



Published in final edited form as:

Appl Psychol Meas. 2010 March 1; 34(2): 122–142. doi:10.1177/0146621609338592.

Three Approaches to Using Lengthy Ordinal Scales in Structural Equation Models: Parceling, Latent Scoring, and Shortening Scales

Chongming Yang, Sandra Nay, and Rick H. Hoyle
Duke University

Abstract

Lengthy scales or testlets pose certain challenges for structural equation modeling (SEM) if all the items are included as indicators of a latent construct. Three general approaches (parceling, latent scoring, and shortening) to modeling lengthy scales in SEM were reviewed and evaluated. A hypothetical population model was simulated containing two exogenous constructs with 14 indicators each and an endogenous construct with four indicators. The simulation generated data sets with varying numbers of response options, two types of distributions, factor loadings ranging from low to high, and sample sizes ranging from small to moderate. The population model was varied to incorporate one of the following: (1) single parcels, (2) various parcels as indicators of two exogenous constructs, (3) latent scores as observed exogenous variables, and (4) four and six of individual items as indicators of two exogenous constructs. The dependent variables evaluated were biases in the covariance and partial covariance population parameters. Biases in these parameters were found to be minimal under the following conditions: (1) when parcels of indicators of five response options were used as indicators of two latent exogenous constructs; (2) when latent scores were used as observed variables at sample sizes above 100 and with indicators that were relatively less skewed in the case of dichotomous indicators; and (3) when four or six individual items with high or diverse factor loadings were used as indicators of two exogenous constructs. These findings provided guidelines for resolving the inconsistency of findings from applying various approaches to empirical data.

Keywords

Testlets; Latent Scores; Scale Length; Growth Modeling; Structural Equation Modeling

Lengthy ordinal scales or testlets pose challenges for structural equation modeling (SEM) if all the items are used as indicators of a latent construct. For instance, a model could have too many parameters to estimate relative to the available sample size, resulting in reduced power to detect important parameters. In addition, it might not fit the data sufficiently well because individual items may have less than ideal measurement properties, leading to the rejection of a plausible model. Three general approaches that can be used to address these challenges are parceling, shortening, and transforming a lengthy scale into latent score variables through preliminary analyses. Each approach has advantages and disadvantages. The research reported here was designed to evaluate these approaches under certain data conditions to determine which ones best reflect the true model parameters, particularly the covariance of the exogenous constructs and partial covariances between the exogenous and endogenous constructs.

Parceling

The prevailing approach to incorporating a lengthy scale into SEM has been using the mean or sum of the scale as an indicator of a latent construct (Cattell, 1956). Empirical justifications for parceling include increasing reliability, achieving normality, adapting to small sample sizes, reducing idiosyncratic influence of individual items, simplifying interpretation, and obtaining better model fit (Bandalos & Finney, 2001). Methods for parceling include parceling all items into a single parcel, splitting all odd and even items into two parcels, balancing item discrimination and difficulty across three or four parcels (e.g., item-to-construct balance; Little, Cunningham, Shahar & Widaman, 2002), randomly selecting a certain number of items to create three or four parcels (e.g., Krishton & Widman, 1994; Nasser & Wisenbaker, 2003), and parceling items that have similar factor loadings (i.e., contiguity; Cattell & Burdsal, 1975). Desirable conditions for parceling that have been identified so far include having more than 12 items (Marsh, Hau, Balla, & Grayson, 1998) and having items that reflect a unidimensional construct (Hall, Snell, & Foust, 1999),

Psychometrically, parceling has been controversial despite the above advantages and its prevalence in practice (Bandalos & Finney, 2001; Little et al., 2002). One concern was that parceling results in a loss of information about the relative importance of individual items (Marsh & O'Neill, 1984), because items are implicitly weighted equally in parcels (Bollen & Lennox, 1991). Another concern was that parceling of ordinal scales results in indicators with undefined values, potentially changing the original relations between the indicators and latent variables, for instance, from nonlinear to linear relations (Coanders, Satorra, & Saris, 1997). Parceling binary or trichotomous items could result in limited range as opposed to the latent trait scale, thereby biasing variance and covariance parameters in SEM (Wright, 1999). Compared with using individual items, parceling could underestimate the relations of the latent variables if the reliability of the scale is low (Shevlin, Miles, & Bunting, 1997).

Empirically, the effects of parceling ordinal indicators on covariance and partial covariance of latent constructs have been unknown under some conditions. Previous studies on the effects of parceling on estimates of covariance parameters have typically considered only parceling of continuous indicators with equal discrimination functions or factor loadings (e.g., Alhija & Wisenbaker, 2006; Bandalos, 2002; Hau & Marsh, 2004). Under these conditions, certain parceling strategies yielded little differences in the covariance of two latent constructs. However, most lengthy ordinal scales do not have equal item discrimination functions. In the case of continuous indicators, equal loadings imply that the latent construct can explain equal variances and covariances within each parceling condition. Thus, it is not surprising that parceling has performed well in reflecting the covariance or correlation of latent constructs. In the case of ordinal indicators, a recent study parceled 12 ordinal items with six categories, normal distribution, and loadings varying from .65 to .85, and found no difference in predictive utility in a one-exogenous and one-endogenous construct model (Sass & Smith, 2006). While this finding was supportive of parceling, such results could have resulted from the fact that ordinal indicators in the study approached properties of normal continuous indicators (Coanders et al., 1997) and both exogenous and endogenous indicators were parceled identically. Given the limited conditions controlled in previous studies, it remains unclear to what extent various parceling strategies have reflected the true covariances and partial covariances of a multiple exogenous constructs model, when ordinal indicators had a broader range of categories (e.g., two, three, five, and seven) and varied discrimination functions. It could be assumed that the partial covariance parameters would not be unduly biased by parceling normally distributed multiple-category indicators, because such indicators have approximately linear relations with their latent construct; moreover, parceling as a linear transformation typically does not change the indicators' linearity and have performed well (Coanders et al., 1997; Ferrando, 2009). However, no assumptions could be made about the

magnitude and directions of biases of parceling indicators of two or three categories and varied measurement qualities.

Latent Scoring

Lengthy ordinal scales can be transformed through item response theory (IRT) modeling into latent scores for further modeling (Hambleton, Swaminathan, & Rogers, 1991). IRT describes the probabilities that individuals respond to a set of test items given a particular level of ability or personality trait. Samejima (1969) proposed the following two-parameter model (a_i & b_k) for ordinal scales:

$$P = \frac{1}{1 + \exp(-a_i(\theta_j - b_{k-1}))} - \frac{1}{1 + \exp(-a_i(\theta_j - b_k))}$$

where P is the probability that person j responds to particular category of k of an item at a given trait level θ_j , $k = 1, 2, \dots, m$ categories, a_i is a discrimination parameter of an item, b_k is a threshold at which person j has a .50 probability for the chosen response category, and \exp stands for exponentiation. After estimating unknown a_i , b_k , and θ_j , the latent score (θ_j) for each individual can be saved for subsequent modeling. Latent scores obtained through this process are theoretically interval with a normal distribution that best reflects a population. Such transformation has been found more likely to eliminate artifactual effects (Embretson, 1996) and detect legitimate effects (Fletcher, 2005) than using raw scores.

The two-parameter IRT model can be equivalent to confirmatory factor analysis (CFA) with categorical indicators (Takane & De Leeuw, 1987) within the advanced latent variable modeling framework of Mplus (Muthén & Muthén, 1998–2006). CFA in Mplus typically models the probability (P) of choosing a response category (μ) given the individual's latent trait level (η) with a probit link (Φ) and a residual (δ):

$$P(\mu=1|\eta) = \Phi[(-\tau + \lambda\eta)\delta^{-1/2}]$$

where thresholds (τ), factor loadings (λ), and factor scores (η) are conceptually equivalent to the threshold (b), the discrimination parameter (a), and the latent trait score (θ) in the two-parameter IRT model, respectively (Reise, Widaman, & Pugh, 1993). Although certain transformations are needed to obtain exactly the same parameter estimates of a ($=\lambda\delta^{-1/2}$) and b ($=\tau/\lambda$) from a typical two-parameter IRT modeling (Muthén & Muthén, 2006), factor scores obtained from CFA with categorical indicators are equivalent to the latent scores from an IRT modeling. Since CFA has been recommended as the first step of SEM to ensure unidimensionality and measurement quality (Anderson & Gerbing, 1988), and factor scores are byproducts of this process, it would be of pragmatic value to ascertain whether using factor scores to reduce test length for SEM would result in accurate estimates of covariance and partial covariance of latent constructs. (Factor scores and latent scores are used interchangeably hereafter.)

