MULTIVARIATE BEHAVIORAL RESEARCH, *41*(2), 189–225

A Comparison of Bifactor and Second-Order Models of Quality of Life

Fang Fang Chen, Stephen G. West, and Karen H. Sousa
Arizona State University

Bifactor and second-order factor models are two alternative approaches for representing general constructs comprised of several highly related domains. Bifactor and second-order models were compared using a quality of life data set ($N = 403$). The bifactor model identified three, rather than the hypothesized four, domain specific factors beyond the general factor. The bifactor model fit the data significantly better than the second-order model. The bifactor model allowed for easier interpretation of the relationship between the domain specific factors and external variables, over and above the general factor. Contrary to the literature, sufficient power existed to distinguish between bifactor and corresponding second-order models in one actual and one simulated example, given reasonable sample sizes. Advantages of bifactor models over second-order models are discussed.

Researchers interested in assessing a construct often hypothesize that several highly related domains comprise the general construct of interest. In the motivating example, which is the focus of this article, Stewart and Ware (1992) proposed measures of health-related quality of life. They hypothesized that this general construct was comprised of at least four highly related domain specific factors (cognition, vitality, mental health, and disease worry). Building on earlier work (e.g., Hays & Stewart, 1990; Lubeck & Fries, 1993; McHorney, Ware, & Raczek, 1993, Stewart & Ware, 1992; Ware, Davies-Avery, & Brook, 1980), these authors selected 17-items to represent the four domains that were hypothesized to comprise the general construct. Other researchers have also hypothesized general constructs

---

Fang Fang Chen is now at the Department of Psychology, University of Delaware. Stephen G. West is in the Department of Psychology at Arizona State University, and Karen H. Sousa is with the College of Nursing at Arizona State University.

This project was partly supported by National Institute of Nursing Research (NR04817) to Karen Sousa. We thank Roger Millsap for his comments on an earlier draft of this manuscript.

Correspondence should be sent to Fang Fang Chen, University of Delaware, Department of Psychology, Wolf Hall, Newark, DE 19716. E-mail: xiyu@psych.udel.edu or Stephen G. West, Psychology Department, Arizona State University, Tempe, AZ 85287–1104. E-mail: sgwest@asu.edu

that are comprised of several closely related domains. For examples, Costa and McCrae (1992, 1995) hypothesized a structure for each of the five dimensions in their revised NEO personality inventory. Jackson, Ahmed, and Heapy (1976) proposed a general structure that was comprised of several domains for achievement motivation. In these hypothesized structures, the central scientific interest lies in the general construct. There may also be a secondary interest in whether a more focused domain specific construct may make a unique contribution, over and above the general factor, to the prediction of external criteria.

Within confirmatory factor analysis, two alternative models (described in the next section) have been proposed to represent the factor structure of items that assess several highly related domains that are hypothesized to comprise a general construct. Bifactor models, also known as general-specific or nested models[1], are less familiar because they have been used primarily in the area of intelligence research in the recent literature (e.g., Gustafsson & Balke, 1993; Luo, Petrill, &Thompson, 1994). Second-order models are more familiar as they have been applied in a wider variety of substantive areas, such as personality (DeYoung, Peterson, & Higgins, 2002; Judge, Erez, Bono, & Thoresen, 2002), self-concept (Marsh, Ellis, & Craven, 2002), and psychological well-being (Hills & Argyle, 2002). Our first goal in this article is to illustrate the potential advantages of bifactor models over second-order models when researchers have an interest in predicting external criteria. Our second goal is to address concerns raised by Mulaik and Quartetti (1997) that bifactor and second-order models may not be statistically distinguishable given sample sizes commonly used in the behavioral sciences. Following our explication and comparison of the models, we address the important issue of orthogonality constraints associated with both models in the general discussion.

CONNECTING BIFACTOR MODELS AND SECOND-ORDER MODELS

Bifactor models are potentially applicable when (a) there is a general factor that is hypothesized to account for the commonality of the items; (b) there are multiple domain specific factors, each of which is hypothesized to account for the unique influence of the specific domain over and above the general factor; and (c) researchers may be interested in the domain specific factors as well as the common factor that is of focal interest. Figure 1 illustrates a bifactor model of health quality of life derived from the second-order model originally proposed by Stewart and Ware (1992). In this case there is a single quality of life factor that underlies each of

[1]Holzinger and Swineford (1937) originally termed this model the *bifactor model*. Several recent presentations of this model have used the term *nested* or *general-specific* model.

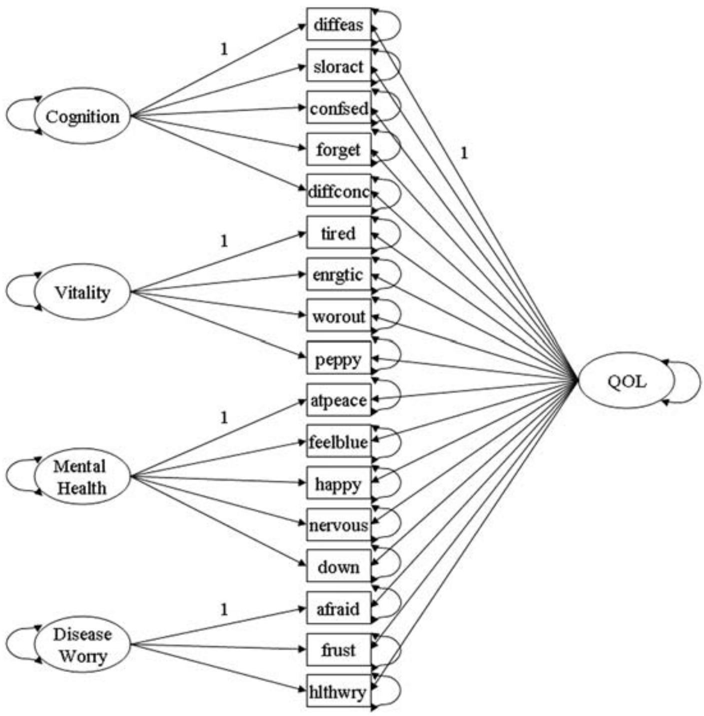

FIGURE 1 A bifactor model of quality of life. *Note.* Items with acronym in parentheses. Items that are reverse coded are denoted with (R). (a) Cognition subscale: "Have difficulty reasoning and solving problems?" (diffeas); "React slowly to things that were said or done?" (sloact); "Become confused and start several actions at a time?" (confused); "Forget where you put things or appointments?" (forget); "Have difficulty concentrating?" (diffconc). (b) Vitality subscale: "Feel tired?" (tired); "Have enough energy to do the things you want?" (R) (enrgtic); "Feel worn out?" (worout); "Feel full of pep?" (R) (peppy). (c) Mental health subscale: "Feel calm and peaceful?"(R) (atpeace)? "Feel downhearted and blue?" (feelblue); "Feel very happy"(R) (happy); "Feel very nervous?" (nervous); "Feel so down in the dumps nothing could cheer you up? (down). (d) Disease worry subscale: "Were you afraid because of your health?" (afraid); "Were you frustrated about your health?" (frust); "Was your health a worry in your life?" (healthwry). Second-order factor: Quality of Life (QOL)—high scores indicate high quality of life.

the items. Separately, there are domain specific factors of cognition, vitality, mental health, and disease worry, each of which accounts for unique variance in its own separate set of domain-related items. We consider the canonical bifactor model in which the relations among the general and domain specific factors are assumed to be orthogonal, as the domain specific factors are related to the contribution that is over and above the general factor.

In RAM notation, the curved double headed arrows represent factor variances for each of the latent variables and measurement error variance for each of the measured variables. Factor loadings designated with 1 are marker variables; all other factor loadings are estimated.

For the model in Figure 1, let vector Y represent observed variables; matrix $\boldsymbol{\Lambda}_y$ represents the factor loadings of the general and domain specific factors; vector $\boldsymbol{\eta}$ represents the general and domain specific factors; and vector $\boldsymbol{\epsilon}$ represents residual variance (uniquenesses).

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \\ y_{16} \\ y_{17} \end{bmatrix} \boldsymbol{\Lambda}_y = \begin{bmatrix} \lambda_{g1,1} & \lambda_{s1,1} & 0 & 0 & 0 \\ \lambda_{g2,1} & \lambda_{s2,1} & 0 & 0 & 0 \\ \lambda_{g3,1} & \lambda_{s3,1} & 0 & 0 & 0 \\ \lambda_{g4,1} & \lambda_{s4,1} & 0 & 0 & 0 \\ \lambda_{g5,1} & \lambda_{s5,1} & 0 & 0 & 0 \\ \lambda_{g6,1} & 0 & \lambda_{s6,2} & 0 & 0 \\ \lambda_{g7,1} & 0 & \lambda_{s7,2} & 0 & 0 \\ \lambda_{g8,1} & 0 & \lambda_{s8,2} & 0 & 0 \\ \lambda_{g9,1} & 0 & \lambda_{s9,2} & 0 & 0 \\ \lambda_{g10,1} & 0 & 0 & \lambda_{s10,3} & 0 \\ \lambda_{g11,1} & 0 & 0 & \lambda_{s11,3} & 0 \\ \lambda_{g12,1} & 0 & 0 & \lambda_{s12,3} & 0 \\ \lambda_{g13,1} & 0 & 0 & \lambda_{s13,3} & 0 \\ \lambda_{g14,1} & 0 & 0 & \lambda_{s14,3} & 0 \\ \lambda_{g15,1} & 0 & 0 & 0 & \lambda_{s15,4} \\ \lambda_{g16,1} & 0 & 0 & 0 & \lambda_{s16,4} \\ \lambda_{g17,1} & 0 & 0 & 0 & \lambda_{s17,4} \end{bmatrix} \boldsymbol{\eta} = \begin{bmatrix} \eta_{g1} \\ \eta_{s1} \\ \eta_{s2} \\ \eta_{s3} \\ \eta_{s4} \end{bmatrix} \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \\ \epsilon_{10} \\ \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \epsilon_{15} \\ \epsilon_{16} \\ \epsilon_{17} \end{bmatrix}.$$

The observed variables can be expressed in the following equation:

$$Y = \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\epsilon}.$$

The first term represents the contribution of the general and domain specific factors, the second term represents the contribution of the residual variance.[2]

[2]The residual variance is composed of measurement error and variance that is not captured by the general factor and the specific factors.

