Journal of Personality and Social Psychology
2009, Vol. 97, No. 6, 1142–1156

0022-3514/09/$12.00 DOI: 10.1037/a0017159

The Nature and Structure of Correlations Among Big Five Ratings: The Halo-Alpha-Beta Model

Ivana Anusic and Ulrich Schimmack
University of Toronto Mississauga

Rebecca T. Pinkus
SUNY Buffalo

Penelope Lockwood
University of Toronto St. George

In light of consistently observed correlations among Big Five ratings, the authors developed and tested a model that combined E. L. Thorndike's (1920) general evaluative bias (halo) model and J. M. Digman's (1997) higher order personality factors (alpha and beta) model. With 4 multitrait–multimethod analyses, Study 1 revealed moderate convergent validity for alpha and beta across raters, whereas halo was mainly a unique factor for each rater. In Study 2, the authors showed that the halo factor was highly correlated with a validated measure of evaluative biases in self-ratings. Study 3 showed that halo is more strongly correlated with self-ratings of self-esteem than self-ratings of the Big Five, which suggests that halo is not a mere rating bias but actually reflects overly positive self-evaluations. Finally, Study 4 demonstrated that the halo bias in Big Five ratings is stable over short retest intervals. Taken together, the results suggest that the halo-alpa-beta model integrates the main findings in structural analyses of Big Five correlations. Accordingly, halo bias in self-ratings is a reliable and stable bias in individuals' perceptions of their own attributes. Implications of the present findings for the assessment of Big Five personality traits in monomethod studies are discussed.

Keywords: Big Five, personality structure, higher order factors, evaluative bias, halo

One area of personality psychology is concerned with the description and measurement of stable individual differences in experiences and behaviors. Presently, the most widely used model of individual differences is the Big Five model (Costa & McCrae, 1985; Goldberg, 1981; John, Angleitner, & Ostendorf, 1988). The Big Five model measures individual differences in terms of five scales: Neuroticism (N), Extraversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C). Although some researchers have questioned whether the Big Five provide a useful level in a hierarchical model of personality traits (e.g., Ashton & Lee, 2007; Block, 1995), the Big Five dominate contemporary personality psychology and the present article assumes that these five dimensions of personality represent a meaningful level in a hierarchical taxonomy of personality traits.

Ivana Anusic and Ulrich Schimmack, Department of Psychology, University of Toronto Mississauga, Mississauga, Ontario, Canada; Rebecca T. Pinkus, Department of Psychology, University at Buffalo, The State University of New York; Penelope Lockwood, Department of Psychology, University of Toronto St. George, Toronto, Ontario, Canada.

This research was supported by an undergraduate Social Sciences and Humanities Research Council (SSHRC) scholarship awarded to Ivana Anusic and an SSHRC standard research grant awarded to Ulrich Schimmack.

Correspondence concerning this article should be addressed to Ulrich Schimmack, Department of Psychology, University of Toronto Mississauga, 3359 Mississauga Road North, Mississauga, ON L5L 1C6, Canada. E-mail: uli.schimma@utoronto.ca or to Ivana Anusic, who is now at the Department of Psychology, Michigan State University, East Lansing, MI 48824. E-mail: anusiciv@msu.edu

Theoretically, the Big Five are conceptualized as orthogonal dimensions (e.g., Costa & McCrae, 1995; Goldberg, 1993). However, empirical studies with the Big Five measures typically show weak to moderate correlations among the Big Five personality dimensions (Digman, 1997). Although this phenomenon is robust and well known, the nature and structure of these correlations remain unclear.

Some researchers have proposed that the correlations among the Big Five arise because these five factors are not at the basic level of personality description and that even broader factors exist at higher levels (DeYoung, 2006; Digman, 1997; Musek, 2007). Other researchers have argued that these correlations are largely due to biases (Biesanz & West, 2004; Paulhus & John, 1998; Saucier, 1994b). The nature of the correlations has important implications for personality research. For example, systematic biases in Big Five ratings may produce spurious correlations with other self-report measures in monomethod studies (Schimmack, Schupp, & Wagner, 2008).

