

Random Intercept Item Factor Analysis

Albert Maydeu-Olivares
University of Barcelona and Instituto de Empresa

Donna L. Coffman
Pennsylvania State University

The common factor model assumes that the linear coefficients (intercepts and factor loadings) linking the observed variables to the latent factors are fixed coefficients (i.e., common for all participants). When the observed variables are participants' observed responses to stimuli, such as their responses to the items of a questionnaire, the assumption of common linear coefficients may be too restrictive. For instance, this may occur if participants consistently use the response scale idiosyncratically. To account for this phenomenon, the authors partially relax the fixed coefficients assumption by allowing the intercepts in the factor model to change across participants. The model is attractive when m factors are expected on the basis of substantive theory but $m + 1$ factors are needed in practice to adequately reproduce the data. Also, this model for single-level data can be fitted with conventional software for structural equation modeling. The authors demonstrate the use of this model with an empirical data set on optimism in which they compare it with competing models such as the bifactor and the correlated trait–correlated method minus 1 models.

Keywords: multilevel modeling, reverse score items, method factor, difficulty factors

Supplemental material: <http://dx.doi.org/10.1037/1082-989X.11.4.344.supp>

One of the most widely used statistical models in the social sciences is without doubt the common factor model. In this model a set of observed variables is expressed as a linear function of a smaller set of common factors plus an error term. In addition, it is assumed that the relationship between each observed variable (e.g., item score) and the underlying common factors is linear and that the linear coefficients (intercepts and factor loadings) are invariant across respondents. In other words, the intercepts and factor

loadings do not change from respondent to respondent. In this article, we focus on the latter assumption.

Is the assumption that the coefficients are invariant across respondents reasonable in practice? This assumption has often been overlooked but has been well known for a long time. Wolfle (1940), in an extensive review of the factor analysis literature, discussed this assumption and suggested that until factor methods allow for the possibility of individual differences in coefficients, the coefficients should be interpreted as averages of individual coefficients that vary over individuals to an unknown degree. According to this assumption, the expected relationship between a respondent's common factor level and his or her response to a measured variable is the same for all respondents with a fixed level on the common factor. However, when the measured variables are items on a questionnaire, it is common to observe that different respondents use the response scale differently in a consistent fashion.

Consider three examples. The first is *differential thresholds*. Participants in a study assessing perceived pain are asked to rate their pain using a continuous scale on a Palm Pilot. It is likely that some participants are more sensitive to pain than others, in which case a 10 on the pain scale for one participant is not the same as a 10 for another participant. For a scale with items measuring perceived pain, it is reasonable to assume that respondents with a similar pain threshold will use the response scale in a similar way. However, respondents with widely different pain thresholds

Albert Maydeu-Olivares, Department of Psychology, University of Barcelona, Barcelona, Spain, and Marketing Department, Instituto de Empresa, Madrid, Spain; Donna L. Coffman, The Methodology Center and The Prevention Research Center, Pennsylvania State University.

This research was supported by the Department of Universities Research and Information Society of the Catalan Government and by Grants BSO2000-0661 and BSO2003-08507 from the Spanish Ministry of Science and Technology awarded to Albert Maydeu-Olivares. Donna L. Coffman was supported by National Institute on Drug Abuse Training Grant T32 DA017629-01A1 and Center Grant P50 DA10075. We thank David Gallardo-Pujol for creating the figures and Bethany C. Bray and Tammy L. Root for valuable feedback on a draft.

Correspondence concerning this article should be addressed to Albert Maydeu-Olivares, Faculty of Psychology, University of Barcelona, P. Valle de Hebrón, 171, 08035, Barcelona, Spain. E-mail: amaydeu@ub.edu

are likely to use the response scale differently. As another example, consider a questionnaire designed to assess perceived peer use of alcohol in which the response scale varies from *never drinks* to *drinks a lot*. Although all participants may be in general agreement as to the meaning of *never drinks*, *drinks a lot* is more ambiguous and is open to different interpretations by respondents.¹ In fact, participants may use the response scale differently anytime there are either no anchors for points on the response scale or the anchors are vague (Krosnick & Fabrigar, in press). The second example is *acquiescence* or *yea-saying* (Billiet & McClelland, 2000; Marsh, 1996). In personality questionnaires as well as in surveys, some respondents have a tendency to endorse all questions. As a result, answers may depend on the wording of the question for some but not for all individuals. The third example is *social desirability* (Paulhus, 2002). Some respondents are more likely to endorse the items that tap socially desirable topics. They present themselves in a manner that will be viewed favorably by others, whereas some respondents appear to be more resistant to this bias. These examples illustrate that the assumption of invariant coefficients is not always reasonable.

In this article, we propose a random intercept factor analysis model in which the intercepts, but not the factor loadings, are allowed to vary across individuals to accommodate these phenomena. The intercepts are treated as a random effect. Thus, individual intercepts are not directly estimated, but rather their variance in the population is estimated. Although common factor models with random components have been proposed before in the literature (see, e.g., Ansari, Jedidi, & Dube, 2002, and the references therein), such proposals have been within the context of multilevel data. In contrast, the model proposed here is for single-level data. Throughout the remainder of this article, the term *fixed coefficients* is used to refer to parameters that are the same for all individuals. In contrast, the term *random coefficients* is used to refer to parameters that are allowed to vary across individuals.

A random intercept factor analysis model may be called for when modeling individuals' observed responses to stimuli, such as their responses to the items of a questionnaire. Specially, m factors may be expected on the basis of theory but an additional factor is needed to adequately fit the data because of variability in the intercepts across participants. The additional factor is often thought of as a methodological artifact and thus spurious from a substantive viewpoint. Spurious factors have been discussed extensively in the factor analysis literature. One type of spurious factor that often appears in practice is related to the wording of items. For example, all the negatively worded items may load on one factor, and all the positively worded items may load on another factor, even though all items on the measurement instrument measure the same construct (Marsh, 1996). When this

occurs, one possibility is to model the data with two correlated factors, in which the positively worded items load on one factor and the negatively worded items load on the second factor. Another possibility is to use a bifactor model. In the bifactor model, a general factor loads on all items, and two additional factors are used to model the residual covariation among the items that is not captured by the general factor. One additional factor accounts for the residual covariation among the positive items, whereas the second additional factor accounts for the residual covariation among the negative items. A third possibility is to use a multimethod data approach. Thus, a model for multimethod data is used to fit the substantive traits being measured with two methods (positive and negatively worded items). An interesting model to fit in this context is Eid's (2000) correlated trait-correlated method minus one, CT-C(M-1), model because its model structure is a special case of the model structure implied by the bifactor model. A fourth possibility that we introduce in this article is to use a one factor model with a random intercept.

In the remainder of this article, we first review the common factor model with fixed coefficients. We then consider the bifactor model. Finally, we introduce our random intercept extension of the common factor model. We show that by introducing suitable assumptions, the random intercept factor model is equivalent to a factor model having an additional factor with constant factor loadings that is orthogonal to all "substantive" common factors. The model can be fit with standard programs for structural equation modeling. To illustrate our presentation, we apply a random intercept factor model to an empirical data set, in which we compare this model with the bifactor model, a model with more factors than would be expected from substantive theory, and a CT-C(M-1) model. We conclude with suggestions for applied researchers on how to choose among these models.

The Common Factor Model

Consider the responses of N individuals to a set of p items. We denote the response of participant j to item i using y_{ij} . Algebraically, the standard m -dimensional factor model for y_{ij} can be written as

$$y_{ij} = \mu_i + \boldsymbol{\lambda}'_i \boldsymbol{\eta}_j + e_{ij}. \quad (1)$$

In Equation 1, μ_i denotes the intercept for item i , $\boldsymbol{\lambda}_i$ denotes the vector of factor loadings for item i , $\boldsymbol{\eta}_j$ denotes participant j 's vector of common factors, and e_{ij} denotes an error term for participant j on item i . Notice that in this model the common factors, $\boldsymbol{\eta}$, and error terms, \mathbf{e} , include a subscript j .

¹ We are grateful to John Graham for this example.

Variables that include a subscript j are random; they change across participants. In contrast, the intercept μ_i and the factor loadings λ_i for each item do not include a subscript j . They are fixed coefficients that are invariant across participants. Figures 1 and 2 depict graphically a one-factor model (Model A) and a two-factor model (Model B), respectively.

We can write Equation 1 in a more compact matrix form using

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \mathbf{e}, \tag{2}$$

where \mathbf{y} , $\boldsymbol{\mu}$, and \mathbf{e} are $p \times 1$ vectors; $\boldsymbol{\Lambda}$ is a $p \times m$ matrix of factor loadings; and $\boldsymbol{\eta}$ is an $m \times 1$ vector of factor scores. The common factor model assumes the following in the population of respondents.

1. The mean of the common factors is zero.
2. The mean of the error terms is zero.
3. The error terms are uncorrelated with each other.
4. The error terms are uncorrelated with the common factors.

These assumptions, coupled with Equation 2, imply the following structure for the mean vector and covariance matrix of the observed variables:

$$\boldsymbol{\mu}_y = \boldsymbol{\mu} \quad \boldsymbol{\Sigma}_y = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}, \tag{3}$$

where $\boldsymbol{\Psi}$ denotes the $m \times m$ covariance matrix of the common factors, and $\boldsymbol{\Theta}$ denotes the $p \times p$ covariance

matrix of the error terms. $\boldsymbol{\Theta}$ is a diagonal matrix by Assumption 3.

In an m -factor model, two individuals with the same level on the latent factors have the same expected score on the items. However, they may not have the same observed score on the items because of the random error term, \mathbf{e} , in Equation 2. This error term includes both measurement error and individual differences on additional factors that are not explicitly modeled. Measurement error is random noise; individual differences on unmodeled additional factors may be systematic. Measurement error and systematic individual differences cannot typically be separated and are referred to as *unique factors* because they are unique to each item. Sometimes, however, some systematic individual differences are common to a small set of items, which can be captured by *specific factors*, yielding the *bifactor model*.

