Modeling General and Specific Variance in Multifaceted Constructs: A Comparison of the Bifactor Model to Other Approaches

Fang Fang Chen,1 Adele Hayes,1 Charles S. Carver,2 Jean-Philippe Laurenceau,1 and Zugui Zhang3

1University of Delaware
2University of Miami
3Christiana Care Health System

ABSTRACT This article recommends an alternative method for testing multifaceted constructs. Researchers often have to choose between two problematic approaches for analyzing multifaceted constructs: the total score approach and the individual score approach. Both approaches can result in conceptual ambiguity. The proposed bifactor model assesses simultaneously the general construct shared by the facets and the specific facets, over and above the general construct. We illustrate the bifactor model by examining the construct of Extraversion as measured by the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992), with two college samples (N = 383 and 378). The analysis reveals that the facets of the NEO-PI-R Extraversion correlate with criteria in opposite directions after partialling out the general construct. The direction of gender differences also varies by facets. Bifactor models combine the advantages but avoid the drawbacks of the 2 existing methods and can lead to greater conceptual clarity.

Psychological constructs are often characterized by several related facets. Examples can be found in most areas of psychology, including personality traits, depression, subjective well-being, attention, interpersonal functioning, cognitive flexibility, and health behaviors. There is, however, a long-standing and unresolved debate in personality research on how to measure and test such multifaceted constructs (e.g., Briggs & Cheek, 1988; Carver, 1989; Finch & West,

Correspondence concerning this article should be addressed to Fang Fang Chen, University of Delaware, Department of Psychology, Wolf Hall, Newark, DE 19716. Email: xiyu@psych.udel.edu.

Journal of Personality 80:1, February 2012
© 2012 The Authors
Journal of Personality © 2012, Wiley Periodicals, Inc.
DOI: 10.1111/j.1467-6494.2011.00739.x
researchers working with multifaceted constructs often have to choose between two problematic approaches that can yield very different results. Multifaceted constructs are measured with several subscales, each of which intends to tap one facet of the construct. For example, in the motivational example of this article, Extraversion, as measured by the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992), is composed of six related facets: Warmth, Gregariousness, Assertiveness, Activity, Excitement Seeking, and Positive Emotions. The total score approach involves creating a composite score from the individual facets. The individual score approach involves analyzing separately each facet of the construct. The challenge that researchers face is that each approach sacrifices some useful information. A composite captures mostly the shared effects but does not separate the unique effects from the shared variance. On the other hand, analyzing the facets separately can capture their unique contributions, but the specific effects of the facets are often entangled with the effects of the shared general construct. Both approaches can result in conceptual ambiguity.

Which method is preferable? Researchers have debated this issue for quite some time (e.g., Carver, 1989; Hull et al., 1991), and it is still a dilemma today (e.g., Smith et al., 2009). This article suggests that an alternative approach, termed the bifactor model, can offer the best of both worlds. This approach can have broad applicability across areas of psychology, as it combines the advantages, but avoids the drawbacks, of the two existing methods. We first review the strengths and weaknesses of the total score and individual score approaches. We then introduce the bifactor model, which simultaneously assesses the general effect shared by the facets and the specific effects associated with the facets. For comparison purposes, we also present two alternative approaches: the second-order model, which has often been used to represent multifaceted constructs but, to the best of our knowledge, has not been used to separate the unique and general variance in trait constructs, and the residual
regression method, in which the residual of each facet is used to predict external variables after partialling out other facets. Finally, we provide an empirical application of the bifactor model and contrast these findings with those of the individual and total score approaches, second-order models, and a residual regression method.

**Total Score and Individual Score Approaches**

The total score approach forms a composite by taking the sum or mean of the facets of a multifaceted construct, giving each component equal weight. The composite score is then related to predictors or outcome variables of interest. The total score approach has three primary advantages: (a) It is simple both conceptually and data-analytically; (b) it incorporates more items than any facet, and, as the number of items increases, the reliability of a scale also tends to increase if the additional items do not significantly lower the average interitem correlation (Carmines & Zeller, 1979); and (c) it tends to provide greater content validity than its individual facets, if all of the items are related to the overall construct. That is, the conceptual coverage of the total score is broader than any of its facets, and thus it more adequately captures the complexity of the underlying construct (Carmines & Zeller, 1979).

A significant drawback of the total score approach, however, is that it does not provide information on the relations between the individual facets and the outcome variables. When a total score is used to predict an outcome variable, it is unclear whether the outcome is equally associated with all of the facets, associated with a single facet in all settings, or associated with different facets in different settings. This ambiguity could have important implications. If only some facets of a multifaceted construct predict an observed effect, the total score will yield weaker research findings relative to the individual components because nonpredictive facets are included in the total score. More than two decades ago, Carver (1989) and Hull et al. (1991) warned that inclusion of nonpredictive facets can lead to the development of inappropriate theories, wasted research efforts, and inappropriate prevention and therapy programs aimed at modifying irrelevant facets. In principle, it is even possible that the facets are related to an external variable in opposite directions, resulting in a misleading null effect. The total score approach is still commonly used, yet the drawbacks of this approach are rarely mentioned in the discussion of research findings.
The individual score approach tests the facets of a construct individually by associating each facet with external variables. This strategy appears to compensate for the disadvantages of the total score approach by considering the role of each facet in the prediction of the outcome variable (e.g., Carver, 1989; Hull et al., 1991). However, this apparent advantage also contains a disadvantage. The individual score approach is conceptually ambiguous because it cannot separate the specific or unique contributions of a facet from the effect of the overall construct shared by all interrelated facets. Conceptually, two sources contribute to a significant correlation between a facet and an outcome variable: unique variance associated with the facet and commonality shared by all facets. For some facets, both the unique and common variances are related to the outcome variable. For others, only the common variance is related to the outcome variable. The two sources of contribution can be distinguished only when the unique aspect of a facet is separated from the common aspect shared by all facets. In addition, the individual score approach tests multiple correlations (of each component) with an external variable, thus making it difficult to interpret the results. In an attempt to overcome the shortcomings of the individual approach, some researchers have related the residuals of each facet (i.e., partialling out all other facets) to external measures (Jang, McCrae, Angleitner, Riemann, & Livesley, 1998; McCrae & Costa, 1992, 2008; McCrae et al., 1999). We term this approach the residual regression method. However, this approach does not have the modeling advantages that bifactor and second-order models offer, and it cannot be used to study mean differences in the facets.

Bifactor Model: Simultaneous Testing of Total and Individual Effects

Bifactor models, also known as general-specific models or nested models, represent a third approach that is particularly well suited to

1. The individual facets can also be entered simultaneously as predictors of an outcome variable in a multiple regression analysis. This type of analysis, also referred to as the regression approach (e.g., Hull et al., 1991), is often combined with the individual score approach.