Monomethod Studies of Big Five Correlations

Digman (1997) provided the first systematic analysis of correlations among Big Five ratings. He concluded that the observed correlations among Big Five ratings are caused by two orthogonal higher order personality traits, *alpha* and *beta*. According to this model, alpha influences N (negatively), A, and C, and beta influences E and O. Digman further proposed that alpha is related to social development, and that beta is related to striving toward personal growth, in order to provide a theoretical explanation of alpha and beta. We refer to this model as the *alpha-beta* (AB) *model*.

Digman's (1997) most convincing argument in favor of existence of the metatraits was the robustness of the higher order factor model. He cited excellent confirmatory fit indices (CFIs) for confirmatory tests of his model in data sets obtained from adults as well as children using self- and informant ratings of various personality inventories. However, Mutch (2005) failed to replicate the CFI values reported by Digman (1997). Moreover, it has been overlooked that the original article also reported poor fit in terms of another fit index, namely, the standardized root-mean-square residual (SRMR). Although poor CFI values may underestimate model fit because CFI tends to be biased for correlation matrices with weak correlations (Kenny, 2008), reanalyses of Digman's data clearly show poor model fit.

We reanalyzed Digman's (1997) adult data sets using multigroup analysis in MPlus5 (Muthén & Muthén, 1998–2007). Because the standard deviation information was not available for the data sets reported in Digman's (1997) article, we performed all analyses of these data sets with variances of one. The main difference of this analysis compared with an analysis based on actual covariations is the assumption that Big Five scales have approximately equal variances. This assumption is justified because there is no evidence of systematic differences in the variance of Big Five scales. Our results indicated that model fit did not meet standard criteria of good fit; CFI $>$ .95, root-mean-square error of approximation (RMSEA) $<$ .06, SRMR $<$ .08 (Schermelleh-Engel, Moosbrugger, & Müller, 2003); $\chi^2(90, N = 70\text{–}1{,}040) = 887.63$, CFI = .726, RMSEA = .148, SRMR = .153. Failure to replicate Digman's (1997) excellent fit indices casts some doubt on the robustness of the AB model.

One possible reason for poor model fit is that the structure of correlations among the Big Five differs across data sets because the structural relations among the Big Five are influenced by the specific measures used to assess the Big Five (e.g., (Ashton, Lee, Goldberg, & de Vries, 2009). Same adjectives may serve as indicators of different facets in different inventories (Block, 1995). It is likely that these differences in the lower order structure of the Big Five have implications for the higher order structure of the Big Five. Thus, it is unlikely that a single structural model will always fit the data. To test this hypothesis, we fitted a model to Digman's (1997) data sets in which we simply constrained each bivariate correlation to be equal across the data sets. Once more, model fit was not acceptable by standard criteria, $\chi^2(80, N = 70\text{–}1{,}040) = 826.12$, CFI = .744, RMSEA = .151, SRMR = .152. In order to rule out rating method as the source of heterogeneity between the data sets, we repeated the analysis, this time allowing the two informant-rated data sets to have a different pattern of correlations than the self-rated data sets. The fit of this model was also below the standard criteria, $\chi^2(70, N = 70\text{–}1{,}040) = 719.15$, CFI = .777, RMSEA = .151, SRMR = .142. Thus, structural heterogeneity across the data sets cannot be attributed to the use of different raters. These results confirm our hypothesis that structural relations among the Big Five are not consistent across different data sets. For this reason, it is not reasonable to assume that a single structural model can maximize fit across diverse data sets. Thus, in the absence of a generally accepted model of the lower order structure of the Big Five, analyses of the higher order structure can only aim to reveal the main structural relations that emerge consistently across diverse measures and samples.