The Bifactor Model

Holzinger and Swineford (1937) proposed the bifactor model in the context of classical intelligence theory. In this model, each observed variable depends on two factors: a general intelligence factor and a smaller factor characterizing a specific subset of the items. The idea behind the bifactor model is to capture residual covariation among the observed measures that is not captured by the general factor by introducing several smaller factors, each of which is specific only to a small set of measures. General factors are assumed to be uncorrelated with these specific factors, but

Model A: One factor model

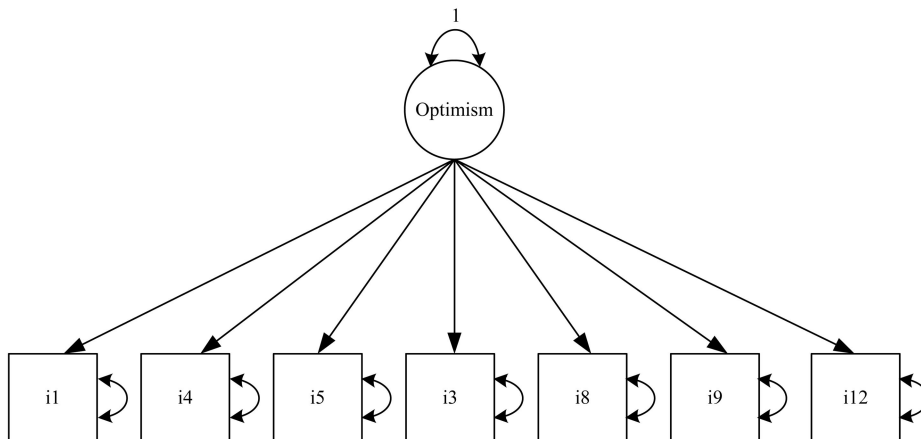


Figure 1. Model fitted to the Life Orientation Test (LOT) data (E. C. Chang et al., 1994). The items are numbered as in the original LOT questionnaire. Items 1, 4, and 5 are positively worded, whereas Items 3, 8, 9, and 12 are negatively worded. Variances are depicted with double-headed arrows from a variable to itself, factor loadings as single-headed arrows, and interfactor correlations as double-headed arrows between two latent variables. The variances of the factors are fixed to 1 for identification.

Model B: Restricted two factor model

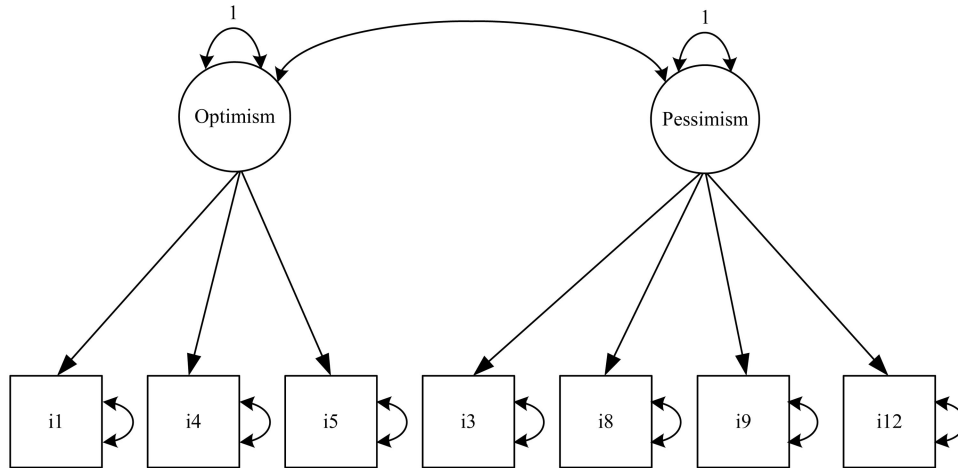


Figure 2. Model fitted to the Life Orientation Test (LOT) data (E. C. Chang et al., 1994). The items are numbered as in the original LOT questionnaire. Items 1, 4, and 5 are positively worded, whereas Items 3, 8, 9, and 12 are negatively worded. Variances are depicted with double-headed arrows from a variable to itself, factor loadings as single-headed arrows, and interfactor correlations as double-headed arrows between two latent variables. The variances of the factors are fixed to 1 for identification.

both the general and specific factors may be of substantive interest. A bifactor model with one general factor and two uncorrelated specific factors is depicted graphically in Figure 3 (Model C).

Unlike the bifactor model, which uses specific factors that may be of substantive interest to capture residual covariation unaccounted for by the common factors, the random intercept factor model incorporates one additional factor to account for individual differences, which is not motivated substantively. Rather, this additional dimension corresponds to the respondents' idiosyncratic usage of the response scale.

The Common Factor Model With a Random Intercept

Like the bifactor model, the random intercept factor model also attempts to model systematic individual differences not captured by the common factors, but it does so differently. In the random intercept factor model, we relax the assumption of an intercept common to all respondents by allowing the intercept to vary from respondent to respondent. This model is depicted in Figure 4 (Model D). Algebraically, instead of Equation 1, we write

$$y_{ij} = \gamma_{ij} + \boldsymbol{\lambda}'_i \boldsymbol{\eta}_j + e_{ij} \quad \gamma_{ij} = \mu_i + \zeta_j. \quad (4)$$

Note that the intercept γ_{ij} contains a subscript j . The intercept γ_{ij} is partitioned into a fixed part, μ_i , common to all respondents but that varies from item to item, and a random part, ζ_j , that varies from respondent to respondent but that is

common to all items. Thus, the term ζ_j captures individual differences in scale usage that are common across items.

In matrix form, we write Equation 4 as

$$\mathbf{y} = \boldsymbol{\gamma} + \boldsymbol{\Lambda} \boldsymbol{\eta} + \mathbf{e} \quad \boldsymbol{\gamma} = \boldsymbol{\mu} + \mathbf{1} \zeta, \quad (5)$$

where, as before, \mathbf{y} , $\boldsymbol{\mu}$, and \mathbf{e} are $p \times 1$ vectors; $\boldsymbol{\Lambda}$ is a $p \times m$ matrix of factor loadings; and $\boldsymbol{\eta}$ is an $m \times 1$ vector. In addition, $\boldsymbol{\gamma}$ and $\mathbf{1}$ are $p \times 1$ vectors, and ζ is a scalar.

In this model, we let the variance of ζ be denoted as φ . We do not estimate the intercept ζ_j for each respondent. Rather, we estimate the variance φ of the intercepts ζ_j in the population of respondents.

In addition to the four common factor model assumptions above, we assume the following in the population of respondents:

5. The mean of the random component of the intercept, ζ , is zero.
6. The random component of the intercept, ζ , is uncorrelated with the error terms.
7. The random component of the intercept, ζ , is uncorrelated with the common factors.

Assumptions 5 and 6 are imposed to identify the model. Assumption 7 implies that individual differences in the use of the response scale are not related to the participants' level on the common factors being measured.

Under the random intercept model, a respondent's expected response to an item given his or her level on the

Model C: Bifactor model

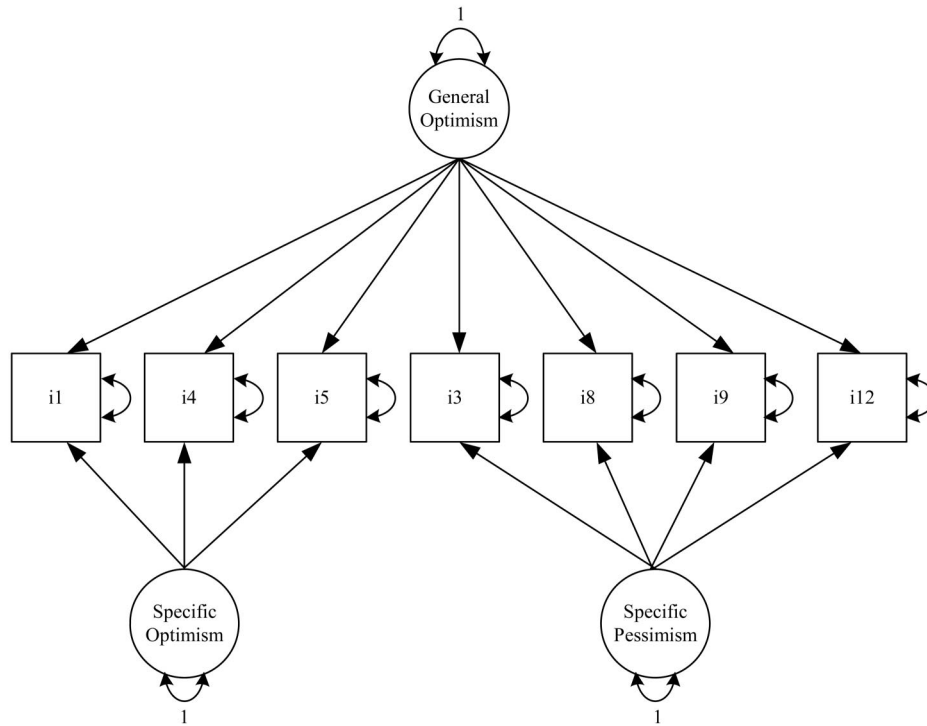


Figure 3. Model fitted to the Life Orientation Test (LOT) data (E. C. Chang et al., 1994). The items are numbered as in the original LOT questionnaire. Items 1, 4, and 5 are positively worded, whereas Items 3, 8, 9, and 12 are negatively worded. Variances are depicted with double-headed arrows from a variable to itself, factor loadings as single-headed arrows, and interfactor correlations as double-headed arrows between two latent variables. The variances of the factors are fixed to 1 for identification.

common factors and his or her level on the random component of the intercept ζ_j is

$$E(y_{ij}) = \mu_i + \lambda'_i \eta_j + \zeta_j. \tag{6}$$

Now, under the model, two respondents, A and B, with the same level on the common factors, η_A and η_B , need not have the same expected scores, $\mu_i + \lambda'_i \eta_j + \zeta_j$, on the item because in general their levels on the random components of the intercept will not be equal, $\zeta_A \neq \zeta_B$. In other words, even when Respondents A and B have the same level on the common factors, η_A and η_B , the expected scores may differ. Because ζ_j is a random variable that changes across respondents but is common to all items for a given respondent, we interpret ζ_j as the idiosyncratic use of the response scale by respondent j .