2. 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.
testing multifaceted constructs (Chen, West, & Sousa, 2006). Bifactor models are not familiar to psychologists in many areas because they have been used primarily in the area of intelligence research (e.g., Gustafsson & Balke, 1993; Luo, Petrill, & Thompson, 1994). It is only recently that the bifactor models have been applied to the study of personality (e.g., Bludworth, Tracey, & Glidden-Tracey, 2010; Brouwer, Meijer, Weekers, & Baneke, 2008). A bifactor model hypothesizes that (a) there is a general factor that accounts for the commonality shared by the facets, and (b) there are multiple specific factors, each of which accounts for the unique influence of the specific component over and above the general factor. We use the construct of Extraversion to illustrate how the bifactor model can be applied to a multifaceted construct. Costa and McCrae (1992) proposed that Extraversion is composed of several related facets (e.g., Warmth, Gregariousness, Assertiveness, Activity, Excitement Seeking, and Positive Emotions). The six facets contribute to an overarching Extraversion factor. Each of the six facets also accounts for unique variance, stemming from its own separate set of facet-related items (see Figure 1). We consider the canonical bifactor model in which the relations among the general and specific facets are assumed to be orthogonal, given that the specific facets account for the variance over and above the general factor.

We argue that the bifactor model has wide applicability in that it provides richer and conceptually less ambiguous information than the total or individual score approach. As with the total score approach, the bifactor model estimates a general latent variable with greater content validity than any of its facets. As with the individual score approach, the bifactor model tests the unique contributions of the facets. However, it overcomes the major drawback of the individual score approach in that it partials out the commonality among the facets when testing the unique association between each facet and an external variable. More specifically, the bifactor model has two central advantages. First, it simultaneously tests the association of an outcome variable with the general latent factor and the unique

3. The benefits of the canonical version of the bifactor model are that only this kind of bifactor model allows us to address the conceptual issues raised in this article, and that the results are easy to interpret. However, if we allow covariances between the general and specific factors, problems of identification will be more likely to occur and the model will not converge. For more detailed discussion of this issue, see Chen et al. (2006).
contributions of the specific factors that are distinct from the general construct. Second, the bifactor model can be used to identify a facet that may no longer remain a unique contributor, after taking into account the common variance shared with other facets. For example, drawing on Spearman's (1927) conceptual framework on intelligence, suppose that variance in verbal ability were entirely redundant with that of general intelligence, whereas spatial, mathematical, and analytic abilities remained as specific factors, even after partialling
out general intelligence. In this case, verbal ability would no longer be a specific factor in the final bifactor model. Together, the advantages of the bifactor model can lead to greater conceptual clarity than the total or individual score approach. The bifactor model clearly distinguishes the variances explained by the common factor and the specific factors, whereas the other two approaches do not make such distinctions. Before testing the proposed bifactor models in our studies, it is important to discuss a related approach, second-order models.

**Second-Order Model**

Hull et al. (1991) recommended second-order models as an alternative approach for testing multifaceted constructs. A second-order model hypothesizes that there is a higher order factor that accounts for the commonality shared by the facets, which consists of the lower order factors. As the bifactor models, second-order models are suitable when measurement instruments assess several related facets. Second-order models are more familiar than bifactor models, as they have been applied in a wider variety of research areas, such as personality (Judge, Erez, Bono, & Thoresen, 2002), self-concept (Marsh, Ellis, & Craven, 2002), and psychological well-being (Hills & Argyle, 2002).

There are several reasons to recommend bifactor models over second-order models for testing multifaceted constructs because bifactor models are more suitable for answering the questions posed in this article. Chen et al. (2006) provided detailed descriptions and examples of the similarities and differences between the bifactor and second-order models, and we will briefly introduce the major advantages of bifactor models. First, only bifactor models can separate the specific factors from the general factor. The strength of the relation between the specific factors in bifactor models and their associated items is reflected in the factor loadings. These relations cannot be tested in the second-order model, as the specific factors are represented by the unique variances (i.e., disturbances) of the first-order factors. Second, only bifactor models can identify whether a facet still exists after partialling out the general factor. Third, only bifactor models can test mean differences on facets over and above the general factor. In contrast, in second-order models, only the second-order latent means can be directly compared, as the specific factors are represented by disturbances. Fourth, the bifactor model is
more applicable when testing whether a subset of the specific factors predicts external variables over and above the general factor, as the specific factors are directly represented as independent factors rather than disturbances, as in second-order models. In second-order models, specific factors (i.e., disturbances) may also be used to predict external criteria, over and above the second-order factor, but such tests may require the use of nonstandard structural equation models. 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 (SEM) software packages, such as Mplus and LISREL. The results from such nonstandard models may be difficult to explain to substantive researchers who are more familiar with the ideal that latent factors, not disturbances, can be used as predictors in SEM. Finally, when a construct is composed of two facets, only bifactor models can be applied. Second-order models will not be identified because they require at least three facets.  

Application of the Bifactor Model to the Analysis of Extraversion

To illustrate the application of the bifactor model for testing multifaceted constructs in psychological research, we present two studies on Extraversion. We analyzed and compared the data using the total score, individual score, residual regression, second-order model, and bifactor model approaches. Extraversion is one of the major five personality factors and is linked to important psychological consequences. Extraversion is widely treated as a multifaceted construct. For example, in the NEO-PI-R Extraversion incorporates six facets: Warmth, Gregariousness, Assertiveness, Activity, Excitement Seeking, and Positive Emotions. The multifaceted nature of the Extraversion scale of the NEO-PI-R makes it well suited for illustrating the application of the bifactor model to study multidimensional constructs.

To demonstrate the applicability of bifactor models for testing multifaceted constructs, this investigation has three goals: (a) to
develop a bifactor model for Extraversion; for comparison purposes, a second-order model and residual regression model are also tested for Extraversion; (b) to test the unique contributions of the general factor and the specific factors of Extraversion in relation to external variables; and (c) to compare latent mean differences in gender on the general and specific factors of Extraversion and consider the possibility that the direction of gender differences may vary across general and specific factor means.

Relations of Extraversion to External Variables

One of the robust findings in personality psychology is that Extraversion is related to subjective well-being, such as positive affect, and the behavioral approach tendency (for reviews, see Ozer & Benet-Martínez, 2006; Steel, Schmidt, & Shultz, 2008). However, these relations have been studied mostly at the general trait level of Extraversion. For scales with a multifaceted structure, such as measures of Extraversion, it is equally meaningful to examine relations with outcomes at the facet level.