Sample size could be an important factor to consider in using factor scores. Item response modeling requires large samples ($N > 350$) to yield accurate parameter estimates (Reise & Yu, 1990). Similarly, small samples that are typical of social science research might yield less accurate parameter estimates from CFA with categorical indicators. Although the minimum sample size for CFA with categorical indicators depends on model size, distribution of the variables, strengths of relationships, and proportion of missing values (Muthén & Muthén,

2002), factor scores could be assumed to perform well in representing the true relations among latent variables with relatively large samples.

Shortening a Scale

A few indicators of a latent construct with certain content and predictive validity may be selected from a lengthy scale to yield a shortened scale (Moore, Halle, Vandivere, & Marina, 2002), which can be easily incorporated into SEM. According to behavior domain theory (Guttman, 1955; McDonald, 1996), an underlying construct could have an infinite number of indicators. Therefore, a shortened scale is merely a smaller sample of all possible indicators. One need not be overly concerned that the shortened scale may not be commensurate with the large scale in its content validity, because neither is a perfect measure. Empirically, a six-item scale selected from 27 items with empirical data could be equivalent to the full scale (Moore et al., 2002). It has been shown that a smaller number of continuous indicators with moderate diversity of loadings performed as well as six indicators with high diversity of loadings in recovering the true correlation of the two constructs (Little, Lindenberger, & Nesselroade, 1999), although it was unclear to what extent such finding could be generalized to ordinal scales. Short scales can be used in large-scale surveys, in which many constructs are assessed, if they have desirable measurement properties and similar predictive utility to their large source scales (e.g., Stephenson, Hoyle, Palmgreen, & Slater, 2003).

Shortening a widely used scale has raised other issues besides concerns about content validity. One issue was how to determine the minimum number of indicators. From the perspective of model identification, Kenny (1979) advocated that four indicators are the best for a latent construct and “anything more is gravy.” Shortening a large scale to fewer than four items might undermine its content validity, although its predictive validity could be retained (Little, 1999). Another issue was which strategies to apply to reducing the scale length. One could rely on the magnitude of correlations between the endogenous and exogenous construct indicators, as in the study by Moore et al. (2002), or on the width of the behavioral domain the indicators reflect (Little et al., 1999), or on discrimination functions (factor loadings) because measurement errors of the exogenous constructs attenuate associations with other constructs (Bollen & Lennox, 1991). Viewed from these findings and considerations, four and six ordinal indicators appeared to be reasonable choices for shortening. Therefore, four or six items with different measurement properties were selected from a hypothetical lengthy ordinal scale in this study to examine their efficacy in recovering population parameters.

In sum, under conditions of varied discrimination functions or factor loadings, numbers of response categories, distributions, sample sizes, and full SEM models, performances of various approaches to incorporating lengthy ordinal scales into SEM were unknown in reflecting the true covariance and partial covariance of latent constructs. An efficient method to gain such knowledge was to simulate all these conditions with artificial data generated from a known population model, so that various approaches could be applied to these artificial data and evaluated against the population model (Paxton, Curran, Bollen, Kirby, & Chen, 2001). Findings from the simulations can also be applied to empirical data to solve practical issues.

Design and Analysis

Population Model, Parameters, Scales, and Data

The hypothetical population model was set to have two exogenous constructs (F1 and F2) and one endogenous construct (F3), as shown in Figure 1. This model allows estimation of the covariance between the exogenous constructs and the effects of the exogenous constructs on the endogenous construct (partial covariances), as has been used in previous simulation studies (e.g., Bandalos, 2001; Muthén, 1984). Each exogenous construct was measured by 14 items

(x_1 – x_{14} and x_{15} – x_{28}) and the endogenous construct was measured by four items (y_1 – y_4). The factor loadings and structural coefficients are also shown in Figure 1. The population parameters estimated included the partial covariance from F1 to F3 ($\gamma_1 = .50$), the partial covariance from F2 to F3 ($\gamma_2 = .40$), and the covariance between F1 and F2 ($\phi = .20$). The variances of the exogenous constructs and disturbance (d) of the endogenous construct were assigned at .50 and .60, respectively. Multivariate continuous data were first generated from the population model and then categorized into two, three, five, and seven category ordinal variables. The cutting points (thresholds) of the underlying dimension, which are equivalent to z values of a normally distributed random variable, were selected to specify the proportion of each response category and thereby vary the distributions of the categorized variables. The threshold and distribution of each response category are listed in Table 1. Sample sizes considered were 100, 350, and 600, which reflected the size of typical clinical studies and medium-to-large intervention studies. Using the Mplus program (version 4), 100 data sets were generated for each of the 24 conditions (2 levels of distribution \times 4 levels of response category \times 3 levels of sample size). For the sake of comparisons, various approaches to modeling the lengthy scales (parceling, latent scoring, and shortening) were viewed as repeated treatments of the same datasets that had been generated and categorized for different conditions. The acceptable range of average bias in this study was chosen to be $\pm .14$, rounding to .1 in absolute value.

Parcel and Latent Score Creation and Item Selection for Shortened Scales

Parcels were created from indicators of only the exogenous constructs in this study. A *single parcel* was created for each of F1 and F2 by taking the means of all 14 indicators of F1 and F2, respectively. Two *odd/even parcels* were created for each of F1 and F2 by taking the means of the odd- and even-numbered indicators. Three *item-to-construct-balance parcels* were created by selecting an item for each parcel in the order of discrimination functions and repeating the selection process in a reversed order, as indicated by letters in the Balance column in Table 1. Four *random parcels* were created by randomly selecting indicators without replacement. Four and six *adjacent-loading parcels* were created by taking the means of indicators that had contiguous loadings. The final indicators for each parcel are listed in Table 1 and denoted by identical alphabetic letters. Latent scores for the two exogenous constructs (F1 and F2) were obtained by fitting a measurement model of the three constructs of the population model to the generated categorical data. The indicators were specified as categorical and the latent (factor) scores produced from this process were saved in the data files. Four shortened scales were created by selecting four or six individual indicators of each of F1 and F2. The factor loadings of these exogenous construct indicators were chosen to be diverse, high, moderate, and low, as listed in Table 1.

Model Estimation

In the case of single parcels or latent scores, the population model was altered to have two single parcels or two latent scores as the observed exogenous variables (in place of F1 and F2) predicting the endogenous construct (F3). With two or more parcels, the population model was altered to have parcels as indicators of F1 and F2. Means and standard deviations of the parameter estimates and the number of converged solutions were recorded in this process. Mean biases (parameter estimate – population parameter) were reported or calculated for subsequent analyses (Paxton, et al., 2001).

The performances of various estimation methods in SEM have been compared for models with categorical indicators. The number of categories typically selected has been two, three, five, or seven. Under various distributions, the weighted least squares estimator with degrees of freedom adjusted for means and variances (WLSMV) in Mplus has proven to be fairly robust to various categorical indicators with large sample sizes (Muthén & Kaplan, 1985, 1992). In

order to eliminate any effect of estimation method, indicators of the endogenous latent variables were specified as categorical and all the models were estimated with the WLSMV estimator.

Empirical Data

The empirical data were adopted from the Child Development Project, an ongoing longitudinal study of children's social and emotional development (Lansford, Malone, Dodge, Crozier, Pettit, & Bates, 2006). A total of 585 families were recruited from two cohorts in consecutive years, 1987 and 1988, from Nashville and Knoxville, Tennessee, and Bloomington, Indiana. Data collection began the year before the children entered kindergarten and data have been collected annually ever since. A linear growth model of aggression was chosen for the empirical example, because it could provide an opportunity to examine the effects of various approaches on the mean levels of latent constructs. Aggression was measured from kindergarten through adolescence using the Child Behavior Checklist (Achenbach, 1978), with less than 30% attrition at the final measurement of aggression. Our main interest involved examining whether findings from the simulation study could be applied to resolving any practical issues.

Results

Based on converged solutions, the average biases of each approach in the three parameters (γ_1 , γ_2 , and ϕ) for each data condition were recorded and are depicted in Table 2. The standard deviations and numbers of converged solutions for each condition are listed in Table 3.