Equation 5, coupled with Assumptions 1–7, implies the following structure for the mean vector and covariance matrix of the observed variables (see the Appendix):

$$\mu_y = \mu \quad \Sigma_y = \mathbf{1}\phi\mathbf{1}' + \Lambda\Psi\Lambda' + \Theta, \tag{7}$$

where $\mathbf{1}$ is a $p \times 1$ vector of ones. This model is identified

by the usual rules for the identification of the factor model (see Bollen, 1989).

In the traditional m -factor model without a random intercept, the term ζ_j is embedded into the error term $e_{ij}^* = \zeta_j + e_{ij}$. This can be seen by reexpressing the two equations in Equation 4 as a single equation:

$$y_{ij} = \gamma_{ij} + \lambda'_i \eta_j + e_{ij} = \mu_i + \lambda'_i \eta_j + (\zeta_j + e_{ij}) = \mu_i + \lambda'_i \eta_j + e_{ij}^*. \tag{8}$$

In contrast to the traditional m -factor model, in the random intercept model a portion of the error term e_{ij}^* is extracted and modeled explicitly. Because ζ_j is a random variable with mean zero and variance ϕ , the inclusion of ζ_j in the model not only reduces the residual variance of the items but also accounts for a portion of the items' covariation. This is reflected in Equation 7, where the term $\mathbf{1}\phi\mathbf{1}'$ is added to the covariance structure of the common factor model (Equation 3). Also, a random intercept m -factor model will always fit as well as or better than a traditional m -factor model because of the estimation of one more parameter. The traditional m -factor model will only fit as well as a random

Model D: One factor model with a random intercept

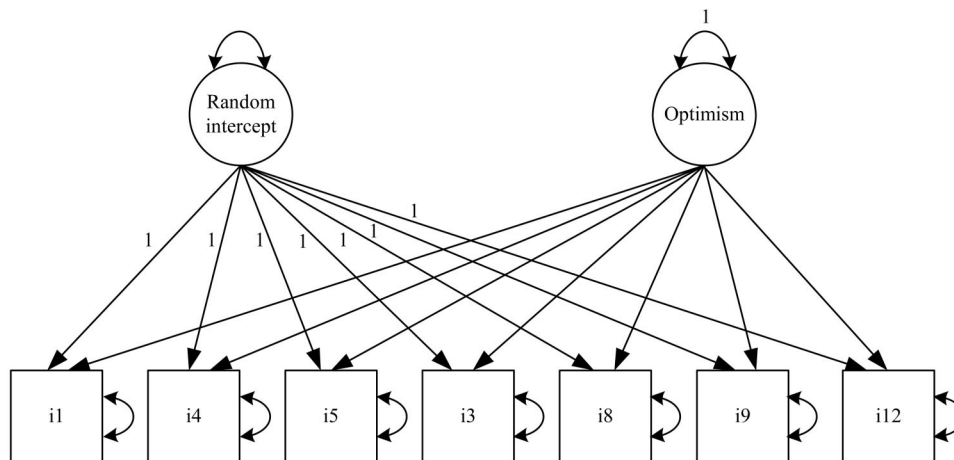


Figure 4. Model fitted to the Life Orientation Test (LOT) data (E. C. Chang et al., 1994). The items are numbered as in the original LOT questionnaire. Items 1, 4, and 5 are positively worded, whereas Items 3, 8, 9, and 12 are negatively worded. Variances are depicted with double-headed arrows from a variable to itself, factor loadings as single-headed arrows, and interfactor correlations as double-headed arrows between two latent variables. The variances of the factors are fixed to 1 for identification. This model is estimated as a model with two factors in which the factor loadings for one of the factors are all fixed to 1 and its variance (the variance of the random intercepts) is estimated.

intercept m -factor model when the variance of the random intercept, φ , is zero.

We show in the Appendix that a common factor model with m factors and a random intercept with Assumptions 1–7 cannot be empirically distinguished from a common factor model with $m + 1$ factors in which the first factor has constant factor loadings and is uncorrelated with all remaining m factors. Thus, if one finds the mean and covariance structure in Equation 7 to be a reasonable model for the observed data, the interpretation of the first factor as a random intercept or as a substantive factor must be made exclusively on substantive, not statistical, grounds. The random intercept factor model with Assumptions 1–7 implies Equation 7, but the converse is not true. This interpretation problem is not unique for this model. Rather, it holds for any mean and structural equation model (see Browne, 1982). In other words, the random intercept model incorporates the random intercept as a latent variable. It may be interpreted as a substantive factor, for example, reading ability. However, the substantive factor would have equal factor loadings for all measured variables, and the factor would be uncorrelated with all other factors in the model. Theory, rather than statistical results, determines the interpretation of the additional factor.

When no restrictions are imposed on $\boldsymbol{\mu}$, Equation 7 can be estimated by minimizing a discrepancy function (e.g., maximum likelihood) with conventional software for struc-

tural equation modeling. The term $\mathbf{1}\varphi\mathbf{1}'$ is similar to the term $\boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}'$; to estimate $\mathbf{1}\varphi\mathbf{1}'$, one simply declares an additional factor, which corresponds to the random intercept ζ . As depicted in Figure 4 (Model D), the factor loadings for this additional factor are set equal to 1 for all items, and the variance of this additional factor, φ , is left free to be estimated.

In closing this section, we point out that the random intercept factor model is not invariant to reverse coding of the items. This is important because, most often, questionnaires consist of positively worded and negatively worded items to avoid acquiescence effects. The negatively worded items are then reverse coded prior to analyzing the data. In the Appendix, we show how to fit a random intercept factor model when a subset of the items has been reverse coded prior to the analysis. Also, because the model is not scale invariant, it should be fitted to a covariance matrix. Fitting it to a correlation matrix results in incorrect test statistics, point estimates, and standard errors (see Cudeck, 1989).

Some Caveats Regarding the Application of the Random Intercept Model to Likert-Type Items

Responses to questionnaire items are often measured on a graded response scale, such as a Likert (1932) scale. Because Likert scores are actually ordered categories, in principle they should be treated as such either by an extension

of the common factor model using categorical variable methods (i.e., by estimating the model from polychoric correlations) or by a direct treatment of nonlinear functions suitable for categorical responses (i.e., by item response modeling), rather than by modeling them with a linear model such as the common factor model and its extensions, the bifactor model, the CT-C(M-1) model, and the random intercept model. In practice, however, when the number of ordered categories per item is five or more, only trivial differences in parameter estimates are generally observed when estimating a common factor model from product-moment versus polychoric correlations (Bentler & Chou, 1987; DiStefano, 2002; Dolan, 1994; Olsson, 1979). Also, the linear model can be a reasonable approximation to Likert-type items in practice, representing average behavior of the items over the range of the test (McDonald, 1999; see also Maydeu-Olivares, 2005), and the approximation may be preferred when the sample size is not large. Consequently, we report factor analyses using the product-moment correlations and sample standard deviations here (i.e., sample covariances). However, we wish to emphasize that prior simulation research has found differences between analyzing product-moment correlations (or covariances) and polychoric correlations that become increasingly substantial as the number of response options decreases below five.

Even with five or more ordered categories, the univariate and multivariate skewness and kurtosis of the variables should be thoroughly checked as skewed and/or kurtotic data can arise with Likert data. When Likert items show substantial skewness and/or kurtosis, the standard errors and goodness-of-fit tests obtained under normality assumptions will be incorrect. In this case, corrections to the standard errors and test statistic that take into account the nonnormality of the data are available (e.g., Satorra & Bentler, 1994) and have been implemented in software packages for structural equation modeling.

An Example: Modeling the Life Orientation Test (LOT)

The LOT (Scheier & Carver, 1985) is a 12-item questionnaire. Eight of the items are designed to measure general-

ized outcome expectancies; the remaining 4 items are filler items. One of the original items contained an idiomatic expression and has since been removed from the LOT (see Scheier, Carver, & Bridges, 1994). Hence, we did not include this item. The remaining 7 items are given in Table 1. The response scale for the items is graded, consisting of five points (0–4). Four of the items are positively worded, and the remaining 3 items are negatively worded.

The LOT was designed to measure a single construct. However, several factor analytic studies (e.g., E. C. Chang, D’Zurilla, & Maydeu-Olivares, 1994; L. Chang & McBride-Chang, 1996; Marshall, Wortman, Kusulas, Hervig, & Vickers, 1992; Robinson-Whelen, Kim, MacCallum, & Kiecolt-Glaser, 1997; Scheier & Carver, 1985) have revealed that a one-factor model does not fit well. Instead, a correlated two-factor model in which all positively worded items load on one factor and all negatively worded items load on another factor provides a substantially better fit to the data. However, we believe that it is hard to justify theoretically that optimism and pessimism are two distinct constructs. Obviously, one may be optimistic about the outcome of one situation and pessimistic about the outcome of another situation. But across situations (and the LOT measures generalized outcome expectancies), it is not clear how a respondent can simultaneously be both optimistic and pessimistic.

The data are from a study by E. C. Chang et al. (1994) and consist of responses to LOT items for 389 respondents. The item means, standard deviations, correlations, skewness, and kurtosis are presented in Table 2. Positive and negatively worded items correlate negatively in Table 2 because the data were not recoded prior to analysis. Here, we reanalyze E. C. Chang et al.’s (1994) data by fitting the following factor models: a one-factor model (Model A), a two-factor model (Model B), a bifactor model (Model C), a random intercept one-factor model (Model D), and a CT-C(M-1) model (Model E; Eid, 2000). These five models are displayed in Figures 1–5.