The other commonly used model is the second-order model. Like bifactor models, second-order models are also used when it is hypothesized that measurement instruments assess several highly related domains. Second-order models are potentially applicable when (a) the lower-order factors are substantially correlated with each other, and (b) there is a higher-order factor that is hypothesized to account for the relationship among the lower-order factors. Figure 2 illustrates the second-order model of quality of life proposed by Stewart and Ware (1992). In this hierarchical structure, quality of life is a general factor that accounts for the commonality among lower order factors representing each of the four domains: cognition, vitality, mental health, and disease worry. Multiple items on the measure in turn, represent each of the lower order factors. Such second-order models can be estimated and the fit of the second-order structure can be statistically tested so long as four or more first order factors are hypothesized.

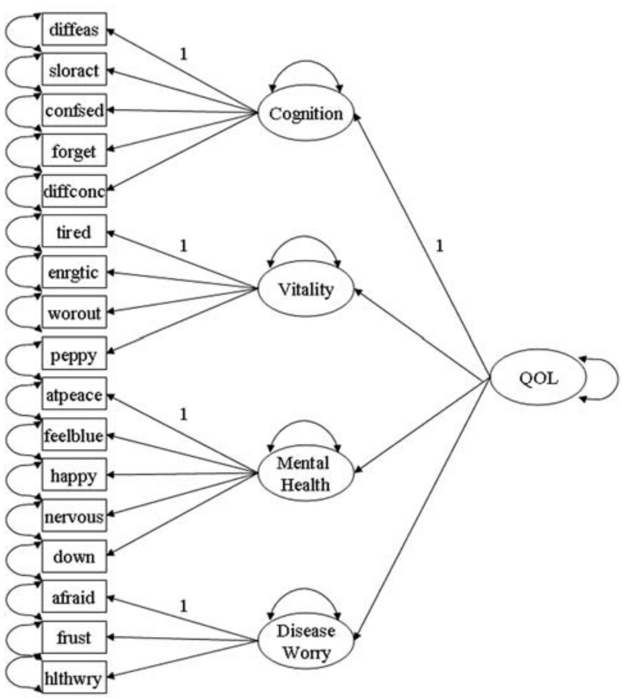

FIGURE 2 A standard second-order factor model of quality of life. *Note*. In RAM notation, the curved double headed arrow represents the factor variance for the second order factor (QOL), disturbance variance for each of the first-order factors, and measurement error variance for each of the measured variables. Factor loadings designated with 1 are marker variables; all other factor loadings are estimated.

For the model in Figure 2, let the Y vector represent observed variables; $\mathbf{\Lambda}_y$ represents the loadings of the measured variables on the first-order factors; the $\boldsymbol{\eta}$ vector represents the lower-order factors; the $\mathbf{\Gamma}$ vector represents the loadings of the lower-order factors on the higher-order factor; the $\boldsymbol{\xi}$ vector represents the higher-order factor; the $\boldsymbol{\zeta}$ vector represents the disturbances of the lower-order factors (unique variance that is not shared with the common higher-order factor); and the $\boldsymbol{\epsilon}$ vector represents the residuals. The equations for the second-order model are:

$$\boldsymbol{\eta} = \mathbf{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$$

$$Y = \mathbf{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon}.$$

The first equation above represents the structure for each of the lower order factors; the second equation above represents the measurement model for the observed variables.

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \\ Y_8 \\ Y_9 \\ Y_{10} \\ Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{15} \\ Y_{16} \\ Y_{17} \end{bmatrix} \mathbf{\Lambda}_y = \begin{bmatrix} \lambda_{1,1} & 0 & 0 & 0 \\ \lambda_{2,1} & 0 & 0 & 0 \\ \lambda_{3,1} & 0 & 0 & 0 \\ \lambda_{4,1} & 0 & 0 & 0 \\ \lambda_{5,1} & 0 & 0 & 0 \\ 0 & \lambda_{6,2} & 0 & 0 \\ 0 & \lambda_{7,2} & 0 & 0 \\ 0 & \lambda_{8,2} & 0 & 0 \\ 0 & \lambda_{9,2} & 0 & 0 \\ 0 & 0 & \lambda_{10,3} & 0 \\ 0 & 0 & \lambda_{11,3} & 0 \\ 0 & 0 & \lambda_{12,3} & 0 \\ 0 & 0 & \lambda_{13,3} & 0 \\ 0 & 0 & \lambda_{14,3} & 0 \\ 0 & 0 & 0 & \lambda_{15,4} \\ 0 & 0 & 0 & \lambda_{16,4} \\ 0 & 0 & 0 & \lambda_{17,4} \end{bmatrix} \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{bmatrix} \mathbf{\Gamma} = \begin{bmatrix} \Gamma_{1,1} \\ \Gamma_{2,1} \\ \Gamma_{3,1} \\ \Gamma_{4,1} \end{bmatrix} \boldsymbol{\xi} = [\xi] \; \boldsymbol{\zeta} = \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \end{bmatrix} \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \\ \epsilon_{10} \\ \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \epsilon_{15} \\ \epsilon_{16} \\ \epsilon_{17} \end{bmatrix}.$$

Early work suggested that the bifactor and second-order models are mathematically equivalent (Wherry, 1959). However, newer work by several researchers

(Gustafsson & Balke, 1993; McDonald, 1999; Mulaik & Quartetti, 1997; Yung, Thissen, & McLeod, 1999) has independently pointed out that these two models are *not* generally equivalent. A bifactor model and a second-order model are mathematically equivalent *only* when proportionality constraints are imposed using the Schmid-Leiman transformation method (Schmid & Leiman, 1957). The Schmid-Leiman transformation imposes two specific constraints: (a) the factor loadings of the general factor in the bifactor model must be the product of the corresponding lower-order factor loadings and the second-order factor loadings in the second-order models (Mulaik & Quartetti, 1997); and (b) the ratio of the general factor loading to its corresponding domain specific factor loading is the same within each domain specific factor (Yung, et al., 1999).

Yung et al. (1999) have used the *generalized* Schmidt-Leiman transformation, in which no proportionality constraints are imposed, to demonstrate that second-order models are in fact nested within the bifactor models (see also Rindskopf & Rose, 1988). For every bifactor model (see Figure 1), there is an equivalent full second-order model with direct effects (factor loadings) from the second-order factor to every observed variable, over and above the second-order effect on the lower order factors (see Figure 3). A standard second-order model (see Figure 2) is a special case (constrained version) of the full second-order model with the direct effects from the second-order factor to the observed variables eliminated. In other words, the "reduced" second-order model is more restricted than the full second-order model, which is equivalent to the bifactor model. Consequently, the "reduced" second-order model is more restricted than the bifactor model (A is nested in B, B is equal to C, so that A is nested in C).

Although the second-order model is not mathematically equivalent to the bifactor model, these two models have similar interpretations (Gustafsson & Balke, 1993). First, the second-order factor in the second-order model corresponds to the general factor in the bifactor model; second, the disturbances of the first-order factors in the second-order model resemble the domain specific factors in the bifactor model; and third, in the bifactor model the general factor and the domain specific factors are assumed to be orthogonal paralleling the representation in the second-order model in which the second-order factor and the disturbances (unique factors) are defined to be orthogonal. However, the differences between the two models become more important when researchers are also interested in the contribution of the one or more of the domain specific factors over and above the general/second-order factor.

ADVANTAGES OF BIFACTOR MODELS

Bifactor models have several potential advantages over second-order models, particularly when researchers may be interested in the predictive relationships be-

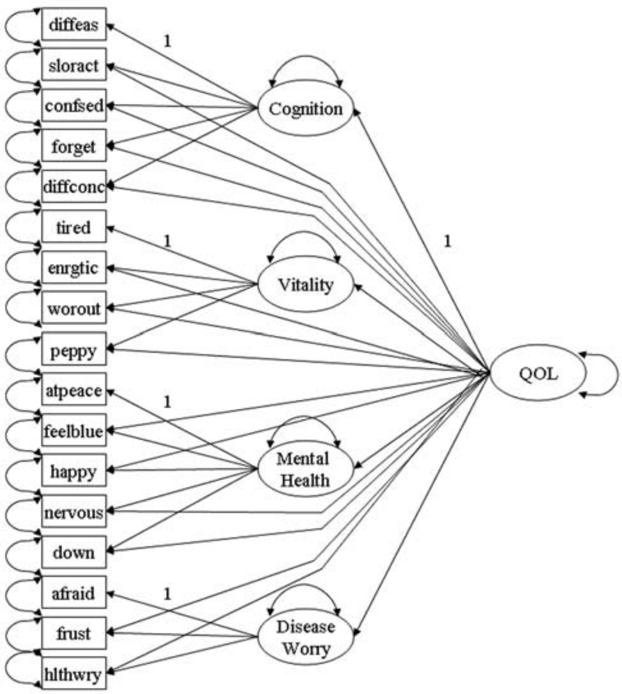

FIGURE 3 A full second-order factor model of quality of life with direct effects from the second-order factor to the observed variables. *Note*. In RAM notation, the curved double headed arrow represents the factor variance for the second-order factor (QOL), disturbance variance for each of the first-order factors and measurement error variance for each of the measured variables. Factor loadings designated with 1 are marker variables; all other factor loadings are estimated. To achieve identification, there is no path from QOL to each of the measured variables that serves as a marker for a first order factor.

tween domain specific factors and external criteria, over and above the general/second-order factor. First, a bifactor model can be used as a less restricted baseline model to which a second-order model can be compared, given that the second-order model is nested within the bifactor model (Yung et al., 1999). A likelihood ratio test (chi-square difference test) can be used to distinguish the two models.

Second, the bifactor model can be used to study the role of domain specific factors that are independent of the general factor. For example, drawing on Spearman's (1927) conception of general and domain specific factors, suppose there is factor of general intelligence of focal interest as well as domain specific factors of intelligence-related abilities such as verbal, spatial, mathematical, and analytic. However, suppose that verbal ability reflects only general intelligence, whereas spatial, mathematical, and analytic abilities still exist as specific domains, even after partialling out general intelligence. In this example, verbal ability will *not* exist as a domain specific factor in the bifactor model. If verbal ability is then included as a hypothesized domain specific factor in the bifactor model, problems in estimation will occur because of identification problems due to factor overextraction (Rindskopf, 1984). These identification problems will be manifested in two primary ways: (a) factor loadings of the verbal domain specific factor will be small and nonsignificant; and (b) the variance of the verbal factor will not be significant. In this example the common variance in the verbal ability items is entirely explained by the general intelligence factor. However, such problems will typically not be easily discovered in the second-order model, as verbal ability will legitimately exist as a lower-order factor. The second-order factor model will manifest the nonexistence of the domain specific factor in the variance of the disturbance of the lower order factor. The variance of the disturbance may not be significant as it represents the domain specific verbal ability factor. The lack of significance in the variance of the disturbance will typically not cause any problem in a model, and therefore, the possibility that one or more of the domain specific factors may not exist can be easily glossed over by researchers examining the output from second-order models.