Some studies proposed a revised model of the correlations among Big Five ratings in monomethod studies (DeYoung, 2006; DeYoung, Peterson, & Higgins, 2002; Musek, 2007). These studies demonstrate that model fit increases when alpha and beta are allowed to be correlated with each other. We refer to this model as the *correlated alpha-beta* (CAB) *model*. However, the authors of these studies did not test whether the advantages of the CAB model generalized to other data sets or were unique to the specific data sets of their particular study. We conducted another multigroup analysis to examine the fit of the CAB model to Digman's (1997) original data. We also did not constrain coefficients across studies on the basis of the earlier findings that structural information varied across data sets. Model fit was considerably better than model fit of the earlier analyses, although it did not meet standard levels of acceptable fit for CFI and RMSEA, $\chi^2(45, N = 70\text{–}1{,}040) = 368.30$, CFI = .889, RMSEA = .133, SRMR = .063.

Although the CAB model is an improvement over the AB model, it has some limitations. First, the correlation between alpha and beta was not theoretically motivated. Second, an important empirical finding was that alpha and beta were positively (rather than negatively) correlated, but no theoretical explanation for the sign of the correlation has been offered. Thus, the CAB model does improve fit, but an explanation for the correlation between alpha and beta is lacking.

Multimethod Studies of Correlations Among the Big Five

A major limitation of the studies reviewed so far is the exclusive reliance on ratings by a single rater, mostly self-report. As a result, it remains unclear whether observed correlations among Big Five measures reveal structural relations among the Big Five personality traits or the influence of method artifacts on ratings of the Big Five (Campbell & Fiske, 1959). A general agreement in the literature is that method artifacts at least partially contribute to monomethod correlations among personality ratings in general (Campbell & Fiske, 1959), and ratings of the Big Five specifically (Biesanz & West, 2004; DeYoung, 2006; Paulhus & John, 1998). Thus, firm conclusions about the nature of Big Five correlations require a multitrait–multimethod (MTMM) examination.

The main assumption of MTMM analysis is that correlations based on the same method may be biased by shared method variance, whereas correlations across independent methods are less biased by method variance (Eid, Lischetzke, Nussbeck, & Trierweiler, 2003; Kenny & Berman, 1980). Even if the assumption of independent methods is too strict, cross-method correlations provide less biased information about the structure of personality traits than same-method correlations (Campbell & Fiske, 1959). Therefore, MTMM is a powerful tool to examine the structure and nature of correlations among Big Five ratings.

Researchers of two recent MTMM studies examined the effect of rater perspective on the correlations among the Big Five (Biesanz & West, 2004; DeYoung, 2006). Biesanz and West (2004) found average absolute correlations of .34, .38, and .33 among the Big Five in self-ratings and informant ratings by peers and parents, respectively (e.g., self-rating of N with self-rating of A, $r = -.36$). In contrast, the cross-method correlations were small and nonsignificant (e.g., self-rating of N and parent rating of A, r = .03). Thus, correlations among the Big Five appeared to be mostly due to biases in ratings of the Big Five by a single rater, and there was no positive evidence for valid correlations among the Big Five factors.

DeYoung (2006) conducted two sets of analyses. In a set of monomethod analyses, the CAB model fitted reasonably well to the data. However, multimethod analyses suggested that alpha and beta are orthogonal factors. This finding suggests that the correlation between alpha and beta reflects mostly biases unique to a single rater, whereas alpha and beta reflect higher order personality traits that produce covariations among the Big Five.

Whereas DeYoung (2006) and Biesanz and West (2004) disagreed about the presence of alpha and beta as valid higher order personality traits, their findings provide consistent evidence that the correlation between alpha and beta in the CAB model is predominantly due to biases that are unique to personality ratings by a single rater. In other words, both studies suggest that N, A, and C, on the one hand, and E and O, on the other hand, are practically orthogonal to each other and that correlations in ratings of N, A, and C with ratings of E and O are specific to a single rater.

In summary, any model that aims to explain correlations among Big Five ratings should explain several key findings. First, structural relations vary across studies, presumably due to the specific instrument used for the assessment of the Big Five. Second, correlations between alpha-related scales (N, A, and C) and beta-related scales (E and O) are only present in ratings by a single rater and fail to demonstrate convergent validity across raters. Third, alpha and beta emerge as additional factors and show convergent validity across raters in some studies (DeYoung, 2006). Subsequently, we propose a model that is consistent with these key findings.