According to the one-factor model (Model A), a single construct underlies the responses to the LOT items. Because we have not recoded the negatively worded items prior to the analysis, positively and negatively worded items should load onto this factor with different signs. In contrast, in the

Table 1
Life Orientation Test (LOT) Items (E. C. Chang et al., 1994)

Item	Original item number
1. In uncertain times, I usually expect the best. (positive)	Item 1
2. I always look on the bright side of things. (positive)	Item 4
3. I’m always optimistic about my future. (positive)	Item 5
4. If something can go wrong for me, it will. (negative)	Item 3
5. I hardly ever expect things to go my way. (negative)	Item 8
6. Things never work out the way I want them to. (negative)	Item 9
7. I rarely count on good things happening to me. (negative)	Item 12

Note. The original item number is the order in which the item appears on the actual LOT questionnaire.

Table 2
Means, Standard Deviations, and Correlations for E. C. Chang et al.'s (1994) Life Orientation Test Data

Item	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7
Item 1	1.00						
Item 2	.51	1.00					
Item 3	.44	.53	1.00				
Item 4	-.16	-.22	-.26	1.00			
Item 5	-.28	-.38	-.33	.50	1.00		
Item 6	-.24	-.29	-.30	.51	.70	1.00	
Item 7	-.22	-.35	-.30	.44	.54	.52	1.00
<i>M</i>	2.24	2.40	2.56	1.85	1.39	1.32	1.40
<i>SD</i>	1.00	0.99	0.99	1.05	1.03	1.00	1.07
Skewness	-0.12	-0.35	-0.57	0.25	0.63	0.68	0.71
Kurtosis	-0.65	-0.36	-0.11	-0.72	-0.14	0.01	-0.23

Note. $N = 389$.

two-factor model (Model B), the positively worded items load onto one factor and the negatively worded items load onto another factor, with the factors being negatively correlated. For the bifactor model (Model C), a three-factor model is assumed. The factors are assumed to be uncorrelated. The general factor is assumed to underlie all items, representing overall optimism. The second factor underlies only the positively worded items, whereas the third factor underlies only the negatively worded items. These last two factors are to capture the specific covariation within positive and negatively worded items that is not captured by the overall optimism factor. The random intercept one-factor model (Model D) is similar to the one-factor model (Model A), except that a random component is added to its intercept to allow for consistent individual differences in the use of the response scale. The smaller the variance of this random component, the closer this model is to the one-factor model. In the limit, if this variance is zero, Model D reduces to Model A.

In a bifactor model, the general and the specific factors are substantive factors. However, in this application, the two specific factors underlie the positively and negatively worded items, respectively. Thus, it is more meaningful to consider them as method factors rather than as specific substantive factors. That is, the LOT data can be seen as an instance of multimethod data measuring one trait (optimism) using two methods (positive and negatively worded items). In this context, it is interesting to fit a CT-C(M-1) model to the LOT data (Model E) because this model structure is a special case of the model structure implied by the bifactor model. In the CT-C(M-1) model, (a) there are only $k - 1$ method factors, where k denotes the number of methods considered, and (b) the trait and method factors are uncorrelated. One method is chosen as a comparison standard for all other methods. Here we use the positively worded items as a comparison standard to assess the effect of including negatively worded items. Thus, the method

factor corresponding to the negatively worded items measures deviations from the values predicted by positively worded items.

In all models except for Model D, the intercepts are fixed parameters and their estimates are the item means (see Table 2). Thus, for Model A (the one-factor model), individuals' predicted scores on an item result from a score on the common factor, optimism, in addition to the intercept value for that item (see Equation 1). For example, for all individuals with a score of zero on the latent factor optimism, the predicted score on Item 1 is 2.24 (mean on Item 1), and the predicted score on Item 2 is 2.40 (mean on Item 2). In contrast, for the random intercept one-factor model (Model D), the intercept incorporates two components: one that is constant over individuals but varies from item to item and another component that is constant over items but varies from person to person. According to Model D, an individual's predicted score on a given item is a result of a score on the latent factor optimism, an intercept value for that item, and an intercept value specific to that individual. Thus, the predicted scores on a given item will vary over individuals even though they may have the same level on the latent optimism factor.

All models were estimated in LISREL (Jöreskog & Sörbom, 2001) with the maximum likelihood fit function and normal theory standard errors. Model A was identified by fixing to 1.0 the variance of the factor (see the path diagram given in Figure 1). Model B was identified by fixing to 1.0 the variances of both of the factors (see Figure 2). Model C was identified by fixing the variances of all the factors to 1.0 (see Figure 3). Model D was identified by fixing to 1.0 the variance of the substantive factor (see Figure 4). In addition, all of the factor loadings for the random intercept factor are fixed to 1.0 as shown in the figures. Model E was identified by fixing the variances of the trait factor and the method factor to 1.0 (see Figure 5). All models were fitted to the covariance matrix. The input files used in the example are

Model E: CT-C(M-1) model

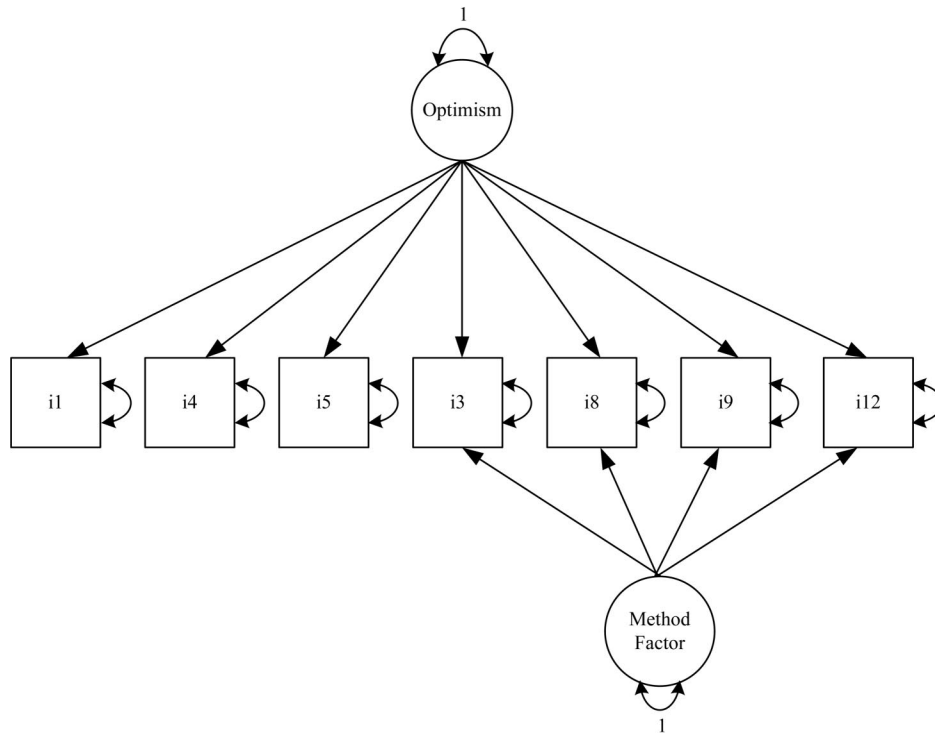


Figure 5. Model fitted to the Life Orientation Test (LOT) data (E. C. Chang et al., 1994). The items are numbered as in the original LOT questionnaire. Items 1, 4, and 5 are positively worded, whereas Items 3, 8, 9, and 12 are negatively worded. Variances are depicted with double-headed arrows from a variable to itself, factor loadings as single-headed arrows, and interfactor correlations as double-headed arrows between two latent variables. The variances of the factors are fixed to 1 for identification. The correlated traits–correlated methods minus one, CT-C(M-1), model is estimated as a model with one substantive factor and one method factor (negatively worded items). The other method factor, positively worded items, is chosen as reference method. The CT-C(M-1) model is identified by fixing the variances of both the substantive factor and the method factor to 1.

available in the supplementary data in the online version of this article.

Results

The comparative fit index (CFI; Bentler, 1990), Akaike’s information criterion (AIC; Akaike, 1987), consistent AIC

(CAIC; Bozdogan, 1987), and chi-square goodness-of-fit tests are given in Table 3. CFI values above .95 are generally considered to indicate good fit (Hu & Bentler, 1999). Both the AIC and the CAIC are information criteria indices used in model comparison. These values may be used regardless of whether the models are nested. The model with

Table 3
Goodness-of-Fit Tests for the Life Orientation Test Data

Model	χ^2	df	p	AIC	CAIC	CFI
A: One-factor model	185.78	14	< .001	243.30	312.79	.87
B: Two-factor model	16.99	13	.20	46.89	121.35	1.00
C: Bifactor model	7.82	7	.35	49.68	153.91	1.00
D: Random intercept one-factor model	19.63	13	.10	50.24	124.70	.99
E: CT-C(M-1) model	10.02	10	.43	46.11	135.46	1.00

Note. N = 389. AIC = Akaike’s information criterion; CAIC = consistent AIC; CFI = comparative fit index; CT-C(M-1) = correlated traits–correlated methods minus one.

the smallest AIC or CAIC value is considered the model that gives the best fit to the data while also taking into account model complexity.

As is shown in Table 3, the one-factor model fits very poorly according to both the chi-square test and the CFI. Also, both the AIC and the CAIC indicate that the one-factor model fits the worst of the five models. The chi-square test and the CFI indicate that the two-factor model, the random intercept model, the bifactor model, and the CT-C(M-1) model fit the data well. The CFI and AIC are not very helpful in choosing among these four models. The CAIC values for the bifactor model and CT-C(M-1) model are larger than for the other two models. The CAIC does not distinguish between the random intercept model and the two-factor model.