Third, in the bifactor model, we can directly examine the strength of the relationship between the domain specific factors and their associated items, as the relationship is reflected in the factor loadings, whereas these relationships cannot be directly tested in the second-order factor model as the domain specific factors are represented by disturbances of the first-order factors.

Fourth, the bifactor model can be particularly useful in testing whether a subset of the domain specific factors predict external variables, over and above the general factor, as the domain specific factors are directly represented as independent factors (Gustafsson & Balke, 1993). Domain specific factors (disturbances) may also be used to predict external criteria, over and above the second-order factor, in second-order models, but such tests may require the use of nonstandard structural equation models (Bentler, 1990). That is, the disturbances of the first-order factors must be used as predictors (Gustafsson & Balke, 1993). However, such nonstandard models are not easily implementable in many of the standard structural equation modeling software packages and the results may be difficult to explain to substantive researchers. Of importance, with either approach only a limited set of domain specific factors may be included with the general factor in the prediction of an external criterion. Otherwise, exact linear dependence of the predictors will result and the model cannot be properly estimated. Most current structural equation modeling software does not include adequate checks to reliably detect this problem.

Fifth, in the bifactor model, we can test measurement invariance of the domain specific factors, in addition to the general factor in different groups (e.g., males vs.

females). In contrast, in the second-order model, only the second-order factor can be directly tested for invariance between groups, as the domain specific factors are represented by disturbances. Measurement invariance involves testing the equivalence of measured constructs in two or more independent groups to assure that the same constructs are being assessed in each group.

Finally, in the bifactor model, latent mean differences in both the general and domain specific factors can be compared across different groups, given an adequate level of measurement invariance. In contrast, in the second-order model only the second-order latent means can be directly compared. For example, using a bifactor model, Gustafsson (1992) was able to compare latent mean differences between white and black participants in general intelligence as well as three domain specific abilities: "broad" verbal ability, "broad" spatial-figure ability, and "narrow" memory span ability. In their Swedish sample it was found that the white participants scored higher than black participants on all abilities except for memory span.

POWER

The second goal of the study is to examine whether there is sufficient power to differentiate the bifactor and second-order models. A common strategy for estimating the power of differentiating two nested models in structural equation modeling was developed by Satorra and Saris (1985). The likelihood ratio chi-square statistic is calculated for the target model and more restricted model, respectively. The difference in the chi-square statistic between the target model and the more restricted model is the noncentrality parameter of the noncentral chi-square distribution. The noncentrality parameter and the difference in degrees of freedom between the two models are used in the estimation of statistical power.

Mulaik and Quartetti (1997) conducted an investigation to examine whether there is enough power to distinguish between the bifactor and second-order models. In the initial example of Mulaik and Quartetti, they generated a population covariance matrix based on a hypothetical second-order model. As expected from theory, the second-order and the bifactor factor models both fit the covariance matrix almost perfectly, because the second-order model is more restricted than the bifactor model. Based on the second-order model, they also set up a Schmid–Leiman decomposition table with *proportionality constraints*. In other examples, they used this Schmid–Leiman decomposition table to generate bifactor models except that they *slightly* perturbed some of the loadings in the table so that the bifactor models would be similar but not equivalent to the corresponding second-order models. Population covariance matrices were generated based on these perturbed bifactor models. In Example 2, an incomplete second-order model with $n = 1{,}000$ was fit to the population covariance matrix based on a perturbed incomplete bifactor model, yielding $\chi^2(101) = 3.59$. In Example 3, a complete second-order model was fit to the population covariance matrix based on a perturbed complete bifactor model, yielding $\chi^2(100) = 18.81$ and 37.66 for sample sizes of 500 and 1000, respectively. The χ^2 statistic obtained from the second-order model was used as the noncentrality parameter to calculate power.

Of importance, instead of calculating the power to distinguish between the bifactor and second-order models (i.e., comparing the two nested models), Mulaik and Quartetti (1997) calculated the power to reject the null hypothesis that the second-order model could have generated the population covariance matrix from the perturbed bifactor model (i.e., an omnibus test of the absolute fit of the second-order model to the data). Two independent population covariance matrices were generated and tested, the first with a sample size of 1,000, and the second with sample sizes of 500 and 1,000. It was found that the power to distinguish the two models was .35 for a sample of 500, and for two independent samples of 1,000, the power ranged from .18 to .75. Based on these results, Mulaik and Quartetti concluded that the power to distinguish the second-order model from the bifactor model is unacceptably low for sample sizes commonly used in behavioral science research.

There are two possible reasons that may account for the low power observed in Mulaik and Quartetti's (1997) study to differentiate the two models. The first reason is that Mulaik and Quartetti focused on the omnibus test of absolute fit rather than the focused test comparing the nested bifactor and second-order models. When we recalculated the power based on the nested models using the data provided in the Mulaik and Quartetti article, power increased from .35 to .84 for the sample of 500, and the power increased from .08 (*not* .18) to .20, and from .75 to > .99, respectively[3], for two independent samples of 1,000**.** The second reason is related to the particular covariance matrices used in the study. As discussed earlier, the Schmid-Leiman transformation puts proportionality constraints on the factor loadings of the general factor and domain specific factors in the bifactor model. In Examples 2 and 3, Mulaik and Quartetti made changes to the Schimid-Leiman transformation table produced by the second-order model; however, *only very small* changes were made. The factor loadings in the perturbed bifactor models deviated by only .01 to .05 from those produced by the Schmid–Leiman transformation. Thus, the factor loadings still generally followed the proportionality constraints. The very small value of the obtained noncentrality parameter is consistent with this view, although it is difficult to interpret in the absence of a familiar metric for model comparisons. MacCallum, Browne, and Cai (2006) have tentatively suggested the change in the RMSEA may provide a useful "standardardized" metric in power analysis for assessing the degree of discrepancy between two confirmatory

[3]We recalculated all of Mulaik and Quartetti (1997) power estimates. All matched except for the value of .18 for which we calculated a value of .08. We believe Mulaik and Quartetti's value of .18 reflects a typographical error.

factor models. The RMSEAs for both the complete second-order and the complete bifactor models (Example 3) evaluated by Mulaik and Quartetti were 0. These results strongly suggest that the lack of power observed by Mulaik and Quartetti was the result of the trivial amount of discrepancy between the two models. We sought to examine the power to detect the difference between the bifactor and second-order models in the more general situation in which proportionality constraints are *not* imposed on the models and a nontrivial amount of misspecification existed.

STUDY 1

The bifactor and second-order factor structures of a 17-item health-related quality of life measurement from the AIDS Time-Oriented Health Outcome Study (ATHOS) were compared. ATHOS is a longitudinal observational database that attempted to represent a population of people with HIV-associated illness cared for by community-based providers. The sample was collected in the early 1990s from three community-based providers in the greater San Francisco area, two private practices in Los Angeles, and five community clinics in San Diego. As noted above, the measurement of health-related quality of life is composed of four subscales: Cognition, Vitality, Mental Health, and Disease Worry. Brief descriptions of the full set of items are given in the caption to Figure 1. Items were answered on a 5-point scale ranging from 1 (*all of the time*) to 5 (*never*) so that high scores on the scale represent high quality of life.

To examine the relationship between a domain specific factor and an external criterion variable, over and above the relationship between the general factor and the criterion variable, a measure of social functioning was also included. Social functioning has been defined as “the ability to develop, maintain, and nurture major social relationships” (Sherbourne, 1992, p. 173). Social functioning has been considered as being indicative of physical and mental health status. Lubeck and Fries (1993) developed a social functioning question as part of their battery of self-administered measures of quality of life. The question is, “Compared to your usual level of social activity, in the past three months, has your level of social activity decreased, stayed the same, or increased because of a change in your physical health or concern about your health?” The responses are 1 “far less active than usual,” 2 “somewhat less active than usual,” 3 “as active as usual,” 4 “somewhat more active than usual,” and 5 “much more active than usual.”

We limited the sample to participants who had either a part-time or full-time job at the time of assessment. We wanted to be certain that no participant in the sample was so debilitated by physical illness as to seriously restrict social functioning. We further limited our sample to White Caucasian males. A relatively small number of Caucasian female (6) and non-Caucasian (67) participants were also included in the original sample. We excluded these individuals to achieve a more homogeneous sample. This resulted in a sample size of 403 with a mean age of 40. All participants were HIV-positive at the time of their enrollment into the study. We used the data from each participant's initial measurement following enrollment.

We investigated the two alternative hypothesized structures for the quality of life instrument. First, we investigated a bifactor structure in which there is a general factor of global quality of life and domain specific factors of cognition, vitality, mental health, and disease worry (see Figure 1). Second, we investigated a second-order factor structure for the quality of life instrument, with cognition, vitality, mental health, and disease worry as the lower-order factors, and global quality of life as the higher-order factor (Stewart & Ware, 1992) (see Figure 2). We conducted tests of the factor structure using confirmatory factor analysis (CFA) based on the covariance structure. Analyses were conducted in two major stages using the Mplus 3.01 program (Muthén & Muthén, 1998) and maximum likelihood estimation. First, CFA procedures were used to test the fit of the bifactor and second-order factor models, respectively. Second, given findings of model tenability, the domain specific factors and the general/second-order factor were used to predict the external criterion variable, social functioning. Maximum likelihood procedures were used because initial examination of the data did not show evidence of excessive non-normality[4] (skewness: median = –.47; range = –1.04 to –.01; excess kurtosis: median = –.04; range = –.40 to .65). The covariance matrix for the 17 items is presented in Table 1.

Test of the Unrestricted Four-Factor Model

Before testing the hypothesized models, an unrestricted exploratory four-factor model was tested. An unacceptable fit of the unrestricted model would prevent further tests of more restricted models (Jöreskog, 1979; Mulaik & Millsap, 2000). We used conventional criteria for three fit indices that are widely used in the literature. Browne and Cudeck (1993) stated that "a value of the RMSEA of about 0.05 or less would indicate a close fit of the model" and "a value of 0.08 or less would indi-

[4]We also addressed non-normality by comparing the results obtained with alternative estimators. First, we used the alternative WLSMV estimator for ordered categorical data available in Mplus (Muthén & Muthén, 1998). The fit does not change substantially except for RMSEA, which becomes worse. For example, for the incomplete bifactor model, with WLSMV estimation, $\chi^2 = 79.69$ ($df = 45$), RMSEA = .086 (vs. .058 with ML); WRMR = .949 (vs. SRMR = .038 with ML, although these indices are not directly comparable because they are in different metrics); CFI = .964 (vs. .966 with ML). Second, we also explored alternative estimators in the Satorra-Bentler family. Neither MLR (with robust standard errors) nor MLMV (with robust standard errors and a mean- and variance-adjusted chi-square test statistic) significantly improved model fit. With MLR, $\chi^2 = 232.76$ ($df = 107$), RMSEA = .054 (vs. .058 with ML); SRMR = .036 (vs. .038 with ML); CFI = .966 (vs. .966 with ML); With MLMV, $\chi^2 =$ 152.29 ($df = 70$), RMSEA = .054 (vs. .058 with ML); SRMR = .036 (.038 with ML); CFI = .965 (vs. .966 with ML). These results are generally consistent with the conclusions of a review by West, Finch, and Curran (1995) on the effects of non-normality on fit.