It is important to study the incremental prediction of the facets over and above the core construct (Ozer & Benet-Martínez, 2006). That is, it is possible that only certain facets are responsible for an observed effect and that other facets contribute nothing of their own. Indeed, one facet can be responsible for correlations with one set of criterion variables, whereas another facet can be responsible for correlations with another set of criterion variables. It is also possible for the two dimensions to correlate in opposite directions with criterion variables (e.g., Tett et al., 2003). Finally, it is even possible that a general construct in a bifactor model may not contribute anything over and above the individual facets.

Gender Differences in Extraversion

Findings are inconsistent with respect to gender differences in the construct of Extraversion. One meta-analysis (Feingold, 1994) found that women are slightly higher on this trait, whereas another meta-analysis (Lynn & Martin, 1997) and an empirical study (Costa, Terracciano, & McCrae, 2001) found the reverse. Given that Extraversion is a combination of typically masculine (e.g., assertiveness) and feminine (e.g., warmth) traits, gender differences likely vary by facets. That is, women are higher on Warmth (Costa et al.,
2001), Gregariousness (Feingold, 1994; Costa et al., 2001), and Positive Emotions (Costa et al., 2001). In contrast, men are higher on Assertiveness (Feingold, 1994; Costa et al., 2001) and Excitement Seeking (Costa et al., 2001). Depending on which facets are emphasized in a given instrument, gender differences on the general construct of Extraversion will vary. For example, Costa et al. (2001) point out that the Extraversion of NEO-PI-R emphasizes warmth more than assertiveness, whereas the opposite is true for the Eysenck scale (Eysenck, 1978). However, results based on bifactor models will be less susceptible to variation in facet composition because gender differences in the general factor will be less likely to be contaminated with gender differences in the facets.

**METHOD**

*Participants*

In Study 1, the participants were 383 undergraduate introductory psychology students (223 women and 160 men) at a medium-sized mid-Atlantic region university, and they were fulfilling their course research credit requirement. The mean age was 19.00 (SD = 1.08); 78.1% were European American, 5% African American, 6.5% Asian American, 3.1% Hispanic American, and 7.3% other. In Study 2, the participants were 378 undergraduate introductory psychology students (220 women and 158 men) at the same university. The mean age of the second sample was 19.13 (SD = 1.40); 73.3% were European American, 5.6% African American, 10.6% Asian American, 4.9% Hispanic American, and 5.6% other.

*Measures*

To cross-validate the bifactor structure of Extraversion, the NEO-PI-R scale of Extraversion was administered in both studies. All of the scales that served as outcomes were included in Study 2. Participants answered the items on a 9-point scale ranging from 1 (*Does not describe me at all*) to 9 (*Describes me very well*), except for positive and negative affect, which is described below.

*Extraversion.* Extraversion was measured by the relevant scales from the NEO-PI-R (Costa & McCrae, 1992). The NEO-PI-R is a measure of the five major domains of personality and six facets that define each domain. For multidimensional constructs, the alpha coefficient is complexly determined, and McDonald’s $\omega_h$ (1999) provides a better estimate for
the composite score and thus should be used (Zinbarg, Revelle, Yovel, & Li, 2005). McDonald’s $\omega_n$ for Extraversion was .92 and .92 for Studies 1 and 2.

**Positive and negative affect.** Positive and negative affect were assessed using the 10-item scales from the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988). The Positive Affect scale consists of *active, alert, attentive, determined, enthusiastic, excited, inspired, interested, proud*, and *strong*. The Negative Affect scale is composed of *afraid, ashamed, distressed, guilty, hostile, irritable, jittery, nervous, scared*, and *upset*. Participants were asked to rate the extent to which they had felt each of the affects “during the past few days.” The response scale ranged from 1 to 5 as follows: 1 (*very slightly or not at all*), 2 (*a little*), 3 (*moderately*), 4 (*quite a bit*), and 5 (*very much*), respectively. Cronbach’s alpha for positive affect and negative affect was .87 and .87.

**The Hopelessness Scale.** This scale (Beck, Weissman, Lester, & Trexler, 1974) assesses the extent to which individuals possess hopeless and unfavorable expectations regarding life outcomes. We selected eight items to measure optimism and seven items for pessimism (Marshall, Wortman, Kusulas, Hervig, & Vickers, 1992). The remaining five items did not load on either optimism or pessimism in a factor analysis (Marshall et al., 1992) and thus were not used in this study. Cronbach’s alpha was .82 for optimism and .88 for pessimism.

**The Satisfaction With Life Scale.** The five-item scale (Diener, Emmons, Larsen, & Griffin, 1985) measures life satisfaction as a cognitive-judgmental process by asking a person to provide an overall judgment of the satisfaction he or she has with life. The scale includes items such as “In most ways my life is close to my ideal,” “I am satisfied with my life,” and “The conditions of my life are excellent.” Cronbach’s alpha for this scale was .92.

**Rosenberg Self-Esteem.** This commonly used 10-item scale (Rosenberg, 1965) measures global self-esteem. Cronbach’s alpha for this scale was .91.

**Behavioral Inhibition System (BIS).** The seven-item BIS scale (Carver & White, 1994) assesses individual differences in sensitivity to threats. An example item is “If I think something unpleasant is going to happen, I usually get pretty worked up.” Cronbach’s alpha for this scale was .77.

**Beck Depression Inventory-II (BDI-II).** The BDI-II (Beck, Steer, & Brown, 1996) assesses the severity of self-reported depression in both
clinical and nonclinical samples. It includes items that assess self-dislike, sadness, pessimism, fatigue, insomnia, and hypochondriasis. These symptoms fall into two related factors: negative attitudes and somatic complaints. Participants were asked to describe “the way you have been feeling during the past two weeks, including today.” Each question has a set of four possible answers, ranging in intensity. For example, the item referring to problems with concentration uses a response scale ranging from 0 to 3: 0 (I can concentrate as well as ever), 1 (I can’t concentrate as well as usual), 2 (It’s hard to keep my mind on anything for very long), and 3 (I find I can’t concentrate on anything). Cronbach’s alphas for the two subscales of the BDI-II were .85 for negative attitude and .89 for somatic complaints.

RESULTS

We used conventional criteria of three model fit indices that are widely used in the literature. The recommended cut-off points are .08 for RMSEA (Browne & Cudeck, 1993), .08 for SRMR (Hu & Bentler, 1999), and .95 for CFI. However, a cut-off point of .90 CFI was used in this study because compared to other fit indices, such as RMSEA, CFI tends to perform poorly when the number of variables (i.e., indicators) per factor is large (e.g., Kenny & McCoach, 2003). In this sample, the maximum number of variables per factor is 23, and thus a less stringent standard of .90 was used for CFI.

Following the convention for testing nested models, chi-square difference tests were used to compare the models (Bentler & Bonett, 1980), given that the second-order model is nested in the bifactor model (Rindskopf & Rose, 1988; Yung, Thissen, & McLeod, 1999).