Parceling

Data in Table 2 reveal a clear pattern and three key findings: First, any parceling of ordinal indicators of two or three response categories overestimated the partial covariances but underestimated the covariance of the exogenous constructs. Second, parceling of indicators with seven response categories underestimated the partial covariances (γ_1 and γ_2) by .10 ~ .24, but more accurately reflected covariance of the two exogenous constructs (ϕ). Third, odd/even split, random parceling (4 or 6 parcels), and item-construct-balance parceling of indicators of five response categories biased the estimates of the effects (γ_1 and γ_2) maximally by .1 in absolute value, and biased the correlation of the two constructs maximally by .13 in absolute value. Parceling of indicators with contiguous factor loadings attenuated the partial covariances (γ_1 and γ_2) by over .1, which appeared to be slightly less advantageous than other parceling approaches. Thus, parceling (odd/even split, random, and item-construct balance) was most effective when categorical indicators had five response categories and parcels were further used as indicators of the latent constructs.

Latent-Scoring

An analysis of variance confirmed that the number of response categories produced some difference in γ_1 [$F(3, 14) = 40.96, p < .01$] and γ_2 [$F(3, 14) = 18.99, p < .01$], type of distribution produced some difference in γ_1 [$F(1, 14) = 16.34, p < .01$], and the two effects were dependent on each other in γ_1 , evidenced by a significant interaction effect between the number of response categories and type of distribution [$F(3, 14) = 10.27, p < .01$]. As shown in Table 2, latent scoring biased the two direct effect estimates maximally by .08 in absolute value in all the data conditions, except in situations with binary indicators and small samples or extremely skewed binary indicators. The average bias in the covariance of F1 and F2 (ϕ) was consistently between .06 and .11 across all the conditions. Thus, the latent scoring approach performed well with large samples and when binary indicators were not extremely skewed.

Shortened Scales

To compare and simplify the information concerning the effects of shortening scales, a repeated measures analysis of variance was used to compare the biases of shortened-scale approaches, including four ways of using four indicators and four ways of selecting six indicators. This analysis yielded a significant within-subject effect (shortening strategies) on γ_1 [$F(7, 147) = 13.98, p < .01$], γ_2 [$F(7, 147) = 10.10, p < .01$], and ϕ [$F(7, 147) = 476.89, p < .01$], and a significant between-subject effect of sample size on γ_1 [$F(1, 21) = 3.95, p < .05$], γ_2 [$F(2, 21) = 8.99, p < .05$], and ϕ [$F(1, 21) = 7.76, p < .01$]. These effects suggest that various shortening strategies and sample sizes yielded different biases in the three parameters. The estimated marginal means showed that selecting four or six items with diverse or high factor loadings for SEM biased all three parameters within the maximum acceptable range of .10. The raw means in Table 2 show that biases could exceed .20 with small sample sizes. The estimated marginal means also show that selecting four or six items with medium factor loadings for the model biased the partial covariances maximally by .08 but underestimated the covariance of F1 and F2 by over .10 on average. Low factor loadings in the four or six items biased γ_1 minimally by .07 on average and the other two parameters (γ_2 and ϕ) minimally by .13. In sum, the optimal conditions for using four or six indicators selected from large scales were when individual items had diverse or high factor loadings and sample sizes were relatively large.

Findings from Empirical Data

The selected approaches produced mixed results when applied to the empirical data. A measurement model of aggression as a latent construct with 20 categorical indicators was estimated and found to fit the data modestly from each wave. Standardized factor loadings for each wave, their reliabilities, and model fit indices are listed in Table 4. Latent scores were obtained from estimating each measurement model. The goodness of fit indices and the factor loadings suggest that these items measured a unidimensional construct. Only three kinds of parceling (single parcel, two odd/even split parcels, and four random parcels) were applied for the measurement of each wave, because the factor loadings of this longitudinal data did not display the same pattern specified for the population model in the simulation. A linear growth model was estimated using the latent scores or single parcels as the observed variables, two odd/even split parcels or four random parcels as indicators of the aggression construct, or six individual items as indicators of the aggression construct. The six individual items were selected to have the highest mean loadings over time, as were marked with \checkmark in Table 4. All the models fit the data acceptably, with CFI = .92 – .95, TLI = .94 – .98, and RMSEA = .06 – .07. The means of the intercept (α_i) and slope (α_s) and the covariance (ϕ_{is}) of the intercept and slope factors differed in their magnitude and statistical significance across different approaches to using the scale. The models of single parcels and odd/even split parcels as indicators of the latent construct found significant upward linear growth, $\alpha_s = .06, z = 3.13, p < .01$ and $\alpha_s = .05, z = 2.96, p < .01$, whereas no growth was found by modeling latent scores, $\alpha_s = .05, z = 1.33, p > .05$, an aggression construct with four random parcels as indicators, $\alpha_s = .01, z = .76, p > .05$, or six individual items as indicators of the aggression construct, $\alpha_s = .06, z = .72, p > .05$.

Discussion

This study was designed to evaluate the effectiveness of three approaches to incorporating lengthy ordinal scales of varied measurement qualities into SEM, with a focus on biases in estimates of partial variances and the covariance between exogenous constructs. Three approaches were similarly effective under their own appropriate conditions in this study. First, parceling approximately reflected the true population parameters when the indicators had five responses categories and parcels were further used as indicators of latent constructs. Desirable parceling procedures included odd/even splitting into two parcels, balancing item

discrimination functions (factor loadings) to form three parcels, and random parceling into four or six parcels. An ineffective procedure under the conditions of this study was forming four or six parcels of items with contiguous factor loadings. Second, latent scoring best reflected the true parameters in most data conditions of this study except in the case of binary indicators of small samples or extremely skewed binary indicators even with large samples. Third, scale length could be shortened to four or six items with diverse or high factor loadings to adequately recover the population parameters with medium to large samples. The best approach depended on the available data conditions.

In contrast to previous findings that used continuous data, identical measurement of both the exogenous and endogenous constructs, or a measurement model rather than a full SEM, biases of parceling in the two partial covariance parameters were found to be severely upward when the indicators had two or three response categories, but slightly downward when the indicators had seven categories. Besides differences in data, model, and identical parceling of indicators of both exogenous and endogenous constructs, these inconsistencies could also be attributed to the fact that parceling changed the original nonlinear functions between categorical indicators and latent constructs into linear functions, resulting in transformation errors. These errors became large with binary indicators, whose variances and means are dependent on each other and typically bound into a range from 0 to 1 (Ferrando, 2009). Parceling of trichotomous indicators could result in distorted variance and covariance, because the range of the parcel was limited to the original scale, as opposed to that of the latent construct it reflected, which is theoretically infinite. Parceling under these conditions would lose information on variances and covariances of the items and thus bias the estimates of partial covariances more severely.

Consistent with previous findings based on continuous data, optimal conditions for parceling in this study were found to be when parcels were created from items having five categories and used as indicators of the latent constructs, regardless of parceling strategies and degree of nonnormality (e.g., Sass & Smith, 2006). Parceling of five-category indicators could increase the range of the parcel to approximately four (± 2) standard deviations of a normal distribution. As a linear transformation of the categorical variables into continuous ones, parceling did not lose much information on variances and covariances and thus yielded estimates that approximated the population partial covariance parameters (Ferrando, 2009).

Small sample size appeared to be disadvantageous to both latent scoring and shortened-scale approaches to SEM. The latent scoring approach produced trivial biases in the estimates of effects, except when the measurement model for obtaining the latent scores had binary indicators and was estimated with small samples, or when the binary indicators were extremely skewed even with large samples. As expected, a small sample typically could not offer sufficient information for any model to describe its population with accurate parameters. This problem could be complicated with binary indicators, whose means and variances are dependent on each other. Simply increasing the sample size did not improve the information that severely skewed binary indicators could provide about the population, partly because certain information might have been lost when multivariate continuous data were categorized into these severely skewed indicators. In spite of these limitations, biases of latent scores in the covariance (φ) and the two partial covariances (γ_1 and γ_2) were not as far from the population parameters as those caused by using parcels. With large samples, shortened scales with either good or diverse measurement qualities also became robust in recovering the population parameters.