In Table 4, we provide the factor loadings for all the models considered. We see in Table 4 that the pessimism specific factor in the bifactor model nearly collapses, as there is only one significant factor loading in this factor. Also, we see that the factor loadings for the optimism specific factor are always larger than the factor loadings for the optimism items in the overall optimism factor. These two facts cast serious doubts on the appropriateness of the bifactor model in this application. The bifactor model appears to be overparameterized in this instance.

One possibility to overcome this overparameterization is to include only one specific factor. This is precisely what the CT-C(M-1) model does in this context. The specific factors are treated as method factors, and one method is chosen as standard. The factor loadings obtained when the positively worded items are used as reference method are shown in Table 4. We see in this table that for the negatively worded items, the trait factor loadings are all smaller than the method factor loadings. This implies that for the negatively worded items, the percentage of variance accounted for by the method is larger than the percentage of variance accounted for by the optimism trait.

Another feature of this model that should give researchers pause when using it is that the model fit changes when a different method is used as the reference (Eid, 2000). In this application, it is possible to use the negatively worded items as an alternative reference method. The alternative CT-C(M-1) model also yields a good fit to the LOT data, but not quite as good in terms of the chi-square test as the CT-C(M-1) model when positive items are used as reference method: $\chi^2(11, N = 389) = 15.40, p = .17, AIC = 49.32, CAIC = 133.70, CFI = 1$.

For the random intercept one-factor model, Table 4 shows that all the loadings are large. Also, the estimate of the variance of the random component of the intercept is 0.13, a rather small value relative to the variance of the substantive factor, which was set to 1.0. However, the value of the variance of the intercept is rather large relative to its standard error ($SE = 0.01, z = 8.87, p < .001$), and setting this

variance equal to zero, which is equivalent to specifying a one-factor model, results in a very poor fitting model.

We can compute the percentage of the items' variance accounted for by the random intercept and the factor. Under an m -factor model with a random intercept, the variance of an observed variable is $\text{var}(y_i) = \varphi + \lambda_i' \Psi \lambda_i + \Theta_{ii}$, where φ is the variance of the random component of the intercept, λ_i' is the $m \times 1$ vector of factor loadings for item i , Ψ is the matrix of interfactor correlations, and Θ_{ii} is the unique variance of the item. The percentage of variance accounted for by the random intercept is $\varphi/\text{var}(y_i)$, whereas the percentage of variance accounted for by the factors is $(\lambda_i' \Psi \lambda_i) / \text{var}(y_i)$. The estimated unique variances for these items are {0.58, 0.41, 0.50, 0.67, 0.31, 0.35, 0.64}. Consequently, with the factor loading estimates reported in Table 4, the percentage of the item's variance accounted for by the random intercept ranges from 10% to 13%. On the other hand, the percentage of the item's variance accounted for by the optimism factor ranges from 29% to 58%. Thus, the random intercept accounts for nonnegligible percentage of the variance of the items, but the substantive factor accounts for about three times the percentage of variance relative to the random intercept.

It is clear that the two-factor model, the random intercept one-factor model, the bifactor model, and the CT-C(M-1) model outperform the one-factor model in terms of goodness of fit. Of the four, the bifactor model involves the use of more parameters than the other three models to attain a similar fit. Thus, it is less parsimonious than its competitors: Its CAIC index is larger than its competitors'. From a substantive viewpoint, an unattractive feature of the bifactor model is that the specific factor pessimism nearly collapses in this application. The CT-C(M-1) model yields the largest p value and smallest AIC index but the second largest CAIC index as it is not as parsimonious as the two-factor model and the random intercept one-factor model. An unattractive feature of the CT-C(M-1) model in this application is that a larger percentage of variance of the negatively worded items is accounted for by the method factor than by the substantive factor. Finally, the two-factor model provides a slightly better fit than the random intercept one-factor model. Yet, taken together, the goodness-of-fit results do not shed much light on which model to choose. Given the similarity of goodness of fit for Models B and D, it appears that choosing between these models needs to be made on substantive rather than on statistical grounds. To shed additional light on the substantive implications of choosing one model over another, we turn to relating the factors estimated in each model to exogenous variables.

Relation to Exogenous Variables

For each model, we computed factor scores via the regression method (for details see Jöreskog & Sörbom, 2001, p. 134). The factor scores were correlated with three exter-

Table 4
Factor Loadings for the Life Orientation Test Data

Item	One-factor model:		Two-factor model			Bifactor model			One-factor random intercept:		CT-C(M - 1) model	
	Optimism		Optimism	Pessimism		Overall optimism	Specific optimism	Specific pessimism	Optimism		Optimism	Method factor
Item 1	0.38 (0.05)		0.64 (0.05)	0		0.35 (0.07)	0.56 (0.07)	0	0.54 (0.05)		0.64 (0.05)	0
Item 2	0.48 (0.05)		0.78 (0.05)	0		0.49 (0.08)	0.61 (0.07)	0	0.66 (0.05)		0.79 (0.05)	0
Item 3	0.46 (0.05)		0.68 (0.05)	0		0.44 (0.07)	0.51 (0.07)	0	0.61 (0.05)		0.68 (0.05)	0
Item 4	-0.64 (0.05)		0	0.65 (0.05)		-0.59 (0.09)	0	0.26 ^a (0.18)	-0.56 (0.05)		-0.32 (0.06)	0.57 (0.05)
Item 5	-0.86 (0.05)		0	0.87 (0.05)		-0.76 (0.10)	0	0.38 (0.23)	-0.78 (0.05)		-0.48 (0.06)	0.72 (0.05)
Item 6	-0.79 (0.05)		0	0.82 (0.05)		-0.63 (0.11)	0	0.64 ^a (0.16)	-0.71 (0.05)		-0.39 (0.06)	0.74 (0.05)
Item 7	-0.70 (0.05)		0	0.70 (0.05)		-0.73 (0.08)	0	0.15 ^a (0.18)	-0.65 (0.05)		-0.44 (0.06)	0.53 (0.05)

Note. $N = 389$. Standard errors for estimated parameters are shown in parentheses. The correlation between the factors in the two-factor model is $-.53$ ($SE = .05$). The variance of the random intercept is 0.13 ($SE = 0.01$, $z = 8.87$, $p < .001$). CT-C(M - 1) = correlated traits-correlated methods minus one.

^a Factor loadings are not significantly different from zero ($p > .05$).

nal criteria: Beck's Hopelessness Scale (BHS; Beck, Weissman, Lester, & Trexler, 1974) and the Avoidance scale (AS) and the Negative Problem Orientation (NPO) scale from the Social Problem Solving Inventory—Revised (D'Zurilla, Nezu, & Maydeu-Olivares, 2002). The BHS is a questionnaire used to predict depression, suicidal ideation, and suicidal intent. The NPO scale measures the general tendency to view problems occurring in living as a significant threat to well-being, to doubt one's personal ability to solve problems successfully, and to easily become frustrated and upset when confronted with problems. Finally, the AS measures a style used to confront problems in living characterized by procrastination, passivity or inaction, and dependency. It is expected that pessimistic individuals are more likely to engage in hopeless thinking and avoidant behaviors and in general to be negatively oriented toward solving problems in everyday living. Optimism is expected to correlate strongly and negatively with these three criteria. Table 5 shows the correlations between factor scores for the different models fitted to the LOT data and these criteria variables.

In their original work on the LOT, Scheier and Carver (1985) treated the set of items as unidimensional. Indeed, they advocated the use of a sum score after reversing the score for the negatively worded items to investigate the relationship between optimism and exogenous criteria. Yet, their factor analyses did not fully support the use of a one-dimensional model. They found that the two-factor model significantly outperformed the fit of the one-factor model. In Table 5 we also provide the correlations between the sum score obtained after reversing the score for the negatively worded items and the criteria.

The one-factor model with a random intercept provides

support to Scheier and Carver's (1985) heuristic. As with other random effects models, such as latent growth curve models or multilevel models, the specific values (i.e. factor scores) of the random component, ζ , for each individual are not estimated when fitting the model. Rather, the variance of the intercept factor, φ , is estimated. To obtain the specific values for each individual, estimates would need to be obtained after the model is fit to the data. We see in Table 5 that the correlations between the factor scores for the optimism factor and the criteria are almost identical to those between the sum score and the criteria. In addition, the correlations between the estimates for the random intercept and the criteria are nonsignificant or, if significant, of negligible magnitude. The random intercept can be interpreted as a method factor negligibly related to external criteria. Thus, because (a) the one-factor model with a random intercept fits the data and (b) the optimism factor scores and the sum score are highly correlated (.991 in this application), researchers are justified to use the sum score as a proxy for the optimism factor to investigate relations between optimism and external criteria. Researchers may also wish to investigate relations between the random intercept estimates and external criteria, but as this application shows, these are likely to be negligible.

In contrast, the interpretation of the relationships between factors and external criteria for Models B, C, and E present substantive questions that are not so easy to answer. Thus, for the two-factor model, two factor scores are computed, measuring optimism and pessimism, respectively. As expected, the optimism factor score correlates strongly and negatively with all three criteria, and the pessimism factor score correlates strongly and positively with all three criteria. Yet, what does it mean substantively that optimism and

Table 5
Correlations Between Factor Scores and Selected Criteria

Model and factor score	BHS	AS	NPO
Two-factor model			
Optimism	-.53**	-.32**	-.44**
Pessimism	.67**	.42**	.59**
Bifactor model			
Specific optimism	-.29**	-.16**	-.22**
Specific pessimism	.35**	.17**	.29**
Overall optimism	-.70**	-.45**	-.62**
Random intercept			
Optimism	-.70**	-.43**	-.61**
Random intercept	.13*	.09	.14*
CT-C(M-1) model			
Optimism	-.51**	-.31**	-.43**
Method factor for negatively worded items	.48**	.30**	.44**
Sum score	-.71**	-.44**	-.61**

Note. $N = 311$. BHS = Beck's Hopelessness Inventory; AS = Avoidance scale of the Social Problem Solving Inventory—Revised; NPO = Negative Problem Orientation scale of the Social Problem Solving Inventory—Revised; CT-C(M-1) = correlated traits-correlated methods minus one.