To make the models more parsimonious, parcels were created by aggregating two items in each parcel (Little, Cunningham, Shahar, & Widaman, 2002). Two different strategies were employed to create the parcels: One was to pair a positively worded item with a negatively worded item, and the other was to couple similarly worded items (i.e., either positive or negative). Given that the results were similar, the final models were based on the first approach. Before creating the parcels, a six-factor model was conducted at the item level to examine the relations between the items and the latent factors. In both studies, items “I have a leisurely style in work and play” and “I usually seem to be in a hurry” in the Activity facet had either a negative or near-zero factor loading. These two items were removed from further analyses. As a result, the facet Activity has three, rather than four, parcels.
Testing the Hypothesized Bifactor Model (Figure 1)

Bifactor models differ from conventional confirmatory factor analysis in two ways. First, each item loads on the general factor as well as on the specific factor, if a specific factor exists. Second, the general factor is uncorrelated with the specific factors, and the specific factors are uncorrelated with each other. Analyses were conducted using Mplus software (Muthén & Muthén, 2010), but bifactor models can be tested with any structural equation modeling program. Mplus syntax for the tested model is presented at http://www.psych.udel.edu/people/detail/fang_fang_chen.

It was hypothesized that responses to the items from the Extraversion scales could be explained by one general factor, plus a number of specific factors corresponding to each of the construct’s facets (see Figure 1). The bifactor model was constructed with the following features: (a) Each item had a nonzero loading on both the general factor and the specific factor that it was designed to measure, but zero loadings on the other specific factors; (b) the specific factors were uncorrelated with each other and with the general factor; and (c) all error terms associated with the items 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 specific factors was also set to 1. The variances of the factors were estimated. The standardized factor loadings of the bifactor model are presented in Table 1.

The model fit the data adequately, $\chi^2 = 630.09$ ($df = 209$, $N = 383$), RMSEA = .073 (CI: .066–.079), SRMR = .058, and CFI = .905 in Study 1; $\chi^2 = 694.01$ ($df = 209$, $N = 378$), RMSEA = .079 (CI: .072–.085), SRMR = .062, and CFI = .895 in Study 2. However, the variance of the Warmth factor was not significant, variance = .06, $t = .06$, $ns$, in Study 1; variance = .45, $t = .09$, $ns$, in Study 2. This suggests that Warmth may not be a specific factor over and above the general factor and that the variance related to Warmth is entirely explained by the general factor.

Testing the Hypothesized Second-Order Model

The second-order model was specified in the following way: (a) Each item had a nonzero loading on the first-order factors that it was designed to measure and a zero loading on other first-order factors;
Table 1
Factor Loadings of the Bifactor Model of Extraversion

| Parcels   | General | Warmth | Gregariousness | Assertiveness | Activity |  |  |  |
|-----------|---------|--------|----------------|---------------|----------| | | |
| Warm1     | .69     | .42    |                |               |          |  |  |  |
| Warm2     | .70     | .10    |                |               |          |  |  |  |
| Warm3     | .79     | .15    |                |               |          |  |  |  |
| Warm4     | .65     | .10    |                |               |          |  |  |  |
| Greg1     | .57     | .25    |                |               |          |  |  |  |
| Greg2     | .63     | .54    |                |               |          |  |  |  |
| Greg3     | .54     | .19    |                |               |          |  |  |  |
| Greg4     | .45     | .39    |                |               |          |  |  |  |
| Assert1   | .28     | .43    | .43            | .55           |          |  |  |  |
| Assert2   | .41     | .81    | .43            | .45           |          |  |  |  |
| Assert3   | .44     | .43    |                |               |          |  |  |  |
| Assert4   | .45     | .41    |                |               |          |  |  |  |
| Act1      | .41     | .25    |                |               |          |  |  |  |
| Act2      | .56     | .83    | .38            | .44           |          |  |  |  |
| Act3      | .39     | .12    | .38            | .44           |          |  |  |  |
| Excite1   | .39     | .38    |                |               |          |  |  |  |
| Excite2   | .22     | .48    |                |               |          |  |  |  |
| Excite3   | .58     | .17    |                |               |          |  |  |  |
| Excite4   | .53     | .44    |                |               |          |  |  |  |
| Emotion1  | .55     | .41    |                |               |          |  |  |  |
| Emotion2  | .64     | .24    |                |               |          |  |  |  |
| Emotion3  | .54     | .11    |                |               |          |  |  |  |
| Emotion4  | .58     | .34    |                |               |          |  |  |  |
| Variances | 1.84    | 1.71   | 1.49           | .69           | .46      |  |  |  |

Note. Warm = Warmth; Greg = Gregariousness; Assert = Assertiveness; Act = Activity; Excite = Excitation Seeking; Emotion = Positive Emotions.

*a* Standardized factor loadings for Study 1.

*b* Standardized factor loadings for Study 2.

*c* Nonsignificant factor loading/variance.
(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—general extraversion. 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. The model fit the data marginally well, $\chi^2 = 713.06$ ($df = 224$, $N = 383$), RMSEA = .076 (CI: .069–.082), SRMR = .065, and CFI = .879 in Study 1; $\chi^2 = 784.70$ ($df = 224$, $N = 378$), RMSEA = .081 (CI: .075–.088), SRMR = .070, and CFI = .874 in Study 2.

The bifactor models fit the data significantly better than the second-order model, $\Delta\chi^2 = 82.97$ ($\Delta df = 15$), $p < .001$ in Study 1; $\Delta\chi^2 = 90.69$ ($\Delta df = 15$), $p < .001$ in Study 2, suggesting that the bifactor models provided a better approximation to the data and thus a better interpretation of the data than the second-order model.

Relations of General and Specific Factors of Extraversion to External Variables (Figure 2)

To examine whether the specific factors were uniquely associated with outcome variables beyond the general factor, the general and specific factors were used to predict the outcome variables simultaneously in the bifactor models. In the second-order models, the disturbances of the lower order factors were used to predict the outcomes, in addition to using the second-order factor. We had to take a very complicated approach to make the Mplus program perform the second-order analysis because they were nonconventional models. For comparison purposes, we also tested the residual approach in which the residual score of each facet (i.e., partialling out the other five facets) was used to predict the outcome variables in regression analysis. To contrast these findings with those from the individual and total score approaches, simple correlations based on those two approaches are also presented. Patterns of the findings based on the bifactor model, second-order model, and residual score approach are similar. For ease of presentation, we will only describe the results from bifactor models\(^5\) (see Table 2).