TABLE 1
Covariance Matrix of Health Related Quality of Life Items (Study 1)

y	0.644																	
*x*1	0.205	0.667																
*x*2	0.160	0.370	0.634															
*x*3	0.220	0.505	0.400	0.763														
*x*4	0.176	0.447	0.335	0.522	0.711													
*x*5	0.195	0.532	0.376	0.511	0.491	0.763												
*x*6	0.280	0.239	0.208	0.263	0.215	0.261	0.571											
*x*7	0.284	0.228	0.218	0.245	0.179	0.230	0.386	0.773										
*x*8	0.270	0.253	0.231	0.283	0.234	0.292	0.428	0.378	0.596									
*x*9	0.305	0.235	0.212	0.256	0.180	0.270	0.390	0.376	0.405	0.641								
*x*10	0.160	0.211	0.190	0.289	0.211	0.278	0.266	0.264	0.256	0.333	0.590							
*x*11	0.203	0.286	0.228	0.292	0.240	0.329	0.321	0.309	0.312	0.329	0.413	0.683						
*x*12	0.193	0.224	0.184	0.239	0.177	0.248	0.289	0.299	0.262	0.378	0.393	0.366	0.657					
*x*13	0.193	0.361	0.282	0.397	0.322	0.399	0.319	0.264	0.306	0.322	0.366	0.412	0.298	0.812				
*x*14	0.258	0.353	0.266	0.329	0.264	0.349	0.327	0.333	0.343	0.352	0.365	0.498	0.380	0.447	0.765			
*x*15	0.250	0.293	0.255	0.325	0.232	0.339	0.277	0.339	0.326	0.306	0.365	0.439	0.325	0.434	0.414	0.936		
*x*16	0.319	0.272	0.233	0.321	0.233	0.321	0.331	0.340	0.378	0.370	0.356	0.422	0.354	0.385	0.397	0.623	0.855	
*x*17	0.235	0.255	0.239	0.291	0.215	0.271	0.277	0.304	0.296	0.335	0.345	0.399	0.336	0.377	0.339	0.669	0.584	0.940

Note. *y*-social functioning; *x*1-diffeas; *x*2-sloract; *x*3-confsed; *x*4-forget; *x*5-diffconc; *x*6-tired; *x*7-enrgtic; *x*8-worout; *x*9-peppy; *x*10-atpeace; *x*11-feelblue; *x*12-happy; *x*13-nervous; *x*14-down; *x*15-afraid; *x*16-frust; *x*17-hlthwry. $N = 403$. See caption to Figure 1 for full description.

cate a reasonable error of approximation" (p. 144). Based on their extensive simulation study, Hu and Bentler (1999) recommended cutoffs of .95 for the CFI, and .08 for the SRMR, respectively.

The four-factor unrestricted model was defined as the follows: (a) there were four common factors: cognition, vitality, mental health, and disease worry; (b) a marker item was chosen for each factor[5], and the marker variable's loading on the factor that was designed to measure was freely estimated and was set to 0 on the other factors (Jöreskog, 1979; Mulaik & Millsap, 2000); (c) other non-marker items were allowed to load on all factors; (d) the four factors were correlated with each other; and (e) error terms associated with each item were uncorrelated. To identify the model, the variances of the factors were set to 1.

As can be seen from Table 2, $\chi^2(74) = 165.99$, $p < .001$; RMSEA was .056 (*ns*); SRMR was .018; and CFI was .978. Given an adequate fit of the unrestricted model to the data, further analyses were conducted.

Test of the Bifactor Model

Testing the Hypothesized Bifactor Model (Figure 1). The bifactor factor model hypothesized that the responses could be explained by one general factor, which we term global health-related quality of life, and four domain specific factors: cognition, vitality, mental health, and disease worry. The model was defined as the following: (a) each item had a nonzero loading on the factor that was designed to measure, and zero loadings on the other factors; (b) the five factors were uncorrelated with each other; and (c) error terms associated with each item were uncorrelated. To identify the model, in addition to setting one of the factor loadings in the general factor to 1, one of the loadings in each of the domain specific factors was also set to 1. The variances of the factors were estimated.

The fit statistics are presented in Table 2 and the unstandardized and standardized factor loadings are presented in Table 3. As can be seen from Table 2, $\chi^2(102) = 206.79$, $p < .001$; RMSEA was .050 (*ns*); SRMR was .033; and CFI was .975, indicating an adequate fit of the data.

It is important to note that 4 of the 5 factor loadings for the mental health factor (Table 3) were not significant, and, contrary to prediction, 3 of the 5 factor loadings were *negative* (items had been reverse scored). In addition, the variance of the mental health factor was nonsignificant and negative, –.06, *ns* (Heywood case), indicating that the model was misspecified. Standardized solutions related to the mental health factor were not calculated because of the negative variance estimate for this factor. These results suggest that mental health may not exist as a domain

[5]The choice of the specific item among the items that are hypothesized to measure a given factor to serve as the marker variable is arbitrary (Mulaik & Millsap, 2000). For simplicity, we chose the first item of each factor.

TABLE 2
Summary of Fit Statistics for Bifactor and Second-Order Models
of Quality of Life (Study 1)

						Bifactor vs. Second		
	x^2	*df*	*RMSEA*	*SRMR*	*CFI*	Δx^2	Δdf	*Power*
Unrestricted four-factor model	165.99*	74	.056 (*ns*)	.018	.978			
Bifactor model (Figure 1)	206.79*	102	.050 (*ns*)	.033	.975			
Second-order model with direct effects (Figure 3)	206.79*	102	.050 (*ns*)	.033	.975			
Second-order model (Figure 2)	279.35*	115	.060	.040	.961	72.56*	13	>.99
Incomplete bifactor model (Figure 4)	252.10*	107	.058 (*ns*)	.036	.966			
Incomplete second-order model (Figure 5)	292.62*	117	.061	.053	.959	40.52*	10	>.99

Note. $N = 403$. RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; CFI = comparative fit index.
$^*p < .001$.

specific factor, over and above the general factor, and that the variance related to mental health is explained by the general factor.

Testing the Incomplete Bifactor Factor Model (Figure 4). The bifactor model was then modified by removing the mental health domain specific factor from the model. This consists of an incomplete bifactor model, and the factor loadings are presented in Table 4. As can be seen from Table 2, $\chi^2(107) = 252.10$, $p < .001$; RMSEA was .058 (*ns*); SRMR was .036; and CFI was .966. These results indicate an adequate fit of the data[6]. No problems with any of the estimates were observed.

Test of the Second-Order Factor Model

Testing a second-order factor model with direct effects that is equivalent to the bifactor model (Figure 3). The second-order factor model hypothesized that the responses to quality of life could be explained by four first-order factors (cognition, vitality, mental health, and disease worry) and further, by one second-order factor of quality of life underlying the first-order factors. Yung et al. (1999) demonstrated that for every bifactor model, there is an equivalent second-order model with direct effects from the second-order factor to the observed

[6]It is possible that another model with additional parameters may fit these data better. To facilitate exploration of alternative models, the covariance matrix is provided in Table 1.

TABLE 3
Factor Loadings From Bifactor Model and Second-Order Factor Model With Direct Effects From the Second-Order Factor to the Observed Variables (Study 1)

	Bifactor Model (Figure 1)					*Second-Order Model With Direct Effects (Figure 3)*				
Item/Factor	*General Factor*	*Cognition*	*Vitality*	*Mental Health*	*Disease Worry*	*Second-Order*	*Cognition*	*Vitality*	*Mental Health*	*Disease Worry*
diffeas	1.000 (.566)[a]	1.000 (.635)				0[c]	1.000 (.851)			
sloract	.819 (.476)	.702 (.458)				.116 (.068)	.702 (.613)			
confsed	1.061 (.562)	1.054 (.626)				.007 (.004)	1.054 (.839)			
forget	.830 (.455)	1.078 (.664)				–.248 (–.136)	1.078 (.889)			
diffconc	1.108 (.587)	1.006 (.598)				.102 (.054)	1.006 (.801)			
tired	1.059 (.649)		1.000 (.562)			0[c]		1.000 (.858)		
enrgtic	1.044 (.549)		.803 (.388)			.193 (.102)		.803 (.592)		
worout	1.070 (.641)		1.021 (.561)			–.012 (–.007)		1.021 (.857)		
peppy	1.166 (.673)		.733 (.388)			.390 (.225)		.733 (.593)		
atpeace	1.207 (.727)			1.000		0[c]			1.000 (.637)	
feelblue	1.460 (.817)			–.415[b]		1.964 (1.099)			–.417 (–.247)	
happy	1.212 (.692)			–1.005[b]		2.426 (1.384)			–1.005 (–.607)	
nervous	1.376 (.706)			–.476[b]		1.951 (1.001)			–.476 (–.258)	
down	1.477 (.781)			.251[b]		1.175 (.621)			.251 (.140)	
afraid	1.367 (.653)				1.000 (.574)	0[c]				1.000 (.870)
frust	1.358 (.679)				.730 (.439)	.360 (.180)				.730 (.665)
Hlthwry	1.248 (.595)				.983 (.564)	–.096 (–.046)				.983 (.854)
Cognition						1.000 (.665)				
Vitality						1.059 (.756)				
Mental health						1.207 (1.141)				
Disease worry						1.367 (.751)				

Note. $N = 403$.
[a]Completely standardized solution is included in parentheses. [b]Factor loading was not significant, and the standardized solution could not be computed due to Heywood case. [c]Factor loading of the direct effect was set to *zero* for identification purposes.