Findings from the simulation study could shed some light on results of linear growth modeling of the empirical data that adopted several approaches to the lengthy aggression scale. Although discrimination functions of the aggression scale did not match exactly those of the simulated data either in magnitude or pattern, no increase in aggression was consistently detected by

modeling the latent scores or factors with six individual indicators of good measurement qualities. These two approaches had performed equivalently well in recovering the population parameters with the simulated data under favorable sample sizes. In addition, no increase was found by the linear growth model of factors with four random parcels as indicators. Four random parcels as indicators of the aggression construct could have provided more reliable measurement than a single parcel or two parcels as indicators of the construct (Kenny, 1979). These consistent findings suggested that aggression might not have increased over time. In contrast, a significant increase in aggression over time was detected with a linear growth modeling of single parcels or a latent construct with two odd/even split parcels as indicators. This finding corresponded to the overestimation of the partial covariances when parcels of trichotomous indicators were used as indicators in the simulation study. Therefore, using findings from the simulation study as guidelines, the efficient approaches to the empirical data problem were (1) latent growth modeling of factors reflected by a shortened scale with the best six indicators, or (2) modeling of latent scores of the full scale, which led to the inference that aggression did not increase over time in this sample.

Conclusions

If all items of a lengthy ordinal scale or testlets with various item discrimination functions are to be included for SEM, parceling is one desirable option under conditions that items have five response categories and are parceled by odd/even splitting the scale, balancing item discrimination functions, or randomizing the items; and two to six parcels obtained through parceling are used as indicators of a construct. Another option is to obtain latent scores through preliminary item response modeling or CFA with categorical indicators based on medium or large samples. However, binary items should not be extremely skewed. If some items of a lengthy scale are not desirable, four or six items of diverse or good measurement properties may be selected robustly as indicators of a construct, preferably based on a medium or large sample.

Acknowledgments

This research was supported by National Institute on Drug Abuse (NIDA) Grants P20 DA017589 and P30 DA023026. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIDA.

References

- Achenbach TM. Child-Behavior Profile: I. Boys aged 6–11. *Journal of Consulting and Clinical Psychology* 1978;46:478–488. [PubMed: 670491]
- Alhija FN, Wisenbaker J. A Monte Carlo study investigating the impact of item parceling strategies on parameter estimates and their standard errors in CFA. *Structural Equation Modeling* 2006;13:204–228.
- Anderson JC, Gerbing DW. Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin* 1988;103(3):411–423.
- Bandalos DL. The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling* 2002;9:78–102.
- Bandalos, DL.; Finney, SJ. Item parceling issues in structural equation modeling. In: Marcoulides, GA.; Schumacker, RE., editors. *New development and techniques in structural equation modeling*. Mahwah, NJ: Erlbaum; 2001. p. 269-275.
- Bollen K, Lennox R. Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin* 1991;110:305–314.
- Cattell RJ. Validation and intensification of the sixteen personality factor questionnaire. *Journal of Clinical Psychology* 1956;12:205–214. [PubMed: 13332084]
- Cattell RJ, Burdsal CA Jr. The radial parceling double factoring design: A solution to the item-vs.-parcel controversy. *Multivariate Behavioral Research* 1975;10:165–179.

- Coanders G, Satorra A, Saris WE. Alternative approaches to structural equation modeling of ordinal data: A Monte Carlo study. *Structural Equation Modeling* 1997;4:261–282.
- Embretson SE. Item response theory and inferential bias in multiple group comparisons. *Applied Psychological Measurement* 1996;20:201–212.
- Embretson, SE.; Reise, SP. *Item response theory for psychologists*. Mahwah, NJ: Erlbaum; 2000.
- Fletcher TD. The effects of parcels and latent variable scores on the detection of interactions in structural equation modeling. *Dissertation Abstracts International*: 66–5B 2005:2872.
- Ferrando P. Difficulty, discrimination, and information indices in the linear factor analysis model for continuous item responses. *Applied Psychological Measurement* 2009;33(1):9–24.
- Guttman L. The determinacy of factor score matrices with implications for five other basic problems of common-factor theory. *British Journal of Statistical Psychology* 1955;8:65–81.
- Hall RJ, Snell AF, Foust MS. Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods* 1999;2:757–765.
- Hambleton, RK.; Swaminathan, H.; Rodgers, HJ. *Fundamentals of item response theory*. Thousand Oaks, CA: Sage Publications; 1991.
- Hau KT, Marsh HW. The use of item parcels in structural equation modeling: Non-normal data and small sample sizes. *British Journal of Mathematical Statistical Psychology* 2004;57:327–351.
- Kenny, DA. *Correlation and causality*. New York: Wiley; 1979.
- Kishton JM, Widaman KF. Unidimensional versus domain representative parceling of questionnaire items: An empirical example. *Educational and Psychological Measurement* 1994;54:757–765.
- Lansford JE, Malone PS, Dodge KA, Crozier JC, Pettit GS, Bates JE. A 12-year prospective study of patterns of social information processing problems and externalizing behaviors. *Journal of Abnormal Child Psychology* 2006;34:715–724. [PubMed: 17053997]
- Little TD, Cunningham WA, Shahar G, Widaman KF. To parcel or not to parcel: Exploring the questions, weighing the merits. *Structural Equation Modeling* 2002;9:151–173.
- Little TD, Lindenberger U, Nesselrode JR. On selecting indicators for multivariate measurement and modeling with latent variables: When “good” indicators are bad and “bad” indicators are good. *Psychological Methods* 1999;4:192–211.
- Marsh HW, Hau K, Balla JR, Grayson D. Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research* 1998;33:181–220.
- Marsh HW, O’Neill R. Self Description Questionnaire III: The construct validity of multidimensional self-concept ratings by late adolescents. *Journal of Educational Measurement* 1984;21:153–174.
- McDonald RP. Latent traits and the possibility of motion. *Multivariate Behavioral Research* 1996;31:593–601.
- Moore KA, Halle TG, Vandivere S, Marina CL. Scaling back survey scales: How short is short? *Sociological Research Methods* 2002;30:530–567.
- Muthén B. A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika* 1984;49:115–132.
- Muthén B, Kaplan D. A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology* 1985;38:171–189.
- Muthén B, Kaplan D. 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* 1992;45:19–30.
- Muthén BO, Muthén LK. How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling* 2002;9(4):599–620.
- Muthén, LK.; Muthén, BO. *Mplus user’s Guide*. 4. Los Angeles, CA: Muthén & Muthén; 1998–2006.
- Muthén, LK.; Muthén, BO. *IRT in Mplus*. 2006. Retrieved December 06, 2008, from <http://www.statmodel.com/download/MplusIRT1.pdf>
- Nasser F, Wisenbaker J. A Monte Carlo study investigating the impact of item parceling on measures of fit in confirmatory factor analysis. *Educational and Psychological Measurement* 2003;63:729–757.
- Paxton P, Curran PJ, Bollen KA, Kirby J, Chen F. Monte Carlo experiments: Design and implementation. *Structural Equation Modeling* 2001;8:287–312.

- Reise SP, Widaman KF, Pugh RH. Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin* 1993;114:552–566. [PubMed: 8272470]
- Reise SP, Yu J. Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement* 1990;65:143–151.
- Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*. 1969;(17)
- Sass DA, Smith PL. The effects of parceling unidimensional scales on structural parameter estimates in structural equation modeling. *Structural Equation Modeling* 2006;13:566–586.
- Shevlin M, Miles JNV, Bunting BP. Summated rating scales: A Monte Carlo investigation of the effects of reliability and collinearity in regression models. *Personality and Individual Differences* 1997;23:665–676.
- Stephenson MT, Hoyle RH, Palmgreen P, Slater MD. Brief measures of sensation seeking for screening and large-scale surveys. *Drug and Alcohol Dependence* 2003;72:279–286. [PubMed: 14643945]
- Takane Y, de Leeuw J. On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika* 1987;52:393–408.
- Wright, BD. Fundamental measurement for psychology. In: Embretson, SE.; Hershberger, SL., editors. *The new rules of measurement: What every psychologist and educator should know*. Mahwah, NJ: Erlbaum; 1999. p. 65-104.