5. Results from the second-order model and residual regression can be obtained from the first author upon request.
Positive affect. Similar to the total score approach ($r = .47$), the general Extraversion factor in the bifactor model predicted positive affect, $\beta = .42$, $t = 8.20$, $p < .001$. Two facets also predicted positive affect: Assertiveness, $\beta = .18$, $t = 3.07$, $p < .01$; and Activity, $\beta = .33$, $t = 6.98$, $p < .001$. In contrast to the individual score approach, the positive relation between facet Gregariousness and positive affect ($r = .22$) was negative in the bifactor model, $\beta = -.12$, $t = -2.54$, $p < .05$.

Negative affect. Similar to the total score approach ($r = -.29$), the general Extraversion factor in the bifactor model predicted negative affect, $\beta = -.45$, $t = -8.71$, $p < .001$. In contrast to the individual score approach, two facets positively predicted negative affect: Gregariousness, $\beta = .44$, $t = 5.86$, $p < .01$; and Activity, $\beta = .21$, $t = 3.93$, $p < .001$. 

Figure 2
Predictive relations of general and specific factors of Extraversion to life satisfaction. Note. Standardized coefficients are presented. *$p < .05$. **$p < .01$. ***$p < .001$. 

Chen, Hayes, Carver, et al. 234
Table 2
Relations of the General and Specific Factors of Extraversion to External Variables

<table>
<thead>
<tr>
<th>Extraversion</th>
<th>General</th>
<th>Warmth</th>
<th>Gregariousness</th>
<th>Assertiveness</th>
<th>Activity</th>
<th>Excitement Seeking</th>
<th>Positive Emotions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bifactor model</td>
<td>.42***</td>
<td>-.12*</td>
<td>.18**</td>
<td>.33***</td>
<td>(.05)</td>
<td>.18*</td>
<td></td>
</tr>
<tr>
<td>Total/individual score</td>
<td>.47***</td>
<td>.36***</td>
<td>.22***</td>
<td>.36***</td>
<td>.54***</td>
<td>.28***</td>
<td>.43***</td>
</tr>
<tr>
<td>Negative Affect</td>
<td>Bifactor model</td>
<td>-.45 ***</td>
<td>.44***</td>
<td>(.02)</td>
<td>.21***</td>
<td>(-.02)</td>
<td>(-.10)</td>
</tr>
<tr>
<td>Total/individual score</td>
<td>-.29***</td>
<td>-.25***</td>
<td>-.24***</td>
<td>-.28***</td>
<td>-.19***</td>
<td>-.10*</td>
<td>-.26***</td>
</tr>
<tr>
<td>Optimism</td>
<td>Bifactor model</td>
<td>.56***</td>
<td>-.31***</td>
<td>(-.01)</td>
<td>(.08)</td>
<td>(.01)</td>
<td>(.06)</td>
</tr>
<tr>
<td>Total/individual score</td>
<td>.50***</td>
<td>.48***</td>
<td>.24***</td>
<td>.31***</td>
<td>.50***</td>
<td>.29***</td>
<td>.51***</td>
</tr>
<tr>
<td>Pessimism</td>
<td>Bifactor model</td>
<td>-.60***</td>
<td>.40***</td>
<td>-.12*</td>
<td>(.08)</td>
<td>(.11)</td>
<td>(-.13)</td>
</tr>
<tr>
<td>Total/individual score</td>
<td>-.50***</td>
<td>-.43***</td>
<td>-.35***</td>
<td>-.44***</td>
<td>-.39***</td>
<td>-.14***</td>
<td>-.52***</td>
</tr>
<tr>
<td>Satisfaction With Life</td>
<td>Bifactor model</td>
<td>.52***</td>
<td>-.20**</td>
<td>(.10)</td>
<td>.11*</td>
<td>(.02)</td>
<td>.21**</td>
</tr>
<tr>
<td>Total/individual score</td>
<td>.51***</td>
<td>.46***</td>
<td>.28***</td>
<td>.35***</td>
<td>.46***</td>
<td>.27***</td>
<td>.52***</td>
</tr>
<tr>
<td>Self-Esteem</td>
<td>Bifactor model</td>
<td>.64***</td>
<td>-.34***</td>
<td>(.06)</td>
<td>(.06)</td>
<td>(-.09)</td>
<td>(.00)</td>
</tr>
<tr>
<td>Total/individual score</td>
<td>.55***</td>
<td>.47***</td>
<td>.38***</td>
<td>.44***</td>
<td>.50***</td>
<td>.20***</td>
<td>.52***</td>
</tr>
<tr>
<td>BIS</td>
<td>Bifactor model</td>
<td>(.00)</td>
<td>(.00)</td>
<td>-.46***</td>
<td>-.20***</td>
<td>-.31***</td>
<td>(-.17)</td>
</tr>
<tr>
<td>Total/individual score</td>
<td>-.13*</td>
<td>(.06)</td>
<td>(-.06)</td>
<td>-.32***</td>
<td>-.12*</td>
<td>-.13*</td>
<td>(-.02)</td>
</tr>
<tr>
<td>BDI-II</td>
<td>Bifactor model</td>
<td>-.46***</td>
<td>.20*</td>
<td>(-.09)</td>
<td>(-.05)</td>
<td>(.01)</td>
<td>(-.02)</td>
</tr>
<tr>
<td>Total/individual score</td>
<td>-.32***</td>
<td>-.29***</td>
<td>-.21***</td>
<td>-.22***</td>
<td>-.25***</td>
<td>-.18***</td>
<td>-.28***</td>
</tr>
</tbody>
</table>

Note. BIS = Behavioral Inhibition System. BDI-II = Beck Depression Inventory II.
*p < .05. **p < .01. ***p < .001. Nonsignificant values are included in parentheses.
Optimism. Similar to the total score approach \((r = .50)\), the general Extraversion factor predicted optimism, \(\beta = .56, t = 12.46, p < .001\). In contrast to the individual score approach, the positive relation between facet Gregariousness and optimism \((r = .24)\) was negative in the bifactor model, \(\beta = -.31, t = -4.24, p < .001\).

Pessimism. Similar to the total score approach \((r = -.50)\), the general Extraversion factor predicted pessimism, \(\beta = -.60, t = -12.21, p < .001\). Facet Assertiveness also predicted pessimism, \(\beta = -.12, t = -1.98, p < .05\). In contrast to the individual score approach, the negative relation between facet Gregariousness and pessimism \((r = -.35)\) was positive in the bifactor model, \(\beta = .40, t = 5.93, p < .001\).

Satisfaction with life. Similar to the total score approach \((r = .51)\), the general Extraversion factor predicted life satisfaction, \(\beta = .52, t = 11.43, p < .001\). Two facets also predicted life satisfaction: Activity, \(\beta = .11, t = 2.37, p < .05\); and Positive Emotions, \(\beta = .21, t = 2.59, p = .01\). In contrast to the individual score approach, the positive relation between facet Gregariousness and life satisfaction \((r = .28)\) was negative in the bifactor model, \(\beta = -.20, t = -2.94, p < .01\).