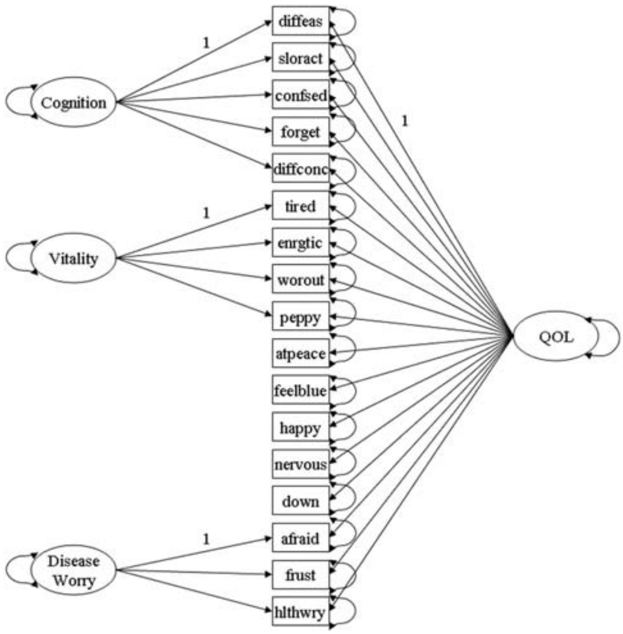

FIGURE 4 An incomplete bifactor model of quality of life. *Note*. In RAM notation, the curved double headed arrows represent factor variances for each of the latent variables and measurement error variance for each of the measured variables. Factor loadings designated with 1 are marker variables; all other factor loadings are estimated.

variables. To illustrate the equivalence between the bifactor and second-order models, direct effects from the second-order factor to the observed variables were added. The model was defined as the following.

1. Each item had a nonzero loading on the first-order factor that it was designed to measure, and zero loadings on the other first-order factors. One of the lower-order factor loadings for each factor was set to 1 and variances of the factors were estimated.
2. The covariance among the first-order factors was explained by a second-order factor. One of the second-order factor loadings for each factor was set to 1, and the variance of the second-order factor was estimated.
3. Of importance and different from the regular second-order model, there was a direct effect from the second-order factor to each item as well; that is, each item had a nonzero loading on the second-order factor. The loadings of items 1 (diffeas), 6 (tired), 10 (atpeace), and 15 (afraid) on the corre-

TABLE 4
Factor Loadings from the Incomplete Bifactor Model and Incomplete Second-Order Factor Model (Study 1)

	Incomplete Bifactor Model (Figure 4)					*Incomplete Second-Order Model (Figure 5)*				
Item/Factor	*General Factor*	*Cognition*	*Vitality*	*Mental Health*	*Disease Worry*	*Second-Order*	*Cognition*	*Vitality*	*Mental Health*	*Disease Worry*
diffeas	1.000 (.555)	1.000 (.646)					1.000 (.874)			
sloract	.824 (.469)	.701 (.465)					.730 (.680)			
confsed	1.078 (.559)	1.040 (.628)					1.012 (.848)			
forget	.839 (.451)	1.064 (.666)					.914 (.797)			
diffconc	1.119 (.581)	1.001 (.605)					1.018 (.853)			
tired	1.072 (.643)		1.000 (.568)					1.000 (.846)		
enrgtic	1.060 (.546)		.803 (.392)					.934 (.679)		
worout	1.081 (.634)		1.026 (.570)					1.014 (.840)		
peppy	1.194 (.676)		.724 (.388)					.987 (.788)		
atpeace	1.296 (.764)					1.000 (.725)				
feelblue	1.489 (.816)					1.265 (.818)				
happy	1.253 (.701)					1.045 (.690)				
nervous	1.388 (.698)					1.187 (.705)				
down	1.504 (.779)					1.281 (.784)				
afraid	1.389 (.650)				1.000 (.579)					1.000 (.860)
frust	1.379 (.676)				.733 (.444)					.914 (.823)
hlthwry	1.271 (.594)				.977 (.564)					.938 (.805)
Cognition						1.000 (.707)				
Vitality						.935 (.783)				
Disease worry						1.199 (.771)				

Note. $N = 403$. Parenthetical values are completely standardized solution.

sponding first-order factors were set to 0 for identification purposes (see caption to Figure 1 for descriptions of items).
4. Error terms associated with each item were uncorrelated.

The fit statistics are presented in Table 2. The unstandardized and standardized factor loadings are presented in Table 3.

As can be seen in Table 2, the fit statistics were exactly the same as in the bifactor model, consistent with the expected mathematical equivalence of the full second-order model. Further, Table 3 shows that the unstandardized factor loadings of the domain specific factors in the bifactor factor were identical to the loadings for the lower-order factors in the full second-order model, although the standardized solutions differ. The relationships between the standardized factor loadings in the bifactor model and the second-order factor model are more complex. Following the development in Yung et al. (1999), these relationships can be specified as follows.

Let $\lambda second$ be the standardized factor loading of the second-order factor, $\lambda first$ be the standardized factor loadings of the first-order factors, and $\lambda direct$ be the standardized factor loading of the direct effect from the second-order factor in the full second-order model.

Let $\lambda general$ be the standardized factor loading of the general factor, and $\lambda specific$ be the standardized factor loading of the domain specific factor in the bifactor model.

The standardized second-order factor loadings can be solved uniquely using the factor loadings of the general factor and domain specific factor of the variables that have been fixed to 0, that is, "diffeas," "tired," "atpeace," and "afraid," respectively:

$$\lambda second = sign(\lambda specific)\sqrt{\frac{\lambda^2 general}{\lambda^2 specific + \lambda general^2}},$$

sign ($\lambda specific$) is +1 if $\lambda specific$ is positive and –1 if $\lambda specific$ is negative.

$$\lambda second1 = sign(.635)\sqrt{\frac{.566^2}{.635^2 + .566^2}} = .665,$$

where .566 and .635 are the general and domain specific factor loadings of item "diffeas."

$$\lambda second2 = sign(.562)\sqrt{\frac{.649^2}{.562^2 + .649^2}} = .756,$$

where .649 and .562 are the general and domain specific factor loadings of item "tired."

$$\lambda second4 = sign(.574)\sqrt{\frac{.653^2}{.574^2+.653^2}} = .751,$$

where .653 and .574 are the general and domain specific factor loadings of item "afraid."

$\lambda second3$ could not be solved because the domain specific factor loading of item "atpeace" was not available due to model misspecification.

After solving the second-order factor loadings, factor loadings of the direct effects can be solved using the following equation:

$$\lambda direct = \lambda general - \frac{(\lambda specific)(\lambda Second)}{\sqrt{1-\lambda^2 Second}}.$$

For example, for item "sloract,"

$$\lambda direct = .476 - \frac{(.458)(.665)}{\sqrt{1-.665^2}} = .068.$$

Using the factor loadings of direct effect and second-order factors, first-order factor loadings can be solved as follows:

$$\lambda first = \lambda specific \frac{1}{\sqrt{1-\lambda^2 Second}}.$$

For example, for item "sloract,"

$$\lambda first = .458 \frac{1}{\sqrt{1-.665^2}} = .613.$$

Testing the standard second-order factor model (Figure 2). The standard second-order model was specified in the following way: (a) each item had a nonzero loading on the first-order factor (cognition, vitality, mental health, disease worry) that it was designed to measure and a zero loading on each of the other first-order factors; (b) error terms associated with each item were uncorrelated; and (c) all covariance between each pair of the first-order factors was explained by a higher-order factor—global health-related quality of life. To identify the model, in addition to setting one of the factor loadings in the second-order factor to 1, one of the loadings in each of the lower-order factors was also set to 1. The variance of the second-order factor was estimated.

As can be seen from Table 2, the $\chi^2(115) = 279.35$, $p < .001$; RMSEA was .060; SRMR was .040; and CFI was .961, indicating an adequate fit of the data. The standardized disturbance of mental health factor was *significant*, but the size was very small, .086, $p < .05$, indicating that 91.4% of variance of the mental health factor was explained by the second-order factor.

Testing the incomplete second-order factor model (Figure 5). As discussed earlier, the standard second-order model (Figure 2) is a reduced form of the full second-order model (Figure 3), which is equivalent to the bifactor model (Figure 1). Therefore, the standard second-order model (Figure 2) is nested within the bifactor model (Figure 1). Yung et al. (1999) have further demonstrated that this relationship can be applied to incomplete bifactor and second-order models. To nest a second-order model within the incomplete bifactor model (Figure 4), an incomplete second-order model (Figure 5) was tested. That is, the full second-order factor model (Figure 3) was modified by removing the mental health first-order factor and, further, removing the direct effects from the second-order factor to the measured items that are associated with mental health factor. The factor loadings are presented in Table 4. As can be seen from Table 2, the $\chi^2(117) = 292.62$, $p < .001$; RMSEA was .061; SRMR was .053; and CFI was .959, indicating an adequate fit of the data. The chi-square difference test between the incomplete second-order model (Figure 5) and the incomplete bifactor model (Figure 4) was significant, $\Delta\chi^2 = 54.44$ ($\Delta df = 10$), $p < .001$, indicating that incomplete bifactor model (Figure 4) fit the data better than the incomplete second-order factor model (Figure 5).

If we assume that the incomplete bifactor model is the true model, the power to differentiate the incomplete bifactor and incomplete second-order models may be calculated using procedures developed by Satorra and Saris (1985). (The SAS program used to calculate the power for the chi-square difference test between the incomplete bifactor and second-order models at the .05 alpha level is given in the Appendix.) Given our sample size of 403, the power to differentiate the two models exceeded .99. Of note, the difference in RMSEAs between the two models was .003, indicating that a relatively small discrepancy could be detected.

Prediction of an External Criterion

Researchers may wish to answer the question of whether one (or more) of the domain specific factors predicts an external criterion over and above the general factor. Researchers may address this question so long as they do not include all domain specific factors (here, four domain specific factors) and the general factor

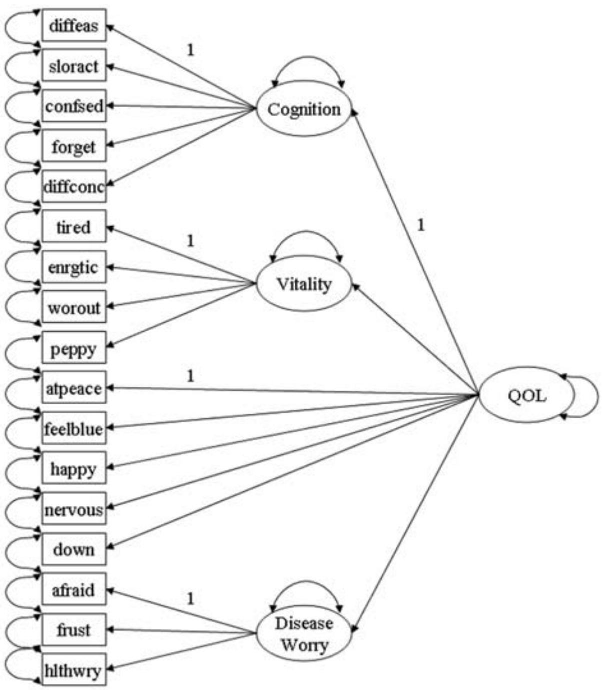

FIGURE 5 An incomplete second-order model of quality of life. *Note*. In RAM notation, the curved double headed arrow represents the factor variance for the second-order factor (QOL), disturbance variance for each of the first-order factors and measurement error variance for each of the measured variables. Factor loadings designated with 1 are marker variables; all other factor loadings are estimated.

simultaneously in the model, a specification that introduces a linear dependency among the predictors.