* $p < .05$. ** $p < .01$.

pessimism are distinct factors (albeit related)? How can a respondent be optimistic and pessimistic across situations? Turning to the bifactor model, the sum score correlates .963 with the general factor in this model. Thus, in Table 6, the factor score for overall optimism behaves very much like the sum score. However, the factor scores for the specific factors also correlate significantly (although more weakly) and in the expected directions with the criteria. Finally, the correlations between factor scores and external criteria for the CT-C(M-1) model (using positively worded items as reference method) present clear interpretation problems. Correlations between the trait factor (optimism) and external criteria are substantially weaker than between the sum score and external variables. In fact, they are of the same magnitude (but opposite directions) as the correlations between the method factor and external variables. Also, the CT-C(M-1) model can be specified with the negatively worded items as reference method. In this case, the factor score correlations with external criteria are similar to those found for the sum score. It correlates .67, .42, and .60 with the BHS, AS, and NPO scale, respectively. Also, the method factor (positively worded items) scores correlate 0.20, $-.10$, and $-.13$ with the external criteria. It is known from Eid's (2000) original work that the two CT-C(M-1) models do not produce the same fit indices. However, even more important, we have shown here that the empirical results with external criteria are highly discrepant. These findings should give researchers pause about using this approach.

In summary, we have shown that by introducing a random component in the intercept of a one-factor model we do not reject Scheier and Carver's (1985) original hypothesis that optimism is indeed a one-dimensional construct. We find our solution more parsimonious from a substantive viewpoint than hypothesizing a two-factor factor model (e.g. E. C. Chang et al., 1994), a one-factor model with correlated errors (e.g., Scheier & Carver, 1985), or a bifactor model. Also, it is free from the interpretation problem of the CT-C(M-1) model.

Did We Have Power to Distinguish Between the Alternative Models?

Because the alternative models we considered in modeling the LOT data are hard to distinguish statistically in the data set we considered, a reviewer raised the issue of whether the models could be distinguished in other data sets. To address this very important question, we performed a small simulation study to investigate the power of the chi-square test statistic to distinguish among the following four alternative models used in the LOT application: (a) a one-factor model; (b) a two-factor model in which the positively worded items load on one factor, the negatively worded items load on another factor, and both factors are correlated; (c) a bifactor model; and (d) a random intercept one-factor model. Multivariate normal data were generated according to Models B, C, or D and estimated using all four models. Sample sizes were in all cases $N = 200$, $N = 400$, $N = 600$, and $N = 1,000$. Estimation was performed using maximum likelihood. Also, we investigated models with similar characteristics (i.e., number of items, parameter values) to those found in the LOT data.

Two-Factor Model as Data Generation (True) Model

To generate data according to this model, we used the following matrix of factor loadings:

$$\Lambda' = \begin{pmatrix} .55 & .65 & .75 & .85 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .55 & .65 & .75 & .85 \end{pmatrix}. \quad (9)$$

The factors were correlated, and factor variances were set equal to one. The unique variances were set to 0.5 for all items. In Equation 9, the factors mimic the optimism and pessimism factors. We generated data to investigate different values for the correlations among the factors ($\rho_s = -.2$, $-.4$, $-.6$, and $-.8$).

Models A–D were estimated for each data set. All factor variances were set equal to 1.0 to identify the models. All

Table 6
Empirical Rejection Rates at $\alpha = .05$ for Alternative Models Fitted to Data Generated According to a Two-Factor Model

Model used for estimation	N	$\rho = -.2$	$\rho = -.4$	$\rho = -.6$	$\rho = -.8$
Two-factor model	200	.06	.06	.06	.06
Bifactor model	1,000	.02 (959)	.03 (979)	.04 (975)	.03 (962)
Random intercept one-factor model	200	.46	.30	.18	.10
Random intercept one-factor model	400	.81	.60	.34	.12
Random intercept one-factor model	600	.96	.82	.52	.16
One-factor model	200	1	1	1	.95

Note. 1,000 replications were used in each condition. All replications converged except when estimating a bifactor model. In this case, the number of converged replications is given in parentheses. When the model used for estimation is a two-factor model, rejection rates should be close to .05. When an alternative model is used for estimation, high rates indicate that the true and estimated models are easily distinguishable, whereas rates close to .05 indicate that the true and estimated models are indistinguishable.

factor loadings and uniquenesses were freely estimated. Empirical rejection rates at $\alpha = .05$ are summarized in Table 6.

First, we investigated whether the empirical rejection rates of the chi-square test are correct when the model used for estimation is the true model. In this case, the percentage of times the chi-square test incorrectly rejects the model should be close to the nominal rate (i.e., close to 5%). Table 6 shows that in this case, the test statistic behaves correctly. Rejection rates are close to nominal rates even with small samples ($N = 200$).

Next, we investigate the percentage of times the test statistic correctly rejects the model when the model used in the estimation is not the model used to generate the data. This is the empirical power to reject a false model. High rejection rates indicate that the true and estimated models are easily distinguishable, whereas rates close to .05 indicate that the true and estimated models are indistinguishable. Table 6 shows that a small sample ($N = 200$) suffices to easily distinguish the two-factor model from the one-factor model. Empirical power is 95% with this sample size even when the correlation between the factors is $-.8$. However, we do not have any power to distinguish a two-factor model from a bifactor model. Power ranges from 2% to 4% even when the sample size is 1,000 observations. Also, a two-factor model is not easily distinguishable from a random intercept factor model when the correlation between the factors is $-.6$ or smaller. When the correlation is $-.6$, we reject the random intercept model more than 50% of the time only when the sample size is 600 observations or higher; when $N = 200$, we would reject the model only 18% of the time, and when $N = 400$, 34% of the time.

Bifactor Model as Data Generation (True) Model

To generate data according to a bifactor model, we used the following matrix of factor loadings:

$$\Lambda' = \begin{pmatrix} .55 & .65 & .75 & .85 & -.55 & -.65 & -.75 & -.85 \\ \lambda & \lambda & \lambda & \lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda & \lambda & \lambda & \lambda \end{pmatrix}. \quad (10)$$

The factors were uncorrelated, and factor variances were set equal to one. The unique variances were set to 0.5 for all items. In Equation 10, the first factor mimics the overall optimism factor, and the other two factors mimic the specific optimism and specific pessimism factors. For simplicity, we used a common factor loading for the specific factors to generate the data. Also, we investigated different values for the loadings of the specific factors ($\lambda_s = 0.15, 0.20, 0.25, \text{ and } 0.30$).

Models A–D were estimated for each data set. All factor variances were set equal to 1.0 to identify the models. All factor loadings and uniquenesses were freely estimated. Empirical rejection rates at $\alpha = .05$ when data were estimated using the different alternative Models A–D are summarized in Table 7.

The test statistic behaves as it should when the model used for estimation is the same used to generate the data. However, note that the estimation does not always converge even when 1,000 observations are used. In fact, when only 200 observations are used, the bifactor model converges in one of the conditions only 54% of the time. In contrast, the alternative models always converge even though they are incorrect models in this case.

However, the statistic has no power whatsoever to distinguish between the bifactor model and the random intercept model. Rejection rates are only 7% even when $N = 1,000$. Also, the chi-square statistic has extremely low power to distinguish between the bifactor model and a two-factor model. Rejection rates range from 8% to 22% when $N = 1,000$. Finally, the chi-square difference test has some power to distinguish between a one-factor model and the bifactor model provided the loadings of the specific factor are not too small. It will correctly reject the one-factor

Table 7
Empirical Rejection Rates at $\alpha = .05$ for Alternative Models Fitted to Data Generated According to a Bifactor Model

Model used for estimation	N	$\lambda = 0.15$	$\lambda = 0.20$	$\lambda = 0.25$	$\lambda = 0.30$
Bifactor model	200	.03 (538)	.03 (594)	.04 (656)	.03 (649)
Bifactor model	1,000	.05 (930)	.03 (964)	.03 (930)	.04 (960)
Random intercept one-factor model	1,000	.07	.07	.07	.07
Two-factor model	1,000	.08	.10	.12	.22
One-factor model	200	.11	.30	.65	.92
One-factor model	400	.20	.60	.93	1
One-factor model	1,000	.52	.97	1	1

Note. 1,000 replications were used in each condition. All replications converged except when estimating a bifactor model. In this case, the number of converged replications is given in parentheses. When the model used for estimation is the bifactor model, rejection rates should be close to .05. When an alternative model is used for estimation, high rates indicate that the true and estimated models are easily distinguishable, whereas rates close to .05 indicate that the true and estimated models are indistinguishable.

model more than 50% of the time when $\lambda = 0.25$ if $N = 200$ (rejection rate 65%) and when $\lambda = 0.20$ if $N = 400$ (rejection rate 60%). The power decreases quickly as the magnitudes of the loadings of the specific factors decrease.

Random Intercept One-Factor Model as Data Generation (True) Model

To generate data according to the random intercept one-factor model, we used the following matrix of factor loadings:

$$\lambda' = (.55 \ .65 \ .75 \ .85 \ - \ .55 \ - \ .65 \ - \ .75 \ - \ .85). \quad (11)$$

The factor variance was set equal to one, and the unique variances were set to 0.5 for all items. This factor mimics the optimism factor. We added a random intercept to the model that was uncorrelated with the substantive factor. We investigated different values for the variance of the random intercept ($\varphi_s = 0.05, 0.10, 0.15,$ and 0.20).