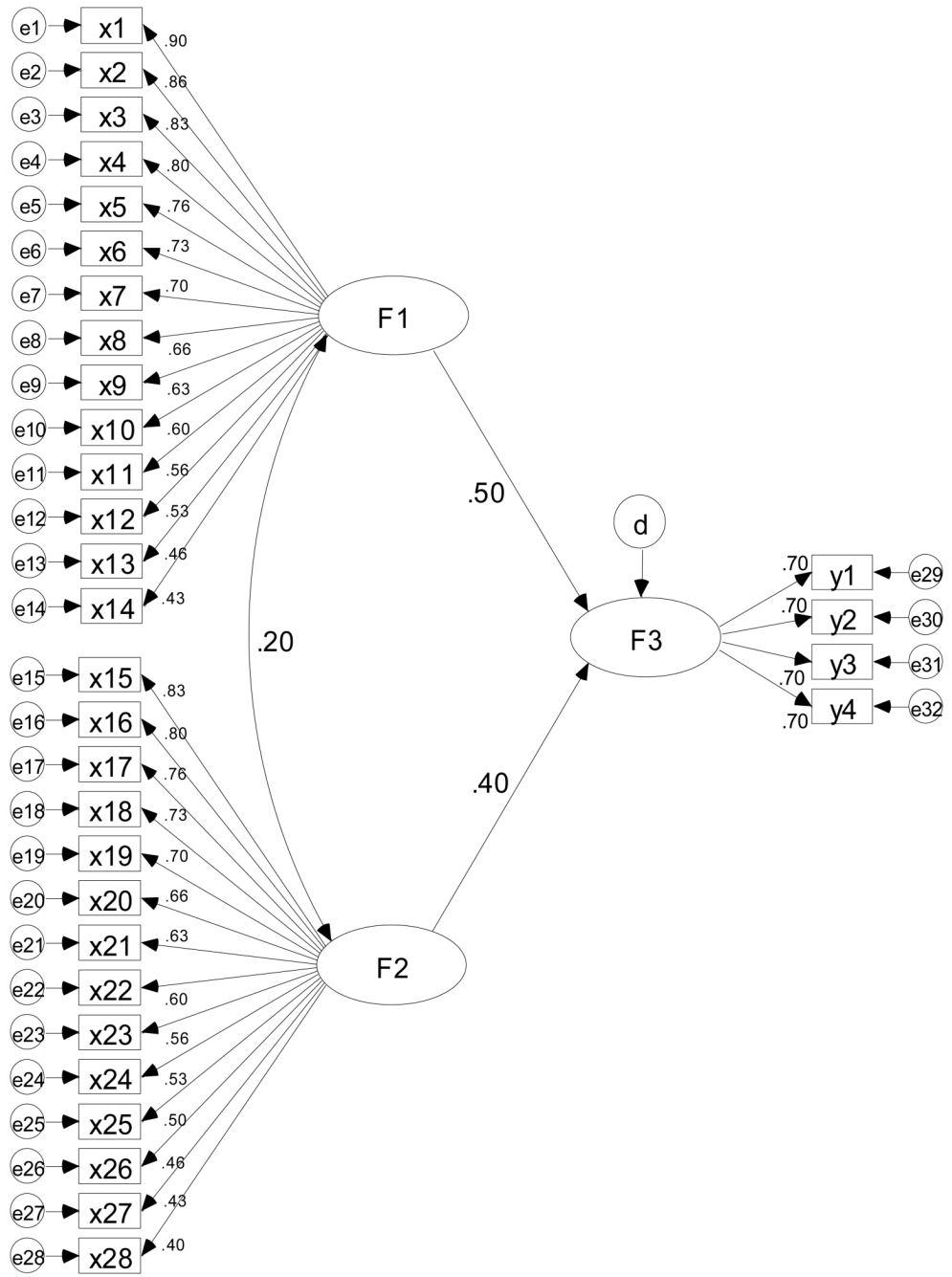


Figure 1.
Population Model

Table 1
Population Indicator Loadings, Items Chosen for Each Parcel, and Item Loadings of Shortened Scales

Population Indicator	Loading	Odd/Even Split	Balance	Parcels			Shortened Scale of Four Indicators				Shortened Scale of Six Indicators				
				Random Four	Random Six	Contiguous Four	Contiguous Six	Diversity	High	Med	Low	Diversity	High	Med	Low
X1	.90	A	A	A	A	A	A	.90	.90	.90	.90	.90	.90	.90	.90
X2	.86	B	B	A	B	A	A	.86	.86	.86	.86	.86	.86	.86	.86
X3	.83	A	C	C	C	A	A	.83	.83	.83	.83	.83	.83	.83	.83
X4	.80	B	C	D	D	A	B	.80	.80	.80	.80	.80	.80	.80	.80
X5	.76	A	B	B	E	B	B	.76	.76	.76	.76	.76	.76	.76	.76
X6	.73	B	A	B	E	B	B	.73	.73	.73	.73	.73	.73	.73	.73
X7	.70	A	A	D	F	B	C	.70	.70	.70	.70	.70	.70	.70	.70
X8	.66	B	B	B	F	B	C	.66	.66	.66	.66	.66	.66	.66	.66
X9	.63	A	C	A	B	C	D	.63	.63	.63	.63	.63	.63	.63	.63
X10	.60	B	C	B	C	C	D	.60	.60	.60	.60	.60	.60	.60	.60
X11	.56	A	B	A	A	C	E	.56	.56	.56	.56	.56	.56	.56	.56
X12	.53	B	A	C	D	D	E	.53	.53	.53	.53	.53	.53	.53	.53
X13	.46	A	C	A	F	D	F	.46	.46	.46	.46	.46	.46	.46	.46
X14	.43	B	B	D	E	D	F	.43	.43	.43	.43	.43	.43	.43	.43
X15	.83	C	D	D	G	E	G	.83	.83	.83	.83	.83	.83	.83	.83
X16	.80	D	E	B	H	E	G	.80	.80	.80	.80	.80	.80	.80	.80
X17	.76	C	F	C	H	E	G	.76	.76	.76	.76	.76	.76	.76	.76
X18	.73	D	F	A	G	E	H	.73	.73	.73	.73	.73	.73	.73	.73
X19	.70	C	E	D	I	F	H	.70	.70	.70	.70	.70	.70	.70	.70
X20	.66	D	D	A	J	F	H	.66	.66	.66	.66	.66	.66	.66	.66
X21	.63	C	D	B	K	F	I	.63	.63	.63	.63	.63	.63	.63	.63
X22	.60	D	E	C	K	F	I	.60	.60	.60	.60	.60	.60	.60	.60
X23	.56	C	F	A	J	G	J	.56	.56	.56	.56	.56	.56	.56	.56
X24	.53	D	F	A	L	G	J	.53	.53	.53	.53	.53	.53	.53	.53
X25	.50	C	E	D	I	G	K	.50	.50	.50	.50	.50	.50	.50	.50
X26	.46	D	D	B	K	H	K	.46	.46	.46	.46	.46	.46	.46	.46
X27	.43	C	F	C	H	H	L	.43	.43	.43	.43	.43	.43	.43	.43

Population	Parcels						Shortened Scale of Four Indicators						Shortened Scale of Six Indicators					
	Indicator	Loading	Odd/Even Split	Balance	Random Four	Random Six	Contiguous Four	Contiguous Six	Diversity	High	Med	Low	Diversity	High	Med	Low		
X28	.40	D	E	B	L	H	L		.40				40			.40		

Note: The same letters in the table designate indicators for the same parcel.

Table 2
 Absolute Bias in Population Model Parameters of Each Approach to Lengthy Scales under Different Data Conditions