Analysis using the conditional incomplete bifactor model (Figure 6). We tested whether the three domain specific factors in the conditional bifactor model would predict social functioning over and above the general quality of life factor. Each of the three domain specific factors and the general factor were used to predict social functioning simultaneously. The fit statistics are presented in Table 5. The χ^2 (120) = 280.47, $p < .001$; CFI was .963, indicating an adequate fit of the data. As can be seen from Table 6, the general quality of life factor predicted social functioning, standardized coefficient $\beta = .42$, $z = 7.35$, $p < .001$. Of the three domain specific factors, "vitality" predicted social functioning over and above the general factor, $\beta = .35$, $z = 6.13$, $p < .001$, "disease worry" also predicted social

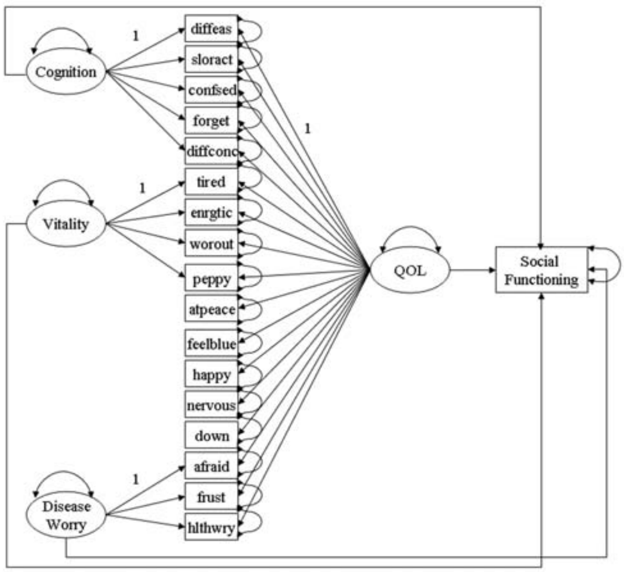

FIGURE 6 Predicting an external criterion: A conditional incomplete bifactor model of quality of life. *Note.* In RAM notation, the curved double headed arrows represent factor variances for each of the latent variables and measurement error variance for each of the measured variables. Factor loadings designated with 1 are marker variables; all other factor loadings and paths are estimated.

functioning, $\beta = .13$, $z = 2.41$, $p < .01$, whereas "cognition" failed to reach statistical significance, $\beta = .08$, $z = 1.73$, *ns*. Note that disease worry was reverse scored so that high scores indicate low levels of disease worry, accounting for the positive sign of the path coefficient.

Analysis using the conditional incomplete second-order factor model (Figure 7). The Bentler and Weeks (1980) representation implemented in EQS (Bentler, 1995) permits similar tests based on second-order models.[7] Once again, we tested whether the domain specific factors would predict social functioning over and above the second-order factor quality of life using the incomplete second-order model. The disturbances of the lower-order factors "vitality," "cogni-

[7]The conditional second-order model is a nonstandard model (Bentler, 1990) in which disturbances are used to predict external variables. As with the bifactor model, specification of a model that includes all four disturbances of the first order factors as well as the general factor produces a linear dependency among the predictors.

TABLE 5
Summary of Fit Statistics for Conditional Bifactor and Second-Order Models of Quality of Life (Outcome on Bifactor/Second and Three Domain Specific Factors) (Study 1)

						Second vs. Bifactor		
	x^2	*df*	*RMSEA*	*SRMR*	*CFI*	$\Delta\chi^2$	Δdf	*Power*
Conditional bifactor model (Figure 6)	280.47*	120	NA	NA	.963			
Conditional second-order model (Figure 7)	316.48*	130	NA	NA	.958	36.01*	10	>.99

Note. $N = 403$.
$^*p < .001$.

TABLE 6
Regression Coefficients for the Conditional Incomplete Bifactor and Second-Order Models of Quality of Life (Study 1)

Factor (N = 403)	*Bifactor Factor*	*Second-Order Factor*
General/Second-order	.73** (.42)	.62** (.42)
Cognition	.13 (.08)	.14 (.09)
Vitality	.65** (.35)	.75** (.37)
Disease worry	.19* (.13)	.23* (.15)

Note. Standardized coefficients are in parentheses. $N = 403$.
$^*p < .01$. $^{**}p < .001$.

tion," and "disease worry," which correspond to the domain specific factors from the bifactor model, and the second-order factor, which corresponds to the general factor in the bifactor model, were used to predict social functioning using the EQS 5.7 program. The fit statistics are presented in Table 5. The $\chi^2(130) = 316.48$, $p < .001$; CFI was .958, indicating an adequate fit of the data.

As can be seen from Table 6, the estimates from the second-order model are very similar to those from the bifactor model. The second-order factor predicted social functioning, $\beta = .42$ (vs. .42 in the bifactor model), $z = 8.27$, $p < .001$. In addition, social functioning was predicted by the disturbances of the first order factors of "vitality," $\beta = .37$ (vs. .35 in the bifactor model), $z = 6.59$, $p < .001$ and "disease worry," $\beta = .15$ (vs. .13 in the bifactor model), $z = 2.76$, $p < .01$, whereas the disturbance of the first order factor "cognition" did not predict social functioning, $\beta = .09$ (vs. .08 in the bifactor model), $z = 1.89$, *ns*.

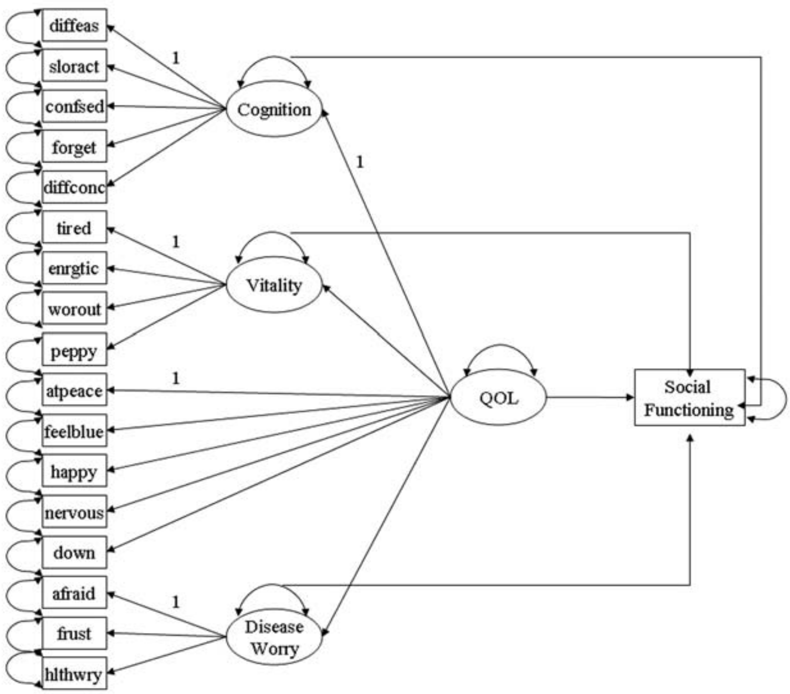

FIGURE 7 Predicting an external criterion: A conditional incomplete second-order model of quality of life. *Note*. In RAM notation, the curved double headed arrow represents the factor variance for the second-order factor (QOL), disturbance variance for each of the first-order factors and measurement error variance for each of the measured variables. Factor loadings designated with 1 are marker variables; all other factor loadings and paths are estimated.

DISCUSSION

The present study compared two alternative approaches to the representation of the factor structure of instruments when it is hypothesized that a general construct is comprised of several highly related domains. Using a quality of life instrument with multiple items representing each domain of the construct, this study indicated that the bifactor model had several advantages over the second-order model when researchers are interested in both the domain specific factors and the general factor. There are five major conclusions from our comparison of the bifactor and second-order models using a quality of life measure.

1. For the hypothesized bifactor model, there was a mathematically equivalent second-order factor model with direct effects from the second-order factor to the

observed variables. This result provides an empirical illustration for the equivalence of the two models demonstrated analytically by Yung et al. (1999).

2. Based on the hypothesized full bifactor model, there was one general quality of life factor with four domain specific factors—cognition, vitality, mental health, and disease worry. The bifactor analysis established that there were only three, rather than four, domain specific factors over and above the general factor. The hypothesized mental health factor did not exist as a domain specific factor over and above the general factor. The hypothesized full bifactor model was then modified by removing the mental health domain specific factor, yielding an incomplete bifactor factor model. In contrast, the second-order model did not provide clear evidence of the collapse of the mental health domain specific factor.

3. A corresponding incomplete second-order model fit the data significantly worse than the incomplete bifactor model, indicating that the constraints in the second-order model were too strict. Given that the incomplete second-order model was nested within the incomplete bifactor model (Yung et al., 1999), the bifactor model was used as a baseline model against which the second-order model was compared.

4. When estimating the predictive relationships between the domain specific factors/disturbances and an external variable, over and above the general/second-order factor, the incomplete bifactor model and second-order model gave similar parameter estimates. However, it is easier for substantive researchers to interpret the results from the bifactor model, as the domain specific factors are represented by common factors, rather than disturbances.

5. In contrast to Mulaik and Quartetti's (1997) results, there was sufficient power to differentiate the bifactor model from the corresponding second-order model with a sample size of less than 500. There are three possible reasons for this discrepancy. First, as noted earlier, the covariance matrix used in Mulaik and Quartetti's study imposed proportionality constraints on the factor loadings of the general and domain specific factors. Given that the proportionality constraints were only slightly perturbed, the discrepancy between the bifactor and second-order models was trivially small. Second, in Mulaik and Quartetti's study, power was calculated for the omnibus test of the absolute fit of the second-order model, rather than for the focused test comparing the fit of the two nested models. When power was recalculated comparing the two nested models, power was substantially increased in that study; or, third, the high power in the present study may be an artifact of the collapse of mental health factor. The use of an actual data set rather than a simulated data set in which the true model is known precluded definite conclusions. Thus the present findings raise issues about Mulaik and Quartetti's conclusion, but the possibility that the present results could be an artifact together with the use of post hoc power analysis (see Lenth, 2001, for a critique on this practice) requires a clearer demonstration.