The Halo-Alpha-Beta Model (HAB)

The results of the two MTMM studies are important, but hardly surprising. In fact, the results could have been predicted on the basis of Thorndike's (1920) seminal work on the *halo error* in person perception. Halo error reflects a disposition to attribute socially desirable characteristics to oneself or to somebody else (Campbell & Fiske, 1959; Thorndike, 1920).

Halo error is a textbook phenomenon, and there is no a priori reason to assume that it would not influence ratings of the Big Five. Indeed, halo provides a simple explanation for the positive sign of the correlation between alpha and beta and the findings that the correlation is only present in self-ratings but not in ratings across informants (Biesanz & West, 2004; DeYoung, 2006). Additional findings in the personality literature provide further evidence for this hypothesis. For example, John and Robins (1993) have shown that interrater agreement for ratings of the Big Five is higher for neutral items than for highly evaluatively desirable items. Other evidence indicates that interrater agreement can be increased by removing the evaluative content from the items (Konstabel, Aavik, & Allik, 2006). However, instructing participants to fake-good, rather than respond honestly, lowers the interrater agreement (Konstabel et al., 2006).

Typically, multiple methods are needed to separate method variance from trait variance. However, one important contribution of MTMM studies of the Big Five was to demonstrate that several of the Big Five scales show virtually no valid relations to each other. In this situation, it is possible to estimate method variance even if only data from a single rater are available, assuming that the results of MTMM studies can be generalized to other studies. The reason is that monomethod correlations of N, A, and C, on the one hand, with E and O, on the other hand, reflect mostly method variance because these corresponding cross-method correlations are approximately zero. Thus, building on MTMM studies, we developed a structural equation model for Big Five ratings by a single rater that identifies method variance due to halo effects in Big Five ratings. We call this model the *HAB model* (Schimmack et al., 2008, Figure 1).

In the HAB model, halo reflects evaluative biases in ratings by a single rater (Campbell & Fiske, 1959; Edwards, 1970; Horst, 1968; Paulhus, 1981). Alpha produces correlations among N, A, and C but is independent of E and O. Beta influences E and O but is independent of N, A, and C. Alpha and beta are assumed to reflect partially valid covariations among the Big Five domains, and partially additional biases. For example, Paulhus and John (1998) identified superheroes as individuals who inflate ratings of E and O and saints as individuals who inflate ratings of A and C.

The HAB model is very similar to the CAB model, and the two models are mathematically identical if additional constraints are imposed (i.e., if loadings of the Big Five factors on alpha, beta, or halo are constrained to be equal). Both models assume that some variance is unique to alpha and beta, whereas some variance is shared between all Big Five scales. If loadings of Big Five ratings on the factors are constrained to be equal, then both models have three degrees of freedom (three factor variances, or two factor variances and one correlation) and will fit an empirical data set equally well.

The main advantage of the HAB model is that it provides a theoretical account for the relation between alpha and beta. Moreover, the HAB model makes it possible to examine the construct validity of the postulated halo factor. In contrast, the CAB model makes it difficult to test theories of the correlation between alpha and beta. Thus, even though the presence of halo effects in personality ratings is hardly a groundbreaking observation (Campbell & Fiske, 1959), the HAB model is the first model that explicitly aims to distinguish halo biases from valid personality factors in monomethod studies of the Big Five. Moreover, our article is the first one to explicitly test a confirmatory model that recognizes the contribution of halo biases to self-ratings and informant ratings of the Big Five.

Estimating the HAB Model

Any structural equation model makes assumptions to move beyond the information that is originally provided by the data (Eid et al., 2003). With the availability of additional evidence, it is often possible to test or relax certain assumptions. A study that is limited to the 10 correlations among Big Five ratings has 10 observed parameters. In other words, the null model that assumes perfectly orthogonal Big Five dimensions has 10 degrees of freedom. The AB model with independent alpha and beta factors, free factor loadings, and no secondary loadings has five degrees of freedom (loadings of N, A, and C on alpha, and loadings of E and O on beta are estimated). The CAB model that adds a free parameter for the correlation among alpha and beta loses another degree of freedom and has four degrees of freedom.