The data generated in this way were estimated with Models A–D described above. All factor variances were set equal to 1.0 to identify the models. All factor loadings and uniquenesses were freely estimated. Empirical rejection rates at $\alpha = .05$ are summarized in Table 8.

Again, when we used the true model for estimation, the test statistic behaved as it should: Rejection rates were close to nominal rates even in small samples ($N = 200$). Again, we have almost no power to distinguish the true model from a bifactor model. Rejection rates for the bifactor model are close to rejection rates for the random intercept model even with a sample size of 1,000 observations. On the other hand, we can easily distinguish a random intercept one-factor model from a one-factor model. Even with a sample size of 200 observations and even when the variance of the random intercept is very small, a one-factor model is rejected 93% of the time. Finally, we have some power to distinguish a random intercept one-factor model and a two-factor model when the variance of the intercept is not too small ($\varphi >$

0.15) and sample size is large ($N > 400$). Below these conditions, power is less than 50%.

Summary of the Simulation Results

The bifactor model is a troublesome model. First, it may not converge. Nonconvergence is more likely when a small sample is used and the factor loadings of the specific factors are equal in the population. Second, a bifactor model is likely to provide a good fit in applications even if it is an incorrect model. The power to reject this model when the true model is a random intercept one-factor model or a two-factor model is extremely low even in large samples. As a result, a researcher who finds a bifactor model to reproduce the data adequately cannot be sure it is the correct model. The bifactor model lacks parsimony and fits a variety of data structures.

The random intercept one-factor model and a two-factor model can only safely be distinguished when sample sizes are 400 observations or greater, provided that the true model is not too extreme (small φ or large ρ in absolute value). Even with samples of this size, a researcher may be confused as to which of the two models is the correct one if (a) the variance φ of the random intercept is smaller than 0.15 or (b) the correlation ρ among the factors is larger than .5 in absolute value. In our LOT example, sample size was roughly 400, the estimated variance of the random intercept was 0.13, and the correlation among the factors was $-.53$. Under these conditions, the power to reject the incorrect model is roughly 50% according to the results of our small simulation study.

Discussion

Researchers often use factor analysis to model the responses of individuals to a set of questionnaire items for which some graded response scale is provided. Provided that the number of categories per item is five or more, the

Table 8
Empirical Rejection Rates at $\alpha = .05$ for Alternative Models Fitted to Data Generated According to a Random Intercept One-Factor Model

Model used for estimation	<i>N</i>	$\varphi = 0.05$	$\varphi = 0.10$	$\varphi = 0.15$	$\varphi = 0.20$
Random intercept one-factor model	200	.07	.07	.07	.07
Bifactor model	1,000	.04 (954)	.05 (965)	.10 (971)	.16 (977)
Two-factor model	200	.09	.17	.27	.38
Two-factor model	400	.13	.33	.54	.71
Two-factor model	600	.18	.48	.77	.91
Two-factor model	1,000	.29	.75	.97	1
One-factor model	200	.93	1	1	1

Note. 1,000 replications were used in each condition. All replications converged except when estimating a bifactor model. In this case, the number of converged replications is given in parentheses. When the model used for estimation is the random intercept one-factor model, rejection rates should be close to .05. When an alternative model is used for estimation, high rates indicate that the true and estimated models are easily distinguishable, whereas rates close to .05 indicate that the true and estimated models are indistinguishable.

common factor model may be a reasonable model in these situations. Yet, in applications researchers sometimes find that when m factors were expected on the basis of substantive theory, $m + 1$ factors are needed in practice to adequately reproduce the data. We have suggested that applied researchers may wish to consider in this situation fitting an m -factor model with a random intercept. Indeed, we have shown that if one is willing to assume that respondents use the response scale of questionnaire items idiosyncratically, a random component needs to be incorporated into the intercept of the model, which results in an additional factor when fitting the data.

Here, we have illustrated the usefulness of this new model by fitting a one-factor model with a random intercept to the LOT, a well-known personality questionnaire for which a unidimensional model was proposed substantively but for which statistically a two-factor model appeared to be needed. In our application, we compared the one-factor random intercept model with a one-factor model, a two-factor model, a bifactor model, and a CT-C(M-1) model. Only the one-factor model could be statistically rejected in this application. The substantive interpretation of and difference between each of these models should be taken into account when choosing among these models in situations in which goodness-of-fit results and model selection criteria are similar.

Compared with the two-factor model, the one-factor model with a random intercept enables researchers to maintain their original substantive hypothesis, namely, that optimism is a unidimensional construct. Also, when the model fits, we have shown that factor score estimates can be highly correlated with summative scores, so that the latter can be used as proxies for the latent factors, just as in factor models without a random intercept. Furthermore, when positive and negatively worded items are simultaneously modeled, the random intercept model may group items worded differently into the same dimension. In contrast, the two factor model groups the items according to their wording: Positively worded items load on one dimension, and negatively worded items load on the other dimension. The two-factor model should be preferred to the one-factor random intercept model only when the additional factor uncovered by the analysis provides additional insight in the phenomenon under investigation, as when there are distinct patterns of relations with exogenous variables (discriminant validity). This is more likely to occur as the correlation between the two factors decreases in magnitude toward zero.

Compared with the bifactor model, the one-factor model with a random intercept is a more parsimonious model. Being a highly parameterized model, the bifactor model may be empirically underidentified, leading to nonconvergent or improper solutions. Also, researchers should be cautious of the bifactor model as it may yield a good fit to the data even when it is an incorrect model. In contrast, the random intercept model is free from these problems. The

bifactor model should be preferred to the random intercept model when researchers are interested in domain specific factors over and above the general factor and, particularly, when researchers are interested in their differential predictive validity (Chen, West, & Sousa, 2006). Another problem frequently encountered in the bifactor model is factor collapse. This occurs when all the factor loading estimates for one or more of the domain specific factors are nonsignificant. A similar problem may be encountered in the random intercept model, namely, all the factor loadings for one or more substantive factors may be nonsignificant. These problems are likely to be caused by model misspecification. Therefore, researchers should be cautious of factor collapse in either model.

Finally, a limitation of the CT-C(M-1) model is that for any given data in which k methods are used, there are $k - 1$ possible CT-C(M-1) models. Each of these models is based on a different reference method, yields a different fit to the data at hand (Eid, 2000), and may lead to a different substantive interpretation. Also, researchers should be cautious about using a CT-C(M-1) model when the method factor accounts for a larger percentage of variance of the observed variables than the trait factors.

In closing, when fitting the random intercept model, we recommend that researchers specify alternative models, examine their goodness of fit to the data, use sequential chi-square difference tests (Bentler & Bonett, 1980) for nested models, examine parameter estimates, and base conclusions on theoretical predictions (as we would recommend for any modeling procedure). In particular, when fitting the random intercept model, researchers should check whether (a) the variance of the random intercept is smaller than the variance of the substantive factors and (b) the loadings for the substantive factors are significant.

Conclusions

We believe that the model proposed here provides applied researchers with a valuable tool to model questionnaire data in which it is likely that there are method effects due to consistent individual differences in the use of the response scale. Although the example presented here had only one substantive factor, this model may be used with any number of substantive factors.

A clear limitation of the random intercept model is the assumption (needed to identify the model) that the random intercept and the substantive factors be uncorrelated. Consider the pain tolerance example presented at the beginning of this article. Imagine college students and construction workers are receiving shocks through their fingertips. The level on the pain latent variable is very likely associated with the intercept because the two groups are likely to differ in the calluses on their fingers, which affect their general tolerance to pain.

The common factor analysis model also makes the as-

sumption that the factor loadings are fixed and do not vary across individuals (Wolfe, 1940). The factor loadings are still invariant across respondents in the model proposed here. Although, as Wolfe (1940) argued, it may not be reasonable to assume that the factor loadings are invariant across individuals, we do not believe that such an extension is feasible with the traditional (frequentist) framework used here. However, factor models with random loadings may be feasible within a Bayesian framework.

Although we have presented the proposed model in the context of modeling questionnaire items, there are other situations in which it may be used. One further example is when several tests are given to measure a latent construct and the tests are speeded. The addition of a random intercept allows for consistent individual differences due to speed. That is, respondents who are slow on one test are likely to be slow on all of them, and those who are fast on one test are likely to be fast on all of them. Essentially, an individual's outcome is due to his or her level on the latent construct, the intercepts for each test, and an adjustment (intercept) for his or her speediness. In closing, we believe that the model proposed here may offer a more parsimonious and accurate view of individual differences in the use of response scales than those models that specify additional substantive factors, correlated residuals, or a bifactor model.