STUDY 2

A small Monte Carlo study was conducted to address two issues: (1) to examine further the power to differentiate bifactor models from second-order models, and (2) to provide a preliminary assessment of the power to distinguish bifactor from second-order models. The use of a known population covariance matrix assured that (a) the full bifactor model in which each of the domain specific factors had variance substantially greater than 0 would be tested, and (b) empirical power estimates rather than post hoc power calculations would be provided.

We used the Monte Carlo feature in Version 3.01 of Mplus (Muthén & Muthén, 1998) to generate the data and Mplus's maximum likelihood (ML) estimation to estimate the model. This method permitted us to fit a model that differed in structure from the model that generated the data (see Muthén & Muthén, 1998, for further details about Mplus Monte Carlo procedures). Following the procedures recommended by Satorra and Saris (1985), the data generation and estimation procedure was comprised of four basic steps.

1. Raw data were randomly generated from a multivariate normal distribution to correspond to the parameterization of a hypothesized *complete bifactor model* (see Figure 8). The hypothesized model was created with all the parameters fixed to the specified values. The parameters selected for the hypothesized bifactor model are similar to the ones presented in Table 3, except that two major changes were made: first, factor loadings for the domain specific factor "mental health" were fixed to be positive and they ranged from .681 to 1.000; and second, the variance of the "mental health" factor was fixed to .257, which is comparable to the variance of other domain specific factors. These changes would be expected to increase the discrepancy between the bifactor and second-order models relative to Study 1. The hypothesized factor loadings, variances, and residual variances are presented in Table 7, and the population covariance matrix in Table 8.
2. The hypothesized bifactor model was fitted to the randomly generated raw data, and the fit statistics were used as baseline fit information.
3. A second-order model was fit to the raw data generated from the bifactor model.
4. The chi-square difference test statistic was used to compare the fit of the two models.

We used a sample size of 200, which is consistent with typical current practice in psychology. Jaccard and Wan (1995) reported that the median sample size of studies using multiple regression analyses reported in American Psychological Association journals in the early 1990s was 175. MacCallum (personal communication, September 22, 2004) reported that of the 109 studies considered in a review of studies reporting structural equation models (MacCallum & Austin, 2000), 50 had sample sizes less than or equal to 200 and 59 had sample sizes of greater than 200.

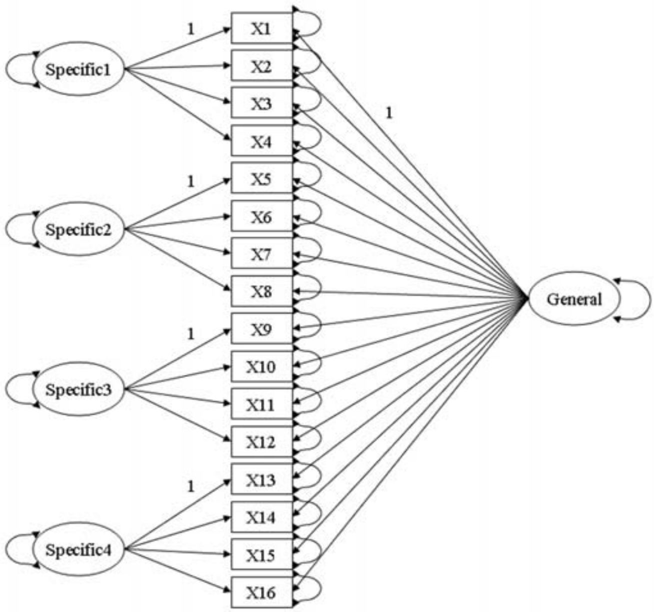

FIGURE 8 A bifactor model. *Note*. In RAM notation, the curved double headed arrows represent factor variances for each of the latent variables and measurement error variance for each of the measured variables. Factor loadings designated with 1 are marker variables; all other factor loadings are estimated.

Data Characteristics

Distribution. Data were generated from a multivariate normal distribution.

Replications. A total of 500 replications that yielded proper solutions were generated for each sample. One replication did not converge and two yielded improper solutions. These replications were replaced.

Baseline Bifactor Model

The baseline model (see Figure 8) was a complete bifactor model with four domain specific factors, and the corresponding second-order model has four lower-order factors (Figure 9). The results of the fit statistics are reported in Table 9. When the bifactor model was fitted to the raw data, the fit was acceptable. Across the 500 replications for $n = 200$, the mean values of the fit statistics were $\chi^2 = 91.52$ ($df = 88$), *ns*; RMSEA = .014; SRMR = .028; and CFI = .998.

TABLE 7
Hypothetical Values for the Complete Bifactor Model (Study 2)

	Unstandardized Factor Loadings					
Item/Factor	*General Factor*	*Domain Specific Factor 1*	*Domain Specific Factor 2*	*Domain Specific Factor 3*	*Domain Specific Factor 4*	*Variance/Residual Variance*
V1	1.00	1.00				.179
V2	.86	.75				.316
V3	1.20	1.03				.362
V4	.90	1.06				.250
V5	1.16		1.00			.272
V6	1.05		1.04			.220
V7	1.09		.76			.352
V8	1.05		1.03			.298
V9	.99			1.00		.447
V10	1.55			.80		.160
V11	1.27			.68		.239
V12	1.26			1.08		.452
V13	1.55				1.00	.325
V14	1.42				.76	.347
V15	1.25				.64	.461
V16	1.15				.93	.297
General factor						.452
Domain Specific Factor 1						.263
Domain Specific Factor 2						2.196
Domain Specific Factor 3						.257
Domain Specific Factor 4						.326

Given a large sample size, the expected value of the χ^2 statistic for a properly specified model should be equal to the degrees of freedom. In contrast, given the modest sample size ($n = 200$) used in the simulation, the average χ^2 statistic would be higher than its expected value (see Hu, Bentler, & Kano, 1992). All parameter estimates in the model, that is, factor loadings, variances/covariances, residual variances, closely matched the population values with no appreciable bias.

Bifactor Model Versus Second-Order Model

For each of the 500 replications, a second-order model (see Figure 9) was also fitted to the raw data generated by the baseline bifactor model. As can be seen from Table 9, with a sample size of 200, the mean fit statistics were $\chi^2(100) = 138.69$; RMSEA was .043; SRMR was .043; and CFI was .988. The mean chi-square difference test between the two models was significant, $\Delta\chi^2 = 47.18$ ($\Delta df = 12$), $p < .001$. The average power to detect the difference exceeded .99.

TABLE 8
Population Covariance Matrix (Study 2)

*x*1	0.894															
*x*2	0.586	0.798														
*x*3	0.813	0.670	1.292													
*x*4	0.686	0.559	0.775	0.912												
*x*5	0.524	0.451	0.629	0.472	3.076											
*x*6	0.475	0.408	0.570	0.427	2.834	3.094										
*x*7	0.493	0.424	0.591	0.443	2.240	2.253	2.157									
*x*8	0.475	0.408	0.570	0.427	2.812	2.851	2.236	3.126								
*x*9	0.447	0.385	0.537	0.403	0.519	0.470	0.488	0.470	1.147							
*x*10	0.701	0.603	0.841	0.631	0.813	0.736	0.764	0.736	0.899	1.410						
*x*11	0.574	0.494	0.689	0.517	0.666	0.603	0.626	0.603	0.743	1.030	1.087					
*x*12	0.570	0.490	0.683	0.513	0.661	0.598	0.621	0.598	0.841	1.105	0.912	1.469				
*x*13	0.701	0.603	0.841	0.631	0.813	0.736	0.764	0.736	0.694	1.086	0.890	0.883	1.737			
*x*14	0.642	0.552	0.770	0.578	0.745	0.674	0.700	0.674	0.635	0.995	0.815	0.809	1.243	1.447		
*x*15	0.565	0.486	0.678	0.508	0.655	0.593	0.616	0.593	0.559	0.876	0.718	0.712	1.084	0.961	1.301	
*x*16	0.520	0.447	0.624	0.468	0.603	0.546	0.567	0.546	0.515	0.806	0.660	0.655	1.109	0.969	0.844	1.177

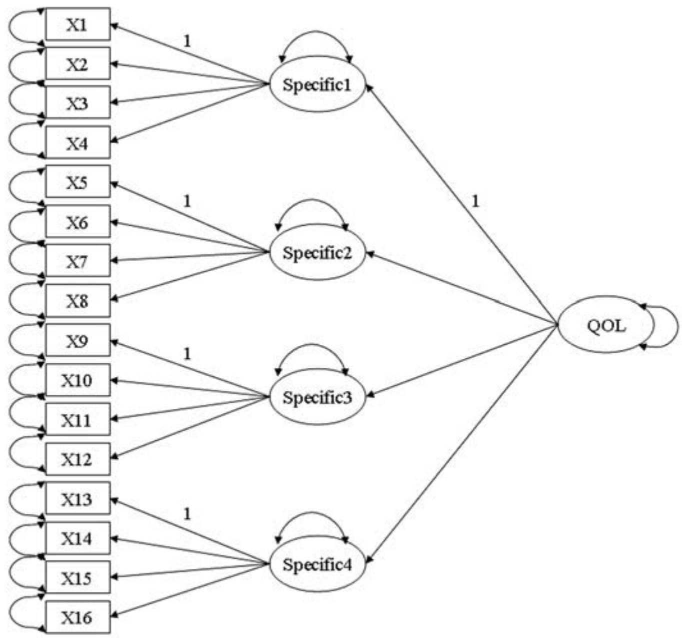

FIGURE 9 A standard second-order model. *Note*. In RAM notation, the curved double headed arrow represents the factor variance for the second-order factor (QOL), disturbance variance for each of the first-order factors, and measurement error variance for each of the measured variables. Factor loadings designated with 1 are marker variables; all other factor loadings are estimated.

Summary

Study 2 replicated the findings of Study 1 regarding statistical power. First, there was sufficient power to detect the difference between the bifactor model and second-order model even with a sample size of 200. Second, the difference between the two models was not an artifact of the collapse of one of the domain specific factors or the use of post hoc power calculations. Indeed, the observed difference between the RMSEAs for the two models (.029) indicates that the inclusion of the fourth domain specific factor lead to a greater discrepancy between the two models, yielding sufficient statistical power to distinguish them even at a modest sample size (n = 200) that typifies current psychological research.