With increasing parameters and decreasing degrees of freedom, models lose power to actually test a theoretical model. In other words, with decreasing degrees of freedom, the number of alternative models that can fit the data equally well increases, and good fit of one specific model becomes less informative without testing alternative models. For this reason, it is desirable to develop parsimonious models with few parameters. It is more impressive if a parsimonious model fits various data sets because it is easier to falsify a parsimonious model with few parameters than a complex model with many parameters and few degrees of freedom.

An unconstrained HAB model would have zero degrees of freedom, which practically guarantees good fit. Moreover, it is impossible to estimate separate factor loadings of E and O on beta if beta is unrelated to any other factor in the unconstrained model (DeYoung, 2006; Finch & West, 1997). For these reasons, we imposed a number of plausible constraints on the HAB model. Specifically, we constrained the factor loadings on halo, alpha, and beta to be equal. As a result, the model gains seven degrees of freedom. In other words, it tries to explain the 10 correlations among the Big Five with only three parameters that represent the shared variance among N, A, and C due to alpha, the shared variance among E and O due to beta, and the shared variance among all five dimensions due to halo. An additional advantage of the constraints is that it is possible to estimate factor loadings of E and O on beta, based on the assumption of equal contributions of beta to E and O.

The availability of seven degrees of freedom also provides some room to relax assumptions if a particular data set does not fit the data. This may be necessary given the heterogeneity of structural information across data sets observed earlier. The HAB model allows us to examine whether these post hoc modifications are unique to specific data sets or reveal more systematic problems of the model, which can be used to modify the model.

Setting equality constraints on factor loadings may be problematic if the proportion of measurement error variance varies between Big Five dimensions. For example, if one dimension contains more measurement error than others, equality constraints would overestimate the contribution of that dimension to the factor in question. However, it is possible to control for this type of error in our model either by using latent Big Five factors that represent reliable variance (when multiple indicators are available) or by adjusting the observed variance by an estimate of reliability (e.g., Cronbach's alpha, when only one indicator per dimension is available). This method ensures that constraints test the assumption that halo makes an equal contribution to the reliable variance in personality ratings. Moreover, reliability estimates of Big Five dimensions do not differ dramatically. As a result, even models that do not correct for random measurement error produce practically identical results, although they underestimate effect sizes.

The use of constraints also has advantages in terms of testing structural theories over models with numerous freely varying parameters. The main advantage is that models with unconstrained parameter estimates are often weak tests of a theoretical model. It is even possible that the model fits the data when the data are actually inconsistent with the theoretical model. For example, DeYoung's (2006) Mini-Marker data set for peer ratings showed weak E–O correlations ranging from .08 to .19. Moreover, these correlations are lower than correlations between scales that do not have a common meta-trait (e.g., correlations between A and O for peer ratings ranged from .23 to .26). Thus, the observed correlations are inconsistent with the theoretical CAB model. Nevertheless, DeYoung (2006) found reasonable fit for a confirmatory test of the CAB model. The apparent inconsistency between the pattern of observed correlations and the fit of the CAB model arises from the flexibility of a model with unconstrained factor loadings on alpha and beta. In this example, the unconstrained CAB model fitted the data by reducing the factor loading of E on beta (loadings ranged from .12 to .31, implying less than 10% explained variance), whereas O had high loadings (ranging from .57 to .67). In other studies, E loads highly on beta, and O has low loadings on beta (DeYoung et al., 2002). Without external correlates, beta is purely defined by the loadings of E and O on beta. Thus, it makes a big theoretical difference for the interpretation of beta whether E and O load equally, E loads more highly than O, or O loads more highly than E on beta. One advantage of a model with constrained loadings is that it forces a researcher to specify a priori the nature of beta.