Self-esteem. Similar to the total score approach \((r = .55)\), the general Extraversion factor predicted self-esteem, \(\beta = .64, t = 15.64, p < .001\). In contrast to the individual score approach, the positive relation between facet Gregariousness and self-esteem \((r = .38)\) was negative in the bifactor model, \(\beta = -.34, t = -4.76, p < .001\).

BIS. Different from the total score approach \((r = -.13)\), the general Extraversion factor did not predict BIS, \(\beta = .00, t = .06, ns\). However, three specific factors did predict BIS: Assertiveness, \(\beta = -.46, t = -8.37, p < .001\); Activity, \(\beta = -.20, t = -3.73, p < .001\); and Excitement Seeking, \(\beta = -.31, t = -4.71, p < .001\).

BDI-II. Similar to the total score approach \((r = -.32)\), the general Extraversion factor predicted BDI-II, \(\beta = -.46, t = -8.82, p < .001\). In contrast to the individual score approach, the negative relation between facet Gregariousness and BDI-II \((r = -.21)\) was positive in the bifactor model, \(\beta = .20, t = 2.02, p < .05\).
To summarize the findings thus far, the general construct of Extraversion predicted all four measures of positive subjective well-being: positive affect, optimism, satisfaction with life, and self-esteem. Facets Activity and Positive Emotions predicted positive affect and satisfaction with life. In addition, the general construct of Extraversion predicted all three measures of negative well-being: negative affect, pessimism, and depressive symptoms. However, facet Gregariousness had a negative relation with all four measures of positive well-being but a positive relation with all measures of negative well-being. It is possible that when social interactions do not involve warmth and interpersonal bond, the need for social activities (i.e., gregariousness) signals loneliness and lack of quality relations, which in turn may have detrimental effects on subjective well-being. Finally, although the general Extraversion did not predict BIS, three facets—Assertiveness, Activity, and Excitement Seeking—were negatively associated with BIS.

The bifactor models provided richer and conceptually clearer information than either the total or individual score approach alone. The pattern of the findings based on the general factor of the bifactor model was generally congruent with that based on the total score approach. However, the strength of the relation often differed because in the total score approach, the common variance shared by the facets was contaminated with the unique variance specific to the facets. In addition, unlike the bifactor model, the total score glossed over important information on the relation between the specific factors and external variables. Compared with the individual score approach, the pattern of the findings was often quite different, sometimes even in opposite directions, because the bifactor approach separated the common variance shared by the facets from the unique variance specific to the facets. The bifactor models also provided the best fit for the data, suggesting that the bifactor models offer more accurate interpretation of the data.

The pattern of results from the second-order and residual approaches was similar to that of bifactor models. However, the bifactor model is more flexible and has unique advantages. First, the standard errors based on the bifactor model are in general smaller than those from second-order models, indicating that the estimates of parameters (e.g., beta weights) are more precise. Second, the bifactor model fits the data better than the second-order model, suggesting that the bifactor model provides more accurate representa-
tation of the data. Third, only bifactor models can estimate the latent means for the facets, as illustrated in the following section. Finally, the residual regression approach does not have all the advantages that latent factor models (i.e., bifactor model and second-order model) embody.

Comparing Factor Means Across Gender on Extraversion

Gender comparisons were also conducted on the general and specific factors of the Extraversion bifactor models. To obtain an estimate of the difference between the factor means, males were chosen as the reference group, and their factor means were set to zero. The factor means of the females were estimated, which is the difference between the factor means of the two groups. The significance test ($z$ test) for the latent means of the female group is the test for significance of the difference between the means of the two groups on the factors (Aiken, Stein, & Bentler, 1994). For comparison purposes, the results based on the latent mean analysis are contrasted with those obtained from the total and individual score approaches using the ANOVA analysis; in addition, results based on previous meta-analysis are presented (see Table 3).

Across both studies, consistent gender differences were found based on the bifactor model. Women scored higher on general Extraversion, $\beta = .48$ (i.e., women were .48 standard deviation higher than men), $z = 4.23, p < .001$ in Study 1; $\beta = .49, z = 4.03, p < .001$ in Study 2. Before testing latent mean differences, measurement invariance was conducted. In Study 1, there was no significant $\chi^2$ difference between the factor loading invariance and configural invariance models, $\Delta \chi^2 = 53.82$ ($\Delta df = 39$), ns; there was no significant change in RMSEA (.080 vs. .078), SRMR (.065 vs. .079), and CFI (.905 vs. .894), indicating that factor loadings were invariant across gender. The $\chi^2$ difference between the factor loading invariance and intercept invariance models was not significant, $\Delta \chi^2 = 62.40$ ($\Delta df = 23$), ns; there was no significant change in RMSEA (.078 vs. .081), SRMR (.079 vs. .086), and CFI (.894 vs. .882), indicating that intercepts were invariant across gender.

In Study 2, the $\chi^2$ difference was significant between the factor loading invariance and configural invariance models, $\Delta \chi^2 = 60.90$ ($\Delta df = 39$), $p < .05$, but there was no significant change in RMSEA (.081 vs. .089), SRMR (.073 vs. .085), and CFI (.895 vs. .886), indicating that factor loadings were invariant across gender. The $\chi^2$ difference between the factor loading invariance and intercept invariance models was not significant, $\Delta \chi^2 = 35.07$ ($\Delta df = 23$), ns; there was no significant change in RMSEA (.089 vs. .089), SRMR (.085 vs. .088), and CFI (.886 vs. .881), indicating that intercepts were invariant across gender.
### Table 3
Gender Differences in General and Specific Factors of Extraversion Based on Bifactor Models and ANOVA

<table>
<thead>
<tr>
<th></th>
<th>Study 1</th>
<th>Study 2</th>
<th>Meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>General</td>
<td>Warmth</td>
<td>Gregariousness</td>
</tr>
<tr>
<td>Bifactor model</td>
<td>.48***</td>
<td>.21*</td>
<td>(-.11)</td>
</tr>
<tr>
<td>Total score</td>
<td>.29**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual score</td>
<td>.59***</td>
<td>.35**</td>
<td>(-.05)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 2</td>
<td>General</td>
<td>Warmth</td>
<td>Gregariousness</td>
</tr>
<tr>
<td>Bifactor model</td>
<td>.49***</td>
<td>.43**</td>
<td>(-.12)</td>
</tr>
<tr>
<td>Total score</td>
<td>.37***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual score</td>
<td>.69***</td>
<td>.34*</td>
<td>(.05)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meta-analysis</td>
<td>Pattern</td>
<td>Mixed</td>
<td>Women higher</td>
</tr>
</tbody>
</table>