Cat.	Dataset Conditions/Approaches	Latent Scores			Single Parcel			Even-Split Parcels			Item-Construct Balance			
		N	$\gamma 1$	$\gamma 2$	ϕ	$\gamma 1$	$\gamma 2$	ϕ	$\gamma 1$	$\gamma 2$	ϕ	$\gamma 1$	$\gamma 2$	ϕ
Two	Distribution (%) (Thresholds)	100	.06	.15	-.09	.47	.42	-.19	.97	.76	-.19	.91	.92	-.19
		350	.04	.08	-.10	.42	.39	-.19	.82	.73	-.19	.75	.73	-.19
	70, 30 (.5244)	600	.05	.06	-.11	.41	.37	-.19	.81	.68	-.19	.76	.68	-.19
		100	.17	.19	-.10	.63	.55	-.20	1.26	1.25	-.19	1.26	1.21	-.19
Three	85, 15 (1.036)	350	.13	.14	-.11	.66	.56	-.20	1.25	1.11	-.20	1.21	1.09	-.19
		600	.16	.11	-.11	.68	.58	-.20	1.28	1.13	-.20	1.28	1.09	-.19
	60, 30, 10 (.2553, 1.282)	100	.01	.02	-.08	.10	.10	-.17	.29	.26	-.17	.29	.28	-.17
		350	.03	.07	-.11	.09	.11	-.18	.29	.30	-.18	.26	.28	-.17
Five	70, 25, 5 (.5244, 1.645)	600	.03	.05	-.10	.09	.11	-.18	.28	.28	-.17	.27	.28	-.17
		100	.02	.03	-.06	.10	.16	-.17	.36	.38	-.17	.31	.41	-.17
	10, 25, 40, 25, 10 (-1.282, -.3853, .6745, 1.282)	350	.05	.08	-.11	.19	.19	-.18	.45	.42	-.18	.41	.43	-.18
		600	.03	.06	-.10	.18	.19	-.18	.42	.42	-.18	.40	.40	-.18
Seven	5, 10, 20, 30, 20, 10, 5 (-1.645, -1.036, -.3853, .3853, 1.036, 1.645)	100	.05	.03	-.10	-.14	-.12	-.12	-.03	-.04	-.12	-.03	-.04	-.11
		350	-.02	.02	-.09	-.15	-.11	-.11	-.07	-.03	-.11	-.08	-.03	-.11
	25, 40, 20, 10, 5 (-.6745, .3853, 1.036, 1.645)	600	-.01	.02	-.10	-.15	-.11	-.12	-.06	-.02	-.11	-.08	-.04	-.11
		100	-.02	.04	-.09	-.16	-.11	-.12	-.06	-.02	-.11	-.08	-.01	-.11
	14, 34, 22, 14, 9, 5, 2 (-1.080, -.0515, .5244, .9945, 1.476, 2.054)	350	-.01	.02	-.10	-.16	-.11	-.12	-.07	-.01	-.12	-.08	-.03	-.11
		600	.00	.01	-.10	-.15	-.11	-.12	-.06	-.02	-.11	-.07	-.04	-.11
	10, 25, 40, 20, 10, 5 (-1.645, -1.036, -.3853, .3853, 1.036, 1.645)	100	-.02	.04	-.09	-.26	-.18	-.04	-.19	-.12	-.03	-.14	-.12	-.03
		350	-.02	.02	-.09	-.24	-.18	-.04	-.18	-.13	-.03	-.20	-.13	-.02
	14, 34, 22, 14, 9, 5, 2 (-1.080, -.0515, .5244, .9945, 1.476, 2.054)	600	-.01	.01	-.10	-.24	-.18	-.04	-.18	-.13	-.03	-.19	-.13	-.02
		100	.03	-.02	-.09	-.25	-.21	-.05	-.16	-.16	-.04	-.17	-.15	-.04
		350	.00	.00	-.10	-.24	-.19	-.05	-.18	-.13	-.04	-.19	-.14	-.03

Dataset Conditions/Approaches		Latent Scores		Single Parcel		Even-Split Parcels		Item-Construct Balance				
600		.00	.02	-.10	-.24	-.19	-.05	-.18	-.13	-.03		
Randomized to Four Parcels		Randomized to Six Parcels		Four Parcels of Contiguous Loadings		Six Parcels of Contiguous Loadings		Four Items of Diverse Loadings		Four Items of High Loadings		
γ_1	γ_2	ϕ	γ_1	γ_2	ϕ	γ_1	γ_2	ϕ	γ_1	γ_2	ϕ	
1.26	.97	-.19	1.04	.87	-.19	.76	.65	-.19	.20	.27	-.08	-.01
.97	.76	-.19	.62	1.13	-.19	.58	.56	-.19	-.12	.03	-.08	-.08
.91	.74	-.19	.95	.66	-.19	.57	-.30	-.19	-.09	-.06	-.08	-.07
1.44	1.61	-.20	1.69	1.25	-.20	.95	1.07	-.19	.08	.19	-.09	-.01
1.46	1.16	-.20	1.49	1.08	-.20	.99	.82	-.19	-.07	.03	-.09	-.01
1.51	1.17	-.20	1.57	1.05	-.20	.97	.81	-.19	-.05	-.04	-.10	-.05
.41	.35	-.18	.45	.27	-.17	.19	.16	-.16	.05	-.10	-.07	.00
.38	.30	-.18	.40	.26	-.18	.16	.18	-.16	-.11	.00	-.10	-.03
.39	.30	-.18	.39	.25	-.18	.15	.17	-.16	-.10	-.04	-.09	-.07
.49	.36	-.18	.49	.41	-.17	.21	.24	-.15	.31	-.56	-.04	.20
.59	.45	-.19	.59	.41	-.18	.28	.28	-.17	-.07	.00	-.10	-.02
.54	.45	-.18	.55	.41	-.18	.25	.28	-.17	-.11	-.05	-.09	-.08
.02	-.01	-.12	.06	-.04	-.12	-.10	-.08	-.08	.01	.21	-.11	-.03
-.02	-.02	-.12	.00	-.05	-.12	-.14	-.09	-.07	-.10	-.04	-.09	-.05
-.02	-.02	-.12	-.01	-.05	-.12	-.14	-.10	-.07	-.09	-.07	-.09	-.05
-.02	.01	-.12	.02	.00	-.12	-.13	-.06	-.07	-.05	-.09	-.08	-.03
-.02	-.02	-.13	.00	-.04	-.12	-.15	-.09	-.08	-.08	-.07	-.09	-.06
-.01	-.02	-.12	.00	-.04	-.12	-.13	-.09	-.07	-.10	-.05	-.09	-.07
-.16	-.10	-.05	-.14	-.11	-.05	-.24	-.16	.04	-.05	.05	-.09	.03
-.15	-.11	-.05	-.14	-.13	-.05	-.23	-.17	.04	-.08	-.07	-.08	-.06
-.15	-.13	-.05	-.14	-.14	-.04	-.24	-.18	.05	-.10	-.06	-.09	-.07
-.13	-.14	-.06	-.10	-.15	-.05	-.22	-.19	.04	-.02	-.05	-.08	-.11
-.14	-.13	-.06	-.14	-.14	-.05	-.23	-.18	.03	-.09	-.03	-.09	-.08
-.14	-.13	-.06	-.13	-.14	-.05	-.23	-.18	.04	-.08	-.06	-.09	-.06
Four Items of Medium Loadings		Four Items of Low Loadings		Six Items of Diverse Loadings		Six Items of High Loadings		Six Items of Medium Loadings		Six Items of Low Loadings		
γ_1	γ_2	ϕ	γ_1	γ_2	ϕ	γ_1	γ_2	ϕ	γ_1	γ_2	ϕ	γ_1

	Four Items of Medium Loadings		Four Items of Low Loadings		Six Items of Diverse Loadings		Six Items of High Loadings		Six Items of Medium Loadings		Six Items of Low Loadings						
.11	.13	-.11	-.05	.62	-.12	-.11	.08	-.08	-.02	.02	-.07	.26	.21	-.10	.38	.49	-.12
-.08	.08	-.11	.27	.26	-.15	-.09	-.05	-.08	-.10	-.07	-.09	-.06	.05	-.12	.08	.20	-.14
-.02	.00	-.12	.12	.12	-.15	-.10	-.07	-.09	-.10	-.08	-.09	-.04	-.02	-.12	.03	.07	-.13
.32	-.18	-.08	.05	.45	-.11	.12	.22	-.08	-.04	.13	-.07	.16	.17	-.08	-.03	.13	-.12
.04	.06	-.12	.32	.19	-.14	-.07	-.01	-.09	-.08	-.04	-.09	.01	.06	-.12	.11	.45	-.14
-.06	.08	-.11	.33	.15	-.15	-.06	-.05	-.09	-.09	-.06	-.09	-.03	.03	-.11	.18	.09	-.14
.05	.26	-.09	.48	-.04	-.13	-.03	-.06	-.06	-.08	-.07	-.06	.02	.12	-.09	.15	.16	-.12
-.02	.07	-.12	.25	.05	-.14	-.10	-.05	-.10	-.10	-.05	-.10	-.02	.02	-.12	.10	.06	-.14
-.04	-.02	-.12	.15	.15	-.15	-.09	-.06	-.09	-.10	-.07	-.09	-.04	-.02	-.12	.09	.09	-.14
.02	.09	-.07	.13	.23	-.11	-.12	.03	-.03	-.15	.13	-.03	.05	.01	-.07	.77	2.08	-.11
-.02	.04	-.12	.25	.13	-.15	-.09	-.04	-.09	-.09	-.04	-.09	-.03	.01	-.12	.05	.18	-.14
-.04	.02	-.12	.10	.16	-.15	-.12	-.05	-.09	-.10	-.08	-.09	-.05	.01	-.12	.02	.11	-.14
.13	.00	-.12	.42	.25	-.14	.09	-.08	-.10	-.04	-.05	-.09	.08	.01	-.12	.26	.15	-.13
-.03	-.01	-.11	.08	.12	-.14	-.09	-.07	-.09	-.10	-.05	-.09	-.04	-.01	-.11	.03	.10	-.14
-.03	-.04	-.11	.15	.06	-.15	-.10	-.06	-.09	-.11	-.06	-.09	-.03	-.04	-.11	.04	.05	-.14
.06	.08	-.11	.15	.45	-.13	-.09	-.06	-.07	-.09	-.04	-.08	.01	.08	-.10	.15	-.01	-.12
-.07	.03	-.12	.12	.14	-.15	-.10	-.06	-.09	-.10	-.07	-.09	-.08	.01	-.11	.07	.10	-.14
-.05	-.01	-.12	.12	.08	-.14	-.10	-.05	-.09	-.09	-.08	-.09	-.04	.00	-.12	.05	.10	-.14
-.03	.04	-.11	.10	.30	-.14	-.06	-.04	-.08	-.12	.04	-.09	-.05	.21	-.11	-.107	.64	-.13
-.04	.02	-.11	.14	.10	-.14	-.10	-.06	-.08	-.10	-.06	-.09	-.03	.00	-.11	.04	.08	-.13
-.05	-.01	-.11	.10	.08	-.14	-.10	-.06	-.09	-.10	-.07	-.09	-.06	.01	-.12	.04	.07	-.14
.04	.04	-.12	.31	.24	-.13	-.04	-.13	-.08	-.04	-.09	-.08	.07	-.04	-.12	.13	.07	-.13
-.04	.01	-.11	.13	.10	-.15	-.10	-.07	-.09	-.10	-.07	-.09	-.03	-.01	-.11	.09	.09	-.14
-.03	-.02	-.11	.12	.06	-.14	-.09	-.06	-.09	-.10	-.06	-.09	-.04	-.02	-.11	.03	.06	-.14