In short, we propose that models with constrained coefficients have a number of desirable properties. The main advantage is that the model actually forces researchers to make explicit a priori predictions about structural relations that are then tested by the model. A good starting point for empirical studies without theoretical predictions about the strength of effects is to assume equal loadings because it is the most parsimonious model. Future studies may suggest improvements to this model if some modifications consistently improve fit.

Overview of Present Research

The main purpose of the present article was to subject the HAB model to a number of empirical tests. The main focus was on the novel halo factor that has not been examined in previous studies of Big Five correlations.

In Study 1, we conducted new MTMM analyses of Biesanz and West's (2004) and DeYoung's (2006) data sets as well as one new MTMM data set that fitted the HAB model to ratings of individual raters. We predicted that alpha and beta would show some convergent validity across raters because they have been shown to be present in the latent factors of the Big Five (DeYoung, 2006). We also predicted that the halo factor should show only negligible degrees of convergent validity across ratings, which would validate it as a bias factor unique to single raters. This is essentially a demonstration of discriminant validity. A bias factor of one rater should be unrelated to a bias factor of another rater.

In Study 2, we examined the convergent validity of the halo factor with an alternative bias measure. It is unlikely that the halo factor in the HAB model is specific to the Big Five ratings. Rather, we suspect that the Big Five ratings are influenced by a more general evaluative bias that may also influence ratings of other personal attributes (Schimmack et al., 2008). Thus, we expect that the halo factor in Big Five ratings will be positively correlated with other favorable ratings of the same individual. We tested this prediction in Study 2 by correlating the halo factor with a measure of biases in ratings of desirable attributes like intelligence and attractiveness.

In Study 3, we examined the relation between halo and self-esteem. Taylor and Brown (1988) proposed that overly positive self-evaluations are related to higher well-being and self-esteem. Thus, we expected halo to be more strongly related to self-esteem than to Big Five ratings. This finding would suggest that halo, at least partially, reflects evaluative biases in self-perceptions rather than mere rating biases.

In Study 4, we analyzed longitudinal data to examine the stability of halo in self-ratings of personality. On the basis of our hypothesis that halo is a bias in perceptions of personality traits that is related to stable personality characteristics like self-esteem, we predicted that halo is highly stable over short retest intervals. We tested this prediction by examining the contribution of stable and occasion-specific influences on halo over a period of several weeks.

Study 1: Convergent Validity of Halo in Big Five Ratings

The main aim of Study 1 was to test our prediction that the halo factor in Big Five ratings represents biases by demonstrating discriminant validity of halo factors of different raters. We conducted MTMM analyses in which we fitted the HAB model to Big Five ratings of each rater and examined the correlations across raters for alpha, beta, and halo. In MTMM analyses, factors that converge across raters are interpreted as valid variance, and factors that do not show convergent validity across raters reflect biases. Even if the strict assumptions of an MTMM analysis are violated, it is likely that convergent correlations across raters reflect more valid than bias variance, whereas the residual variance in mono-method correlations reflects more bias variance than valid variance (Campbell & Fiske, 1959). In this sense, low correlations between halo factors of different raters validate our interpretation of these factors as method factors. The results also provide first evidence of the ability of the HAB model to fit various data sets across studies.

Method

The fit of the HAB model was estimated to four data sets that contained self- and informant ratings of the Big Five. The first data set appeared in Biesanz and West's (2004) article. This data set included ratings by 256 participants who provided self-ratings and were also rated by a parent and an acquaintance, using 97 unipolar adjectives (Goldberg, 1992). The next two sets were provided by DeYoung's (2006) article, in which 483 participants and three of their peers completed the Big Five Inventory (BFI; John & Srivastava, 1999), and 487 participants and three of their peers completed Saucier's (1994a) Mini-Markers. Analyses of these three data sets were performed on covariance matrices on the basis of the published correlation matrices and standard deviations. The final analysis was based on raw data from a dyadic study of 226 spouses who provided self- and informant ratings of each other using a slightly modified version of the BFI (Schimmack, Oishi, Furr, & Funder, 2004). Throughout the present article, standard errors and confidence intervals associated with the standardized solutions are reported, as provided by MPlus5.