*Note. The means based on the bifactor model are standardized latent mean differences. Positive values indicate that women scored higher and negative values indicate that men scored higher. 
*p < .05. **p < .01. ***p < .001. Nonsignificant values are included in parentheses.*
Study 2. Women also scored higher on two facets: Warmth, $\beta = .21, z = 1.98, p < .05$ in Study 1; $\beta = .43, z = 3.24, p < .01$ in Study 2; Positive Emotions, $\beta = .25, z = 2.25, p < .05$ in Study 1; $\beta = .90, z = 3.87, p < .001$ in Study 2. However, after controlling for general Extraversion, men scored higher on two other facets: Assertiveness, $\beta = -.41, z = -3.30, p = .001$ in Study 1; $\beta = -.41, z = -2.98, p < .01$ in Study 2; Excitement Seeking, $\beta = -.48, z = -2.65, p < .01$ in Study 1; $\beta = -.63, z = -4.28, p < .001$ in Study 2. In addition, men scored higher on Activity in Study 1, $\beta = -.32, z = -2.66, p < .01$. These results are in sharp contrast with those based on the individual approach, in which men did not score higher on Assertiveness, Activity, and Excitement Seeking than women.

**DISCUSSION**

What is the best way to analyze multifaceted psychological constructs? How can we test the unique contributions of facets over and above the general construct? Researchers engaged in vigorous debate on this topic 20 years ago, and there is still no clear answer to these questions. We describe the bifactor model as a viable solution that encompasses the advantages of the total and individual score approaches yet also overcomes their shortcomings. The bifactor model simultaneously examines the general factor shared by the facets and the specific factors unique to the facets. This approach has several specific advantages: (a) It can test the predictive validity of the general factor shared by the facets; (b) it can examine the predictive validity of specific factors after partialling out the general factor, which can be important when researchers are interested in the unique contribution of the specific factors; and (c) it can compare mean differences on variables, such as gender, for both the general and specific factors. Taken together, these advantages can lead to greater conceptual clarity because bifactor models can separate the commonality shared by the facets from the unique contribution of each facet. We also illustrated in two samples that the bifactor model provided a significantly better representation of the data than the more traditional total score and individual score approaches.

Although the bifactor model fit the data better than the second-order model, the difference in fit was small, and both these models and the regression model provided similar parameter estimates. The main strength of the bifactor model is that it estimates the relations
between the latent facets and the outcome variables more seamlessly than the second-order model. In addition, the bifactor model provides a mean structure that allows gender and possibly other group comparisons. That is, it is the substantive possibilities provided by the bifactor model that are its real strength.

Numerous discussions of problems in interpreting multidimensional scales can be found in personality literature (Type A: Booth-Kewley & Friedman, 1987; Dembroski & Williams, 1989; self-monitoring: Briggs & Cheek, 1988; internal-external locus of control: Carver, 1997; hardiness: Hull, Van Treuren, & Virnelli, 1987; attributional style: Sweeney, Anderson, & Bailey, 1986; see Carver, 1989, for an overall discussion). However, if we study these relations at the general construct level using the total score approach or at the facet level relying on the individual score approach, it is unclear whether these findings are due to the general factor of the personality construct, the specific factors of these measures, or both. The bifactor model allows one to test these relations simultaneously, clearly separating these two sources of variation.

To illustrate the application of the bifactor model, we selected the Extraversion scale of the NEO-PI-R. We used the scale to test the relations of general Extraversion and its facets to subjective well-being. We also tested gender differences in general Extraversion and its domain facets.

The Relations of Extraversion to External Variables

The second-order model and residual regression analyses were included as comparisons to highlight the flexibility and precision of the bifactor model. As expected, the results from all approaches were similar, so we focus on the findings from the bifactor model. The relations between general Extraversion and positive subjective well-being were all in the expected direction. That is, general Extraversion predicted all four measures of well-being: positive affect, optimism, satisfaction with life, and self-esteem. However, some specific facets were related to subjective well-being in opposite directions: Gregariousness was negatively associated with positive affect, optimism, satisfaction with life, and self-esteem, whereas Activity and Positive Emotions were positively associated with positive affect and satisfaction with life, after partialling out general Extraversion. The positive relation between facet Positive Emotions and satisfaction with life is
consistent with another study in which such a link was observed across three time points (Schimmack, Oishi, Furr, & Funder, 2004). The negative relation between Gregariousness and well-being is the opposite of those found with total and individual score approaches, and it was revealed in all three methods: bifactor models, second-order models, and residual regression. If replicated in future studies, the finding that Gregariousness might be a less adaptive facet of Extraversion could have important implications.

Also as expected, general Extraversion was negatively related to negative affect, pessimism, and depressive symptoms. Overall, the bifactor model provided more refined information than the total score approach and revealed possible mischaracterizations that could arise from both the individual and total score approaches.

It is worth mentioning that the facet Warmth had nonsignificant or trivial residual variance after partialling out general Extraversion in both the bifactor model (as a specific factor) and second-order model (as a disturbance term). Results from the residual regression approach are consistent with this finding in that the residual variance of Warmth did not predict psychological well-being (except for optimism). However, previous research using a residual score approach indicated that the residual variance of Warmth had cross-observer validity (McCrae & Costa, 1991; McCrae & Costa, 2008), was reliable and heritable (Jang et al., 1998), and had unique developmental trajectories (McCrae et al., 1999). One possible explanation for this discrepancy is due to random sampling variation. Specifically, in follow-up studies, we tested the bifactor model of Extraversion in two large samples (Ns = 500, 493). The specific factor of Warmth from the bifactor model was significantly related to well-being in both studies (Chen, Hayes, & Jing, 2011). Nevertheless, consistent with the present studies, Warmth defined the general construct of Extraversion in the follow-up studies as well. That is, the items that measure the Warmth facet had the strongest loadings on the general factor of Extraversion.

**Gender Differences in Extraversion**

The bifactor model analysis also helps to clarify the inconsistency in the literature on gender differences in the general construct of Extraversion. The bifactor models (but not the individual approach) revealed that gender differences varied by facets of Extraversion,
after partialling out overall Extraversion as measured by the NEO-PI-R: women scored higher on Warmth and Positive Emotions, but men scored higher on Assertiveness and Excitement Seeking. These findings are congruent with the theoretical reasoning that men should be higher on typically masculine traits, whereas women are supposed to be higher on feminine traits (e.g., Eagly & Wood, 1991; Maccoby & Jacklin, 1974). These results are also consistent with those of meta-analytic reviews on gender differences in Extraversion (Feingold, 1994) and Costa and colleagues’ work (2001). Given that gender differences can emerge in opposite directions on the facets, results based on the total score approach will inevitably change, as found in the literature, depending on which facets are overrepresented in the measures used. However, bifactor models are less likely to be influenced by such variation in facet composition, as the general construct and facets are separated in such models. These gender differences in facets can only be revealed in bifactor models.