Note: Cat. = Number of categories.

Table 3
Standard Deviation of Parameter Estimates under Various Data Conditions and Approaches and Corresponding Sample Sizes (*n*)

Cat.	Dataset Conditions/Approaches	Latent Scores				Single Parcels				Even-Split Parcels				Item-Construct Balance																				
		$\gamma 1$	$\gamma 2$	ϕ	<i>n</i>	$\gamma 1$	$\gamma 2$	ϕ	<i>n</i>	$\gamma 1$	$\gamma 2$	ϕ	<i>n</i>	$\gamma 1$	$\gamma 2$	ϕ	<i>n</i>																	
Two	100 350 600	.65 .20 .16	.69 .26 .16	.06 .03 .02	100 100 100	.40 .19 .15	.38 .20 .14	.00 .00 .00	100 100 100	.85 .35 .28	.82 .39 .28	.00 .00 .00	99 100 100	.89 .33 .28	.92 .39 .27	.00 .00 .00	100 100 100																	
																		70, 30	1.02	.32	.03	100	.54	.26	.00	100	.65	.61	.00	100	.56	.62	.00	100
Three	100 350 600	.49 .18 .14	.62 .19 .16	.05 .02 .02	100 100 100	.23 .13 .09	.26 .12 .10	.01 .00 .00	100 100 100	.46 .22 .15	.48 .21 .19	.01 .01 .00	100 100 100	.46 .22 .16	.57 .21 .18	.01 .01 .01	100 100 100																	
																		20, 60, 20	.78	1.37	.06	100	.26	.30	.01	100	.64	.80	.01	98	.58	.77	.01	100
Five	100 350 600	.33 .15 .11	.32 .15 .13	.04 .02 .01	100 100 100	.14 .07 .06	.12 .07 .06	.03 .02 .01	100 100 100	.21 .11 .08	.20 .10 .09	.03 .02 .01	100 100 100	.23 .10 .08	.20 .10 .09	.04 .02 .01	100 100 100																	
																		5, 15, 60, 15, 5	.29	.41	.05	100	.12	.15	.03	100	.22	.22	.04	99	.19	.25	.04	100
Seven	100 350 600	.37 .15 .11	.41 .17 .13	.04 .03 .02	100 100 100	.11 .06 .04	.12 .06 .06	.05 .03 .02	100 100 100	.08 .09 .08	.10 .06 .06	.02 .03 .03	100 100 100	.08 .09 .06	.09 .09 .06	.02 .03 .03	100 100 100																	
																		5, 10, 20, 30, 20, 10, 5	.42	.35	.04	100	.10	.11	.05	100	.18	.19	.07	100	.17	.18	.07	100
Randomized to Four Parcels	100 350 600	.12 .15 .15	.02 .02 .02	100 100 100	.04 .04 .05	.02 .02 .02	.01 .01 .01	100 100 100	.07 .08 .08	.03 .03 .03	.03 .03 .03	.03 .03 .03	100 100 100	.06 .06 .06	.08 .08 .08	.03 .03 .03	100 100 100																	
																		Four Parcels of Adjacent Loadings	.76	.74	.01	100	.08	.81	.01	100	1.02	2.19	.11	82	.66	.97	.09	96
Four Items of Diverse Loadings	.90	1.03	.00	100	.76	.74	.01	100	.08	.81	.01	100	1.02	2.19	.11	82	.66	.97	.09	96														
																					Four Items of High Loadings	.90	1.03	.00	100	.76	.74	.01	100	.08	.81	.01	100	1.02

	Randomized to Four Parcels			Randomized to Six Parcels			Four Parcels of Adjacent Loadings			Six Parcels of Adjacent Loadings			Four Items of Diverse Loadings			Four Items of High Loadings								
.41	.40	.00	100	.46	.39	.00	100	.31	.34	.00	100	.30	.34	.00	100	.18	.60	.05	98	.18	.18	.18	.05	100
.32	.30	.00	100	.31	.26	.00	100	.25	.23	.00	100	.25	.22	.00	100	.19	.18	.05	100	.19	.14	.13	.04	100
1.29	2.47	.00	100	1.56	1.58	.00	98	1.57	1.68	.00	99	1.20	1.65	.00	100	.65	1.87	.14	84	.65	1.32	.14	.83	
.64	.59	.00	100	.71	.64	.00	100	.48	.50	.00	100	.49	.50	.00	100	.49	.56	.06	98	.26	.35	.05	100	
.49	.51	.00	100	.56	.49	.00	100	.35	.39	.00	100	.37	.38	.00	100	.24	.24	.11	100	.18	.17	.04	100	
.58	.65	.01	100	.59	.51	.01	100	.41	.44	.02	100	.39	.44	.02	100	1.02	.76	.10	86	.43	.96	.08	98	
.24	.21	.01	100	.24	.22	.01	100	.18	.19	.01	100	.18	.19	.01	100	.20	.25	.04	100	.17	.18	.03	100	
.18	.18	.00	100	.18	.17	.01	100	.12	.15	.01	100	.13	.15	.01	100	.12	.18	.03	100	.13	.13	.03	100	
.74	.91	.01	100	.78	.77	.01	100	.50	.70	.02	100	.52	.73	.02	100	3.05	3.48	.13	81	.82	1.00	.09	96	
.34	.31	.00	100	.33	.33	.00	100	.22	.25	.01	100	.22	.25	.01	100	.34	.33	.05	99	.17	.25	.04	100	
.25	.22	.00	100	.26	.22	.00	100	.18	.17	.01	100	.18	.17	.01	100	.15	.17	.04	100	.11	.13	.32	100	
.24	.26	.04	100	.28	.21	.04	100	.19	.18	.05	100	.19	.17	.05	100	.64	1.48	.07	96	.04	.36	.06	100	
.12	.11	.02	100	.13	.10	.02	100	.09	.09	.03	100	.09	.09	.03	100	.17	.20	.04	100	.13	.13	.03	100	
.09	.09	.01	100	.09	.09	.01	100	.07	.07	.02	100	.07	.07	.02	100	.14	.13	.03	100	.09	.11	.02	100	
.21	.27	.03	100	.27	.28	.04	100	.16	.22	.05	100	.16	.22	.05	100	.48	.86	.08	96	.29	.40	.07	99	
.12	.12	.02	100	.13	.12	.02	100	.09	.10	.03	100	.09	.10	.03	100	.18	.17	.04	100	.15	.13	.03	100	
.09	.10	.01	100	.10	.10	.01	100	.07	.08	.02	100	.06	.08	.02	100	.12	.15	.03	100	.10	.10	.03	100	
.19	.20	.06	100	.21	.22	.06	100	.14	.18	.09	100	.15	.19	.09	100	.53	.59	.07	97	.30	.39	.06	98	
.10	.10	.03	100	.10	.10	.03	100	.08	.08	.05	100	.08	.08	.05	100	.15	.21	.04	100	.12	.12	.04	100	
.07	.06	.03	100	.07	.06	.03	100	.05	.05	.04	100	.05	.05	.04	100	.11	.13	.03	100	.10	.09	.02	100	
.18	.20	.06	100	.06	.13	.15	100	.14	.17	.09	100	.15	.17	.09	100	.53	.76	.08	96	.37	.30	.06	99	
.10	.11	.03	100	.10	.10	.04	100	.07	.08	.05	100	.07	.08	.05	100	.17	.21	.04	100	.15	.13	.03	100	
.08	.07	.02	100	.08	.08	.03	100	.06	.07	.04	100	.06	.06	.04	100	.14	.16	.03	100	.10	.11	.02	100	