In order to obtain more accurate measures of the Big Five and to control for differences in reliability of Big Five ratings, the observed variances were adjusted by the reliability coefficients (i.e., Cronbach's alphas). and the HAB model was fitted to the reliable portion of the variance only. The adjustment was made by estimating the amount of error variance as the total variance (SD^2) minus the total variance times reliability ($SD^2 - (SD^2 \times \alpha)$). The HAB model was fitted individually to each type of rater. The Big Five residuals, alpha, beta, and halo were allowed to correlate with corresponding factors across raters (see Figure 1). Contrary to common MTMM studies, residual correlations for the unique variances in the Big Five reflect shared variance between raters due to valid unique variance in ratings of the Big Five.

The model fit was compared with generally prescribed values of good fit for three indices (Schermelleh-Engel et al., 2003; CFI > .95, RMSEA < .06, SRMR < .08). However, some researchers have cautioned against using these guidelines as strict cutoff criteria and have suggested that the context should play a larger role in interpreting the goodness of fit. For example, Kenny (2008) noted that CFI values tend to be low if the original correlations in the data set tend to be weak. Because the average correlations among Big Five ratings tend to be weak (<.20), CFI may underestimate overall model fit. Moreover, it was demonstrated earlier that correlation matrices are not identical across measures and samples. Thus, overall model fit is less relevant than the nature of inconsistencies between empirical data and the HAB model. For this reason, the source of discrepancies between actual correlations and theoretically predicted correlations when model fit did not met standard criteria of acceptable fit was carefully examined.

Results

Biesanz and West (2004)

The fit of the HAB model to Biesanz and West's (2004) MTMM data was below standard criteria for model fit for CFI and

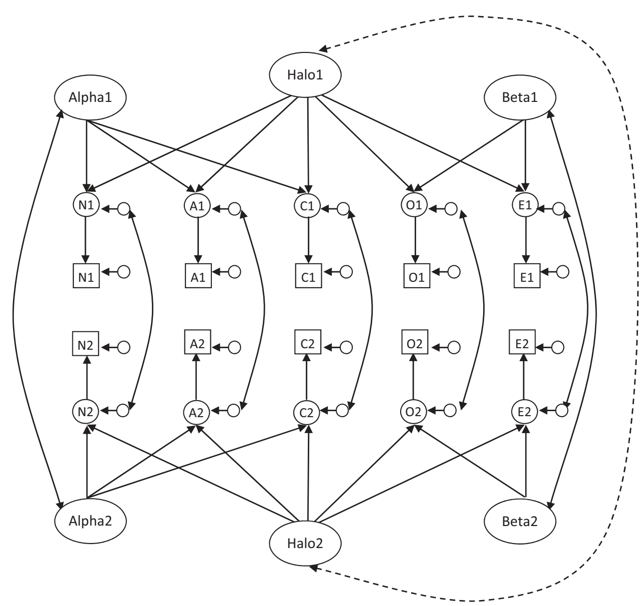

Figure 1. Multitrait–multimethod model for halo-alpha-beta model with two informants. Squares represent observed Big Five variances. Circles for the Big Five variables represent reliability-corrected variances. Open circles represent residual variances. The dotted line represents our prediction that the halo factors across informants will not be correlated. N = Neuroticism; A = Agreeableness; C = Conscientiousness; O = Openness; E = Extraversion.