**Implications of Bifactor Models**

Multifaceted constructs are very common in psychological research, and bifactor models are well suited for separating the unique contributions of facets from a general construct. For example, Watson and Clark (1992) attempted to examine the relationships between negative affect and its four facets: sadness, fear, anger, and guilt. Can these facets be distinguished from general negative affect? Their data in Study 1 can be reanalyzed by constructing a bifactor model: The 4 measures of sadness/depression, the 3 measures of fear/anxiety, and the 3 measures of anger/hostility would form three facets, and all 10 measures would form the general negative affect factor. If the facets do not exist over and above the general factor, it suggests that the specific facets cannot be separated from general negative affect. In contrast, if the facets do exist and can predict outcome variables over and beyond the general negative factor, it indicates that sadness, fear, and anger can be separated from the general construct.

Bifactor models also can be used to test hypotheses regarding the relation between a multifaceted construct and a criterion. For example, researchers have been puzzled by the robust relation of Extraversion to positive affect. The temperament model (McCrae & Costa, 1991; Tellegen, 1985; Watson & Clark, 1984) proposes that individual differences in personality and emotionality ultimately
reflect the same common underlying constructs. The temperament view emphasizes that individual differences in positive emotionality comprise central, organizing features of Extraversion. Thus, they should relate to each other at the higher order or general factor level. This can be directly tested by examining the relationship between a bifactor model of Extraversion and a bifactor model of positive affect. If the hypothesis is supported, the relation between the general construct of Extraversion and that of positive affect should be stronger than the relations at the facet level across the two constructs. In contrast, the instrumental model (McCrae & Costa, 1991) argues that the relations of Agreeableness and Conscientiousness to affect should occur at the lower order or specific factor level. This is because certain characteristics of Agreeableness and Conscientiousness, rather than the general constructs, are expected to foster situations and life experiences that would have strong specific emotional consequences. If that is the case, the relation between the general construct of Agreeableness and that of Conscientiousness and general positive or negative affect should be weaker than the relationships between the facets of these personality dimensions and facets of the affects.

Bifactor models are also particularly well suited to study gender or other group differences across a variety of constructs. Recent evidence from intelligence studies further indicates that the application of bifactor models can challenge well-established findings on gender differences. For example, a meta-analysis of high school mathematics ability (Hyde, Fennema & Lamon, 1990) reports that boys typically outperform girls in mathematics, but the effect size is small. Relying on a bifactor model to separate the general cognitive ability from the specific mathematical ability, Brunner, Krauss, and Kunter (2008) drew a different conclusion: The gender difference in mathematical ability is large rather than small. However, when the same data were analyzed with the traditional total score approach, in which the specific mathematical ability was confounded with the general cognitive ability, only a small gender difference was found. These results suggest that relying on total and individual scores can be misleading when analyzing multidimensional constructs because the total scores do not take into account the unique contribution of each facet, and the individual scores do not separate the common variance shared by the facets from the unique variance specific to the facets.
Finally, bifactor models can be useful in scale construction and evaluation. When developing a new multifaceted scale or evaluating an existing instrument that intends to measure a general construct and certain facets, the strength of the factor loadings on the general and specific factors can be used to guide item selection or reevaluation of an existing scale. Ideally, items should load stronger or at least not weaker on the general construct than on the facets. If certain items primarily load on the general construct with weak loadings on their facets, those facets should be eliminated as specific facets. If certain items primarily load on the specific factors, they should be omitted from the scale, as they do not contribute substantially to the general construct.

Limitations of Bifactor Models

Compared to second-order models, we do not wish to imply that bifactor models are applicable (or optimal) under all conditions. If the general factor is the main focus of the research, the second-order factor model may be more parsimonious. Moreover, the bifactor and second-order representations are not mutually exclusive, and they can coexist in different parts of the same complex model. Eid, Lischetzke, Nussbeck, and Trierweiler (2003) provided some examples in their representations of second-order multitrait-multimethod models. Finally, both approaches offer unique features that only latent variable models can provide, such as taking into account measurement error, testing multiple multifaceted variables simultaneously, and providing model fit.

There are several limitations to consider with the bifactor model. Only a limited set of specific factors may be included with the general factor when predicting 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. As any structural equation models (SEMs), bifactor models require sufficiently large sample sizes, given that more parameters are estimated than in total score or individual score approaches. Minimum sample size depends on several aspects of a given study, including the level of communality of the variables and the level of overdetermination of the factors (MacCallum, Widaman, Zhang, & Hong, 1999). Compared to regular SEMs, such
as second-order factor models, bifactor models may require the provision of start values for the models to converge, and thus it may take more effort to run the models, given the unique nature of the bifactor models.

It is worth noting that bifactor models are difficult to interpret when the specific factors are allowed to correlate (Rindskopf & Rose, 1988), and the model often cannot be identified. In addition, as in other factor models, the bifactor model is not appropriate given weak loadings, nonconvergence, and poor fit. Finally, not all multifaceted constructs can be subject to a bifactor or second-order structure. For example, among three facets of the Self-Monitoring Scale (Briggs, Cheek, & Buss, 1980), Acting is positively related to Other Directedness, which has no relation with Extraversion, which in turn is negatively related to Acting (cf. Finch & West, 1997). In this case, a bifactor model (or a second-order factor) is not applicable. When we tested a bifactor model on the self-monitoring data used by Finch and West (1997), items that measure the Extraversion facet loaded positively and strongly on the general Self-Monitoring factor and the facet did not form a specific factor, over and beyond the general factor; although items that measure the Acting facet had positive loadings on the general factor, the loadings were much stronger on the specific factor; items that measure Other Directedness had negative loadings on the general factor, and the loadings were much stronger on the specific factor. These results suggest that these three facets do not form a general construct. Only when the facets are connected by a common higher order factor can a bifactor model apply. Even with these limitations, however, bifactor models can be applied to a wide range of psychological constructs.

**CONCLUSION**

Bifactor models have several advantages over traditional approaches in testing multifaceted psychological constructs. Bifactor models can address conceptual issues that are unique to multifaceted constructs, such as the defining feature of a construct and the separability or inseparability of the general construct and its facets. As does the total score approach, bifactor models test the commonality among the facets. As does the individual score approach, bifactor models examine the unique contributions of the facets. In addition,
bifactor models go beyond the total and individual score approaches by testing the contribution of specific factors, over and above the general factor shared by the facets. Bifactor models can also test group differences by separating the specific factors from the general factor. Bifactor models thus provides a method for studying multi-dimensional constructs that can increase conceptual clarity. In addition, bifactor models have the potential to reveal new findings or to help reconcile inconsistent results in the literature.

REFERENCES


Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach’s α, Revelle’s β and McDonald’s ωh: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*, 1–11.