References

- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317–332.
- Ansari, A., Jedidi, K., & Dube, L. (2002). Heterogeneous factor analysis models: A Bayesian approach. *Psychometrika*, *67*, 49–77.
- Beck, A. T., Weissman, A., Lester, D., & Trexler, L. (1974). The measurement of pessimism: The Hopelessness Scale. *Journal of Consulting and Clinical Psychology*, *42*, 861–865.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246.
- Bentler, P. M., & Bonett, D. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606.
- Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, *16*, 78–117.
- Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling*, *7*, 608–628.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*, 345–370.
- Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141). Cambridge, England: Cambridge University Press.
- Chang, E. C., D'Zurilla, T. J., & Maydeu-Olivares, A. (1994). Assessing the dimensionality of optimism and pessimism using a multimeasure approach. *Cognitive Therapy and Research*, *18*, 143–160.
- Chang, L., & McBride-Chang, C. (1996). The factor structure of the Life Orientation Test. *Educational and Psychological Measurement*, *56*, 325–329.
- Chen, F. F., West, S. W., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, *41*, 189–224.
- Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, *105*, 317–327.
- DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling*, *9*, 327–346.
- D'Zurilla, T. J., Nezu, A. M., & Maydeu-Olivares, A. (2002). *Manual of the Social Problem-Solving Inventory—Revised*. North Tonawanda, NY: Multi-Health Systems.
- Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5, and 7 response categories: A comparison of categorical variables estimators using simulated data. *British Journal of Mathematical and Statistical Psychology*, *47*, 309–326.
- Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, *65*, 241–261.
- Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, *2*, 41–54.
- Hu, L.-T., & Bentler, P. M. (1999). Cut-off criteria for fit indices in covariance structure analysis: Conventional criteria verses new alternatives. *Structural Equation Modeling*, *6*, 1–55.
- Jöreskog, K. G., & Sörbom, D. (2001). LISREL 8 [Computer software]. Chicago, IL: Scientific Software.
- Krosnick, J. A., & Fabrigar, L. R. (in press). *Designing great questionnaires: Insights from psychology*. New York: Oxford University Press.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *140*, 1–55.
- Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology*, *70*, 810–819.
- Marshall, G. N., Wortman, C. B., Kusulas, J. W., Hergig, L. K., & Vickers, R. R. (1992). Distinguishing optimism from pessimism: Relations to fundamental dimensions of mood and personality. *Journal of Personality and Social Psychology*, *62*, 1067–1074.
- Maydeu-Olivares, A. (2005). Linear IRT, non-linear IRT, and factor analysis: A unified framework. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics. A festschrift to Roderick P. McDonald* (pp. 73–100). Mahwah, NJ: Erlbaum.
- McDonald, R. P. (1999). *Test theory. A unified treatment*. Mahwah, NJ: Erlbaum.
- Olsson, U. (1979). On the robustness of factor analysis against crude classification of the observations. *Multivariate Behavioral Research*, *14*, 485–500.
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 67–88). Hillsdale, NJ: Erlbaum.
- Robinson-Whelen, S., Kim, C., MacCallum, R. C., & Kiecolt-Glaser,

J. K. (1997). Distinguishing optimism from pessimism in older adults: Is it more important to be optimistic or not to be pessimistic? *Journal of Personality and Social Psychology*, 73, 1345–1353.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variable analysis. Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and

health: Assessment and implications of generalized outcome expectancies. *Health Psychology*, 4, 219–247.

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, 67, 1063–1078.

Wolfe, D. (1940). *Factor analysis to 1940* (Psychometric Monograph No. 3). Chicago: University of Chicago Press.

Appendix

Some Technical Details

Mean and Covariance Structure Implied by the Random Intercept Factor Model

Let $\Psi = \text{cov}[\eta'\eta]$, $\Theta = \text{cov}[\mathbf{e}'\mathbf{e}]$, and $\varphi = \text{cov}[\zeta'\zeta]$ (a scalar). We can rewrite Assumptions 1–6 as follows:

1. $E[\eta] = \mathbf{0}$.
2. $E[\mathbf{e}] = \mathbf{0}$.
3. Θ diagonal.
4. $\text{cov}[\mathbf{e}'\eta] = \mathbf{0}$.
5. $E[\zeta] = 0$.
6. $\text{cov}[\eta'\zeta] = \mathbf{0}$ and $\text{cov}[\mathbf{e}'\zeta] = \mathbf{0}$.

Now, putting together the two equations in Equation 5, we have

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{1}\zeta + \Lambda\eta + \mathbf{e}. \quad (\text{A1})$$

Then, the mean structure for \mathbf{y} implied by the random intercept factor model is

$$E[\mathbf{y}] = \boldsymbol{\mu} + \mathbf{1}E[\zeta] + \Lambda E[\eta] + E[\mathbf{e}] = \boldsymbol{\mu}.$$

To obtain the covariance structure implied by this model, we note that by Assumptions 1–6,

$$\begin{aligned} \text{cov}[\mathbf{y}'\mathbf{y}] &= \mathbf{1}\text{cov}[\zeta'\zeta]\mathbf{1}' + \Lambda\text{cov}[\eta'\eta]\Lambda' \\ &\quad + \text{cov}[\mathbf{e}'\mathbf{e}] = \varphi\mathbf{1}\mathbf{1}' + \Lambda\Psi\Lambda' + \Theta. \end{aligned}$$

The random intercept model can be easily estimated using conventional software for structural equation modeling by defining a common factor model

$$\mathbf{y} = \boldsymbol{\mu} + \Lambda^*\eta^* + \mathbf{e}, \quad (\text{A2})$$

where $\eta^* = (\zeta|\eta)'$ is a $(m + 1) \times 1$ vector of random variables, which consists of the random part of the intercept, ζ , and the m common factors, η , and

$$\Lambda^* = (\mathbf{1}|\Lambda) \quad \Psi^* = \begin{pmatrix} \varphi & \mathbf{0} \\ \mathbf{0} & \Psi \end{pmatrix}, \quad (\text{A3})$$

where Λ^* is a $p \times (m + 1)$ matrix of loadings and Ψ^* is the covariance matrix of η^* . Thus, a model with $m + 1$ common factors where the matrices of factor loadings and factor covariances have the pattern in Equation A3, or alternatively

$$\Lambda^* = (\sqrt{\varphi}\mathbf{1}|\Lambda) \quad \Psi^* = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \Psi \end{pmatrix}, \quad (\text{A4})$$

is equivalent to a factor model with m common factors and a random intercept.

To see that the $m + 1$ factor model in Equation A2 with the pattern in Equation A4 or A3 implies the mean and covariance structure in Equation 7, we first note that by Assumptions 1–4, $E[\eta^*] = \mathbf{0}$, $\text{cov}[\mathbf{e}'\eta^*] = \mathbf{0}$, $E[\mathbf{e}] = \mathbf{0}$, $\text{cov}[\mathbf{e}'\mathbf{e}] = \Theta$. Thus, $E[\mathbf{y}] = \boldsymbol{\mu} + \Lambda^*E[\eta^*] + E[\mathbf{e}] = \boldsymbol{\mu}$. As for the covariance matrix, we have

$$\text{cov}[\mathbf{y}'\mathbf{y}] = \Lambda^*\text{cov}[\eta^{*'}\eta^*]\Lambda^{*'} + \text{cov}[\mathbf{e}'\mathbf{e}],$$

which by Equation A4 or A3 simplifies to the covariance structure given in Equation 7.

Effect of Reverse Coding a Subset of Items on the Random Intercept Factor Model

Let \mathbf{y} denote the observed variables prior to recoding and \mathbf{z} denote the recoded variables. We partition \mathbf{z} into \mathbf{z}_1 and \mathbf{z}_2 , where \mathbf{z}_1 denotes the set of items that are not recoded, $\mathbf{z}_1 =$

\mathbf{y}_1 , and \mathbf{z}_2 denotes the set of items that are reverse coded, $\mathbf{z}_2 = (k - 1)\mathbf{1} - \mathbf{y}_2$.

The mean and covariance structures under the random intercept factor model for the original variables are $\boldsymbol{\mu}_y = \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_y = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$, where we partition $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ according to the partitioning of \mathbf{y} as $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$ and $\boldsymbol{\Lambda} = \begin{pmatrix} \mathbf{1} & \boldsymbol{\Lambda}_1 \\ \mathbf{1} & \boldsymbol{\Lambda}_2 \end{pmatrix}$, and $\boldsymbol{\Psi}$ is given by Equation A3.

Now, the reverse coding transformation is

$$\begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ (k-1)\mathbf{1} \end{pmatrix} + \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}.$$

Thus, the mean and covariance structures under the random intercept factor model for the recoded variables are $\boldsymbol{\mu}_z = \boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}_z = \boldsymbol{\Lambda}^*\boldsymbol{\Psi}\boldsymbol{\Lambda}^{*'} + \boldsymbol{\Theta}$, where

$$\boldsymbol{\mu}^* = \begin{pmatrix} \boldsymbol{\mu}_1 \\ (k-1)\mathbf{1} - \boldsymbol{\mu}_2 \end{pmatrix} \quad \boldsymbol{\Lambda}^* = \begin{pmatrix} \mathbf{1} & \boldsymbol{\Lambda}_1 \\ -\mathbf{1} & -\boldsymbol{\Lambda}_2 \end{pmatrix}.$$

Received May 23, 2005

Revision received September 5, 2006

Accepted September 7, 2006 ■

Call for Nominations

The Publications and Communications (P&C) Board has opened nominations for the editorships of **Journal of Applied Psychology**, **Psychological Bulletin**, **Psychology of Addictive Behaviors**, **Journal of Personality and Social Psychology: Interpersonal Relations and Group Processes (IRGP)**, and **Journal of Educational Psychology** for the years 2009-2014. Sheldon Zedeck, PhD, Harris Cooper, PhD, Howard J. Shaffer, PhD, Charles S. Carver, PhD, and Karen R. Harris, PhD, respectively, are the incumbent editors.

Candidates should be members of APA and should be available to start receiving manuscripts in early 2008 to prepare for issues published in 2009. Please note that the P&C Board encourages participation by members of underrepresented groups in the publication process and would particularly welcome such nominees. Self-nominations are also encouraged.

Search chairs have been appointed as follows:

- **Journal of Applied Psychology**, William C. Howell, PhD and J Gilbert Benedict, PhD
- **Psychological Bulletin**, Mark Appelbaum, PhD and Valerie F. Reyna, PhD
- **Psychology of Addictive Behaviors**, Linda P. Spear, PhD and Robert G. Frank, PhD
- **Journal of Personality and Social Psychology: IRGP**, David C. Funder, PhD
- **Journal of Educational Psychology**, Peter A. Ornstein, PhD and Leah L. Light, PhD

Candidates should be nominated by accessing APA's EditorQuest site on the Web. Using your Web browser, go to <http://editorquest.apa.org>. On the Home menu on the left, find "Guests". Next, click on the link "Submit a Nomination," enter your nominee's information, and click "Submit."

Prepared statements of one page or less in support of a nominee can also be submitted by e-mail to Susan J.A. Harris, P&C Board Search Liaison, at sjharris@apa.org.

Deadline for accepting nominations is **January 10, 2007**, when reviews will begin.