TABLE 9
Summary of Fit Statistics for the Simulated Complete Bifactor and Second-Order Models (Study 2)

						Bifactor vs. Second			
Model	x^2	*df*	*RMSEA*	*SRMR*	*CFI*	$\Delta\chi^2$	Δdf	*Rejecting Probability*[a]	*Power*
Bifactor model	91.52 (13.18)	88	.014 (.014)	.028 (.005)	.998 (.003)				
Second-order model	138.69* (18.12)	100	.043 (.011)	.043 (.005)	.988 (.006)	47.18**	12	90.60%	> .99

Note. Standard deviations are in parentheses. $N = 200$. RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; CFI = comparative fit index.
[a]Probability of rejecting the second-order model out of the 500 replications.
$*p < .01$. $**p < .001$.

GENERAL DISCUSSION

The present article compared the results of a bifactor model and second-order factor model using a health-related quality of life instrument with HIV patients. Consistent with theoretical expectation, the bifactor model had several advantages over the second-order model. First, the bifactor model was able to identify that there were only three, rather than four, domain specific factors, over and above the general factor. Second, the bifactor model fit the data significantly better than the corresponding second-order model, indicating that the constraints on the second-order model were too strict. Third, when domain specific factors/disturbances are used to predict an external variable, over and above the general/second-order factor, it is easier for substantive researchers to interpret the results from the bifactor model, which represents the domain specific factors as common factors rather than disturbances. These findings have important implications for psychological research, as researchers may be interested in the predictive validity of the general factor as well as domain specific factors over and above the common underlying factor.

Yung et al.'s (1999) demonstration that the second-order models are nested within corresponding bifactor models made it possible to directly compare the two models. We showed in Studies 1 and 2 that there was sufficient power to distinguish the two models even with a sample size of less than 500. Study 1 compared an incomplete bifactor model with an incomplete second-order model. The results were based on the empirical example of the quality of life instrument. Given our sample size ($n = 403$), the power exceeded .99, clearly sufficient to differentiate the two models. In contrast, Study 2 simulated data for a complete bifactor model and compared it with the corresponding second-order model. Even with a sample size of 200, the power to differentiate the two models in our example exceeded .99.

Despite these positive characteristics of the bifactor model, its limitations should also be noted. We considered here the canonical version of the bifactor model in which the general factor and each of the domain specific factors are assumed to be orthogonal. An attractive feature of this version is that the results are very easy to interpret, as we have shown. Each parameter in the model is uniquely estimable so that theoretically there should not be problems with identification. As we move away from the orthogonal version of the model and allow covariances between factors, problems of identification become more likely. For example, Mulaik and Quartetti (1997) found that when the general factor is allowed to covary with the domain specific factors, the model will not converge. Rindskopf and Rose (1988) found evidence of identification problems when they allowed the domain specific factors to covary. Similar identification issues also affect second-order models in which assumptions of orthogonality are relaxed. However, these identification issues are less apparent because the orthogonality constraints are included in the standard basic assumptions of confirmatory factor analysis. Paralleling the bifactor model assumption that the general and the domain specific factors are uncorrelated, one assumption of the second-order model is that the second-order factor and the disturbances of the first order factor are uncorrelated. If researchers have an interest in investigating models in which non-orthogonal relationships between general and domain specific factors are permitted, variables outside of the measurement model that are known to predict only one of the factors (instrumental variables) may be added to help assure identification. In addition, external criterion variables that are known to be uniquely predicted by one of the factors can also help identify the model. These general strategies are discussed in detail by Graham and Collins (1992). The success of such techniques that "borrow strength" to achieve identification depend strongly on the correct specification of the relationships between the external variables and those of the bifactor model.

The context of our motivating example has several characteristics that make consideration of bifactor and second-order models attractive. Researchers (e.g., Lubeck & Fries, 1993; McHorney, Ware, & Raczek, 1993, Stewart & Ware, 1992) hypothesized a general factor that is comprised of several specific domains. Their central interest was in the effect of the general factor; their secondary interest was in the effects of the domain specific factors. Previous studies had been conducted to support the hypothesized conception. The pool of items representing the domains had been refined so that substantive interest is now focused on the standard, fixed set of items that comprise the scale rather than a hypothetical population of items. Under these conditions, models need to capture the general factor that the measure was designed to assess. Alternative approaches such as the exploratory factor analysis with oblique rotation to simple structure or first order confirmatory factor models in which correlations are estimated between the domain-specific factors (group-factor model, Rindskopf & Rose, 1988) do not capture this focus. Although these alternative approaches are highly useful in many applications, in the present context they tend to yield solutions with highly correlated domain-specific factors and miss the general factor that underlies the data. Under such circumstances, it is possible to have a set of domain-specific factors that collectively account for a substantial portion of the variance in an external criterion but which individually fail to lead to substantial prediction of the criterion due to multicollinearity. In contrast, bifactor models and second-order models maintain the focus of the analysis on the constructs hypothesized by the developers of the scale, and can more adequately capture the relationship between the general factor and external criteria.

Finally, we do not wish to imply that bifactor models are more applicable than second-order models under all conditions. If the general factor is the main focus of the research, the second-order factor model may be more parsimonious, given that the second-order model fits the data equally well as the bifactor model. Moreover, the bifactor and second-order representations are not mutually exclusive, and they can coexist in different parts of the same complex model. Eid, Litschetzke, Nussbeck, and Trierweiler (2003) provided some examples in their representations of second-order multitrait multimethod models.

REFERENCES

Bentler, P. M. (1990). Latent variable structural models for separating specific from general effects. In L. Sechrest, E. Perrin, & J. Bunker (Eds.), *Research methodology: Strengthening causal interpretations of nonexperimental data* (pp. 61–80). U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research.

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

Bentler, P. M., & Weeks, D. G. (1980). Linear structural equations with latent variables. *Psychometrika*, *45*, 289–308.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.). *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Costa, P. T., & McCrae, R. R. (1992). *NEO PI–R professional manual*. Odessa, FL : Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment*, *64*, 21–50.

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences*, *33*, 533–552.

Eid, M., Lischetzke, T., Nussbeck, F., & Trierweiler, L. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M-1) Model. *Psychological Methods*, *8*, 38–60.

Graham, J. W., & Collins, N. L. (1992). Controlling correlational bias via confirmatory factor analysis of MTMM data. *Multivariate Behavioral Research*, *26*, 607–629.

Gustafsson, J. (1992). The relevance of factor analysis for the study of group differences. *Multivariate Behavioral Research*, *27*, 239-247.

Gustafsson, J., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, *28*, 407–434.

Hays, R. D., & Stewart, A. L. (1990). The structure of self-reported health in chronic disease patients. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *2*, 22–30.

Hills, P., & Argyle, M. (2002). The Oxford Happiness Questionnaire: A compact scale for the measurement of psychological well-being. *Personality and Individual Differences*, *33*, 1071–1082.

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, *2*, 41–54.

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

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

Jaccard, J., & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin*, *117*, 348–357.

Jackson, D. N., Ahmed, S. A., & Heapy, N. A. (1976). Is achievement a unitary construct? *Journal of Research in Personality*, *10*, 1–21.

Jöreskog, K. G. (1979). A general approach to confirmatory maximum likelihood factor analysis, with addendum. In K. G. Jöreskog & D. Sörbom (J. Magidson, Ed.), *Advances in factor analysis and structural equation models* (pp. 21–43). Cambridge, MA: Abt Books.

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct? *Journal of Personality and Social Psychology*, *83*, 693–710.

Lenth, R. V. (2001). Some practical guidelines for effective sample size determination. *American Statistician*, *55*, 187–193.

Lubeck, D. P., & Fries, J. F. (1993). Health status among persons infected with human immunodeficiency virus. *Medical Care*, *31*, 269–276.

Luo, D., Petrill, S. A., & Thompson, L. A. (1994). An exploration of genetic g: Hierarchical factor analysis of cognitive data from the western reserve twin project. *Intelligence*, *18*, 335–347.

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

MacCallum, R. C., Browne, M. W., & Cai, L. (2006). Testing differences between nested covariance structure models: Power analysis and null hypotheses. *Psychological Methods, 51,* 201–226.

Marsh, H. W., Ellis, L. A., & Craven, R. G. (2002). How do preschool children feel about themselves? Unraveling measurement and multidimensional self-concept structure. *Developmental Psychology*, *38*, 376–393.

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

McHorney, C. A., Ware, J. E., & Raczek, A. E. (1993). The MOS 36-Item Short-Form Health Survey (SF-36): II Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care*, *31*, 247–263.

Mulaik, S. A., & Millsap, R. E. (2000). Doing the four-step right. *Structural Equation Modeling*, *7*, 36–73.

Mulaik, S. A., & Quartetti, D. A. (1997). First order or higher order general factor? *Structural Equation Modeling*, *4*, 193–211.

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

Rindskopf, D. (1984). Structural equation models: Empirical identification, Heywood cases, and related problems. *Sociological Methods and Research*, *13*, 109–119.

Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research*, *23*, 51–67.

Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, *50*, 83–90.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *5*, 141–152.
Sherbourne, C. D. (1992). Social functioning: Social activity limitations measure. In A. L. Stewart & J. E. Ware (Eds.), *Measuring functioning and well-being: The medical outcome study approach* (pp. 173–181). Durham, NC: Duke University Press.
Spearman, C. (1927). *The abilities of man*. London: MacMillan.
Stewart, A. L., & Ware, J. E. (1992). *Measuring functioning and well-being: The medical outcome study approach*. Durham, NC: Duke University Press.
Ware, J. E., Davies-Avery, A., & Brook, R. H. (1980). *Conceptualization and measurement of health for adults in the Health Insurance Study: Vol. 6. Analysis of relationships among health status measures*. Santa Monica, CA: RAND Corp.
West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56–75). Thousand Oaks, CA: Sage.
Wherry, R. J. (1959). Hierarchical factor solutions without rotation. *Psychometrika*, *24*, 45–51.
Yung, Y. F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika*, *64*, 113–128.

Accepted June 2005

APPENDIX

The SAS program for power calculation.

```
/*Title 'Power Analysis for the Incomplete Bifactor and Second-Order Models'*/
DATA ONE;
DF=10; CRIT=18.3070;    /*critical value at alpha equals .05 for df = 10*/
LAMBDA=40.52;           /*chi-square difference between the two models*/
POWER=(1-(PROBCHI(CRIT,DF,LAMBDA)));
PROC PRINT DATA=ONE;
RUN;
```