Four Items of Medium Loadings			Four Items of Low Loadings			Six Items of Diverse Loadings			Six Items of High Loadings			Six Items of Medium Loadings			Six Items of Low Loadings								
$\gamma/1$	$\gamma/2$	ϕ	n	$\gamma/1$	$\gamma/2$	ϕ	n	$\gamma/1$	$\gamma/2$	ϕ	n	$\gamma/1$	$\gamma/2$	ϕ	n	$\gamma/1$	$\gamma/2$	ϕ	n				
1.44	.93	.09	78	1.91	2.46	.09	61	.82	.89	.09	95	.47	.65	.08	100	1.38	2.47	.09	88	1.79	2.01	.08	78
.26	.33	.05	99	1.25	.97	.04	94	.17	.25	.05	100	.13	.17	.04	100	.19	.24	.04	100	.60	.60	.04	100
.18	.20	.04	100	.34	.35	.03	99	.14	.14	.04	100	.12	.11	.03	100	.16	.18	.03	100	.23	.32	.03	100
1.47	1.41	.15	74	1.62	1.3	.12	53	1.31	1.11	.12	91	.44	1.19	.12	91	1.17	1.56	.10	79	1.11	1.42	.11	67
.41	.66	.05	99	.89	1.56	.05	87	.22	.28	.06	100	.19	.20	.04	100	.26	.45	.05	100	.79	1.35	.05	95

	Four Items of Medium Loadings		Four Items of Low Loadings		Six Items of Diverse Loadings		Six Items of High Loadings		Six Items of Medium Loadings		Six Items of Low Loadings												
.24	.41	.04	100	.74	.58	.04	99	.21	.19	.04	100	.14	.14	.03	100	.18	.22	.03	100	.44	.34	.03	100
.96	1.37	.09	92	1.9	1.2	.07	69	.61	.92	.08	97	.31	.39	.07	98	.97	.99	.08	97	1.48	1.89	.07	84
.22	.36	.03	100	.93	.48	.03	99	.16	.16	.04	100	.14	.13	.03	100	.19	.27	.03	100	.39	.34	.03	100
.13	.18	.03	100	.38	.39	.03	100	.12	.14	.03	100	.12	.12	.03	100	.12	.15	.02	100	.25	.28	.02	100
.99	1.16	.07	86	2.56	1.78	.09	76	.66	.80	.10	97	.43	.73	.09	99	.91	.69	.08	95	4.42	17.5	.08	86
.29	.33	.04	100	.68	.69	.04	96	.15	.20	.04	100	.14	.15	.04	100	.20	.21	.04	100	.71	.92	.04	98
.16	.22	.03	100	.32	.36	.03	100	.13	.13	.03	100	.10	.10	.03	100	.14	.18	.03	100	.21	.23	.03	100
.84	1.06	.06	96	1.17	1.64	.06	86	.60	.50	.06	100	.27	.27	.06	100	.37	.47	.05	99	1.4	1.64	.05	95
.21	.20	.03	100	.34	.44	.03	99	.16	.13	.03	100	.12	.13	.03	100	.17	.17	.03	100	.23	.35	.03	100
.12	.13	.02	100	.25	.23	.02	100	.11	.10	.02	100	.08	.10	.02	100	.12	.12	.02	100	.17	.21	.02	100
.54	.43	.07	100	1.25	1.33	.07	87	.46	.61	.07	99	.23	.30	.06	100	.45	.33	.06	100	1.47	1.66	.06	95
.16	.19	.04	100	.39	.63	.03	100	.14	.18	.03	100	.12	.11	.03	100	.15	.18	.03	100	.24	.39	.03	100
.11	.16	.02	100	.31	.38	.02	100	.10	.12	.02	100	.10	.09	.02	100	.11	.15	.02	100	.17	.27	.02	100
.39	.57	.06	96	1.12	1.48	.06	93	.35	.38	.05	100	.30	.43	.05	100	.58	1.19	.06	100	12.9	4.92	.05	96
.19	.23	.03	100	.38	.40	.03	100	.15	.16	.03	100	.11	.11	.04	100	.17	.17	.03	100	.28	.37	.02	100
.13	.15	.02	100	.25	.24	.02	100	.10	.10	.03	100	.09	.08	.02	100	.11	.12	.02	100	.17	.18	.02	100
.59	.70	.07	96	1.79	2.19	.07	84	.50	.74	.06	100	.33	.26	.05	100	.46	.52	.05	100	.61	1.16	.05	90
.19	.22	.03	100	.29	.50	.03	100	.13	.16	.03	100	.13	.13	.03	100	.18	.18	.03	100	.26	.43	.03	100
.14	.13	.02	100	.29	.25	.02	100	.11	.13	.02	100	.09	.10	.02	100	.12	.12	.02	100	.19	.24	.02	100

Note: Cat. = Number of categories.

Table 4
Standardized Factor Loadings of the Latent Construct of Aggression Measured from Kindergarten to Grade 8

Item & Contents	Kindergarten	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
3. Argues	.86	.88	.93	.91	.92	.87	.88	.87	.91
7. Brags	.75	.73	.70	.75	.70	.80	.75	.73	.81
16. Cruel to others ✓	.88	.94	.92	.94	.93	.92	.89	.87	.94
19. Demands attentions	.76	.81	.84	.80	.79	.81	.82	.74	.89
20. Destroys own things	.80	.76	.78	.80	.79	.82	.76	.63	.87
21. Difficulty following directions	.84	.84	.85	.90	.84	.86	.90	.84	.92
23. Disobedient at School ✓	.87	.85	.91	.91	.90	.91	.88	.87	.90
27. Easily jealous	.69	.82	.81	.69	.73	.71	.74	.81	.82
37. Fights ✓	.90	.89	.91	.92	.89	.83	.85	.87	.88
43. Lies	.79	.75	.87	.79	.84	.74	.82	.72	.83
57. Physically attacks ✓	.87	.81	.89	.91	.88	.91	.87	.92	.88
68. Screams	.82	.69	.89	.90	.83	.82	.88	.81	.93
74. Shows off	.82	.82	.80	.77	.75	.79	.81	.84	.87
86. Stubborn	.82	.85	.87	.84	.86	.85	.86	.84	.76
87. Moody	.68	.78	.82	.79	.81	.81	.87	.82	.76
93. Talks too much	.78	.77	.80	.75	.82	.80	.81	.83	.88
94. Teases	.78	.88	.82	.83	.85	.81	.85	.86	.86
95. Hot temper ✓	.85	.88	.86	.89	.89	.89	.89	.91	.88
97. Threatens ✓	.90	.85	.89	.91	.91	.92	.88	.93	.94
104. Loud	.87	.86	.83	.84	.87	.78	.88	.89	.93
Measurement Model Fit →	$\chi^2 = 258.11$	$\chi^2 = 245.31$	$\chi^2 = 191.04$	$\chi^2 = 188.40$	$\chi^2 = 180.81$	$\chi^2 = 204.35$	$\chi^2 = 179.02$	$\chi^2 = 131.79$	$\chi^2 = 150.45$
	df = 50	df = 48	df = 50	df = 53	df = 42	df = 52	df = 43	df = 32	df = 37
	$p < .01$	$p < .01$	$p < .01$	$p < .01$	$p < .01$	$p < .01$	$p < .01$	$p < .01$	$p < .01$
	CFI = .94	CFI = .94	CFI = .96	CFI = .97	CFI = .96	CFI = .94	CFI = .95	CFI = .96	CFI = .97
	TLI = .98	TLI = .98	TLI = .99	TLI = .99	TLI = .98	TLI = .98	TLI = .98	TLI = .98	TLI = .99
	RMSEA = .08	RMSEA = .08	RMSEA = .07	RMSEA = .07	RMSEA = .08	RMSEA = .08	RMSEA = .08	RMSEA = .08	RMSEA = .08