning around reading achievement: Predictors of high school students' academic resilience. *Journal of Educational Psychology*, *93*(4), 758–771.

Center for Human Resource Research. (2002). *NLSY79 child and young adult data user's guide*. Columbus: The Ohio State University, CHRR NLS User Services.

Chan, D., Schmitt, N., DeShon, R. P., Clause, C. S., & Delbridge, K. (1997). Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions, and test-taking motivation. *Journal of Applied Psychology*, *82*(2), 300–310.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Harcourt Brace Jovanovich.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, *105*(2), 317–327.

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

DeShon, R. P. (1998). A cautionary note on measurement error corrections in structural equation models. *Psychological Methods*, *3*(4), 412–423.

Fuller, J. B., Jr., & Hester, K. (2001). A closer look at the relationship between justice perceptions and union participation. *Journal of Applied Psychology*, *86*(6), 1096–1105.

Gong, Y., Shenkar, O., Luo, Y., & Nyaw, M. K. (2001). Role conflict and ambiguity of CEOs in international joint ventures: A transaction cost perspective. *Journal of Applied Psychology*, *86*(4), 764–773.

Gribbons, B. C., & Hocevar, D. (1998). Levels of aggregation in higher level confirmatory factor analysis: Application for academic self-concept. *Structural Equation Modeling*, *5*(4), 377–390.

Hall, R. J., Snell, A. F., & Foust, M. S. (1999). Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods*, *2*(3), 233–256.

Hill, M. S., & Fischer, A. R. (2001). Does entitlement mediate the link between masculinity and rape-related variables? *Journal of Counseling Psychology*, *48*(1), 36–50.

Hofmann, D. A., & Morgeson, F. P. (1999). Safety-related behavior as a social exchange: The role of perceived organizational support and leader-member exchange. *Journal of Applied Psychology*, *84*(2), 286–296.

Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453.

Kahn, J. H. (2001). Predicting the scholarly activity of counseling psychology students: A refinement and extension. *Journal of Counseling Psychology*, *48*(4), 344–354.

Kishton, J. M., & Widaman, K. F. (1994). Unidimensional verses domain representative parceling of questionnaire items: An empirical example. *Educational and Psychological Measurement*, *54*(3), 757–765.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford.

Lee, R. M., & Lui, H. T. (2001). Coping with intergenerational family conflict: Comparison of Asian American, Hispanic, and European American college students. *Journal of Counseling Psychology*, *48*(4), 410–419.

Lehrman-Waterman, D., & Ladany, N. (2001). Development and validation of the evaluation process within supervision inventory. *Journal of Counseling Psychology*, *48*(2), 168–177.

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods*, *4*(2), 192–211.

Lopez, D. F., & Little, T. D. (1996). Children's action-control beliefs and emotional regulation in the social domain. *Developmental Psychology*, *32*(2), 299–312.

Lopez, F. G. (2001). Adult attachment orientations, self-other boundary regulation, and splitting tendencies in a college sample. *Journal of Counseling Psychology*, *48*(4), 440–446.

Macan, T. H. (1994). Time management: Test of a process model. *Journal of Applied Psychology*, *79*(3), 381–391.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*, 201–226.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149.

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.

Marsh, H. W., Kong, C. K., & Hau, K. T. (2001). Extension of the internal/external frame of reference model of self-concept formation: Importance of native and nonnative languages for Chinese students. *Journal of Educational Psychology*, *93*(3), 543–553.

Masterson, S. S. (2001). A trickle-down model of organizational justice: Relating employees' and customers' perceptions of and reactions to fairness. *Journal of Applied Psychology*, *86*(4), 594–604.

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

Miller, D. C., & Byrnes, J. P. (2001). To achieve or not to achieve: A self-regulated perspective on adolescents' academic decision making. *Journal of Educational Psychology*, *93*(4), 677–685.

Mulaik, S. (2000, October). *The part truth of parcels*. Paper presented at the annual meeting of the Society of Multivariate Experimental Psychology, Saratoga Springs, NY.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Pajares, F., & Miller, M. D. (1995). Mathematics self-efficacy and mathematics performances: The need for specificity of measurement. *Journal of Counseling Psychology*, *42*(2), 190–198.

Paullay, I. M., Alliger, G. M., & Stone-Romero, E. F. (1994). Construct validation of two instruments designed to measure job involvement and work centrality. *Journal of Applied Psychology*, *79*(2), 224–228.

Perrone, K. M., & Worthington, E. L. (2001). Factors influencing ratings of marital quality by individuals within dual-career marriages: A conceptual model. *Journal of Counseling Psychology*, *48*(1), 3–9.

Plummer, B. A. (2000). *To parcel or not to parcel: The effects of item parceling in confirmatory factor analysis*. Unpublished doctoral dissertation, University of Rhode Island.

Portello, J. Y., & Long, B. C. (2001). Appraisals and coping with workplace interpersonal stress: A model for women managers. *Journal of Counseling Psychology*, *48*(2), 144–156.

Raykov, T., & Widaman, K. F. (1995). Issues in applied structural equation modeling research. *Structural Equation Modeling*, *2*(4), 289–318.

Rioux, S. M., & Penner, L. A. (2001). The causes of organizational citizenship behavior: A motivational analysis. *Journal of Applied Psychology*, *86*(6), 1306–1314.

Rose, S. A., & Feldman, J. F. (1997). Memory and speed: Their role in the relation of infant information processing to later IQ. *Child Development*, *68*(4), 630–641.

Shore, L. M., Tetrick, L. E., Sinclair, R. R., & Newton, L. A. (1994). Validation of a measure of perceived union support. *Journal of Applied Psychology*, *79*(6), 971–977.

Smith, G. T., Goldman, M. S., Greenbaum, P. E., & Christiansen, B. A. (1995). Expectancy for social facilitation from drinking: The divergent paths of high-expectancy and low-expectancy adolescents. *Journal of Abnormal Psychology*, *104*(1), 32–40.

Stephenson, M. T., & Holbert, R. L. (2003). A monte carlo simulation of observable- versus latent-variable structural equation modeling techniques. *Communication Research*, *30*, 332–354.

Stice, E., & Barrera, M., Jr. (1995). A longitudinal examination of the reciprocal relations between perceived parenting and adolescents' substance use and externalizing behaviors. *Developmental Psychology*, *31*(2), 322–334.

Tateneni, K., Mels, G., Cudeck, R., & Browne, M. W. (1998). *CEFA: Comprehensive exploratory factor analysis* [Computer software]. Columbus, OH: Author.

Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.

VandeWalle, D., Cron, W. L., & Slocum, J. W. (2001). The role of goal orientation following performance feedback. *Journal of Applied Psychology*, *86*(4), 629–640.

Weisz, J. R., Southam-Gerow, M. A., & McCarty, C. A. (2001). Control-related beliefs and depressive symptoms in clinic-referred children and adolescents: Developmental differences and model specificity. *Journal of Abnormal Psychology*, *110*(1), 97–109.

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.

Whittaker, A. E., & Robitschek, C. (2001). Multidimensional family functioning: Predicting personal growth initiative. *Journal of Counseling Psychology*, *48*(4), 420–427.

Widaman, K. F., Gibbs, K. W., & Geary, D. C. (1987). Structure of adaptive behavior: I. Replication across fourteen samples of nonprofoundly mentally retarded people. *American Journal of Mental Deficiency*, *91*(4), 348–360.

Accepted August 2004