RMSEA, but acceptable for SRMR, $\chi^2(78, N = 256) = 209.16$, CFI = .894, RMSEA = .081, SRMR = .072. Inspections of parameters revealed clear presence of alpha, beta, and halo for each rater. However, the correlations between alpha factors of different raters failed to provide positive evidence of convergent validity of alpha, although the confidence interval (CI) allows for moderate convergent validity ($r = .11$, $SE = .19$, 95% CI = −.25, .47). The model provided support for the validity of beta. The point estimate of convergent validity was surprisingly high, but the 95% CI for this parameter estimate had a wide range that includes more plausible parameters ($r = .90$, $SE = .29$, 95% CI = .34, 1.00). Thus, it is likely that actual convergent validity is lower than the point estimate in this sample. The most important finding was that the halo factors of different raters were unrelated to each other ($r = .08$, $SE = .07$) and that the 95% CI suggests that the true parameter is likely to be small, ranging from −.06 to .22 (see Tables 1 and 2). Due to the less than satisfactory model fit, we examined several possibilities to improve model fit. However, we found no major discrepancies, suggesting that the HAB model captured most of the structural relations relatively accurately.

In general, our results are consistent with Biesanz and West's (2004) original conclusions. However, some minor discrepancies are noteworthy. First, contrary to their analyses, we did find some evidence for a valid beta factor. The reason for the discrepancy is the greater statistical power of our model to find evidence for beta. Second, the CIs indicated that there could be some convergent validity for alpha but that the design lacks statistical power to provide strong evidence that alpha is not a valid personality factor. Most important, however, is the finding that halo is clearly present, and its convergent validity across raters is close to zero. This finding is fully consistent with Biesanz and West's (2004) original conclusions and supports our interpretation of the halo factor as a bias factor.

DeYoung (2006)

BFI. We fitted the standard HAB model to DeYoung's (2006) BFI data. Model fit was acceptable, $\chi^2(145, N = 483) = 314.51$, CFI = .954, RMSEA = .049, SRMR = .063. However, model fit could be improved considerably by relaxing the assumption of equal loadings on alpha, due to a higher observed A–N correlation in contrast with A–C and N–C correlations, a finding that is consistent with DeYoung's (2006) primary analyses. Thus, we relaxed this assumption by allowing C to have a different loading for all raters. The resulting model fit was excellent, $\chi^2(141, N = 483) = 202.47$, CFI = .983, RMSEA = .030, SRMR = .046. For all four raters, the model identified halo, alpha, and beta factors. Once more beta revealed convergent validity across raters, and the point estimate was surprisingly high (beta, $r = .98$, $SE = .10$, 95% CI = .79, 1.00). In addition, alpha factors of different raters showed convergent validity ($r = .48$, $SE = .05$, 95% CI = .38, .58). The most important finding was that halo was again unrelated across raters ($r = .03$, $SE = .08$, 95% CI = −.13, .18). Once more

Table 1
Factor Loadings of the HAB Model in MTMM Analyses

	Biesanz & West (2004)			DeYoung (2006) BFI			DeYoung (2006) Mini-Markers			Schimmack, Pinkus, & Lockwood (2004)		
Factor loading	Halo	Alpha	Beta	Halo	Alpha	Beta	Halo	Alpha	Beta	Halo	Alpha	Beta
Self												
N	.51	.29	—	.36	.46	—	.37	.45	—	.30	.32	—
E	.54	—	.29	.35	—	.37	.33	—	.20	.37	—	.43
O	.74	—	.39	.39	—	.41	.40	—	.24	.45	—	.52
A	.64	.36	—	.48	.60	—	.48	.58	—	.51	.54	—
C	.56	.32	—	.47	.23	—	.39	.07	—	.47	.50	—
Informant 1												
N	.53	.23	—	.39	.62	—	.40	.62	—	.41	.35	—
E	.54	—	.27	.43	—	.32	.43	—	—	.49	—	.39
O	.77	—	.38	.48	—	.35	.50	—	—	.49	—	.39
A	.59	.26	—	.44	.71	—	.45	.71	—	.56	.48	—
C	.57	.25	—	.51	.25	—	.43	.18	—	.48	.41	—
Informant 2												
N	.48	.51	—	.41	.60	—	.40	.64	—			
E	.44	—	.22	.42	—	.42	.42	—	—			
O	.61	—	.30	.49	—	.49	.49	—	—			
A	.53	.57	—	.46	.68	—	.47	.75	—			
C	.45	.49	—	.51	.35	—	.43	.17	—			
Informant 3												
N				.41	.65	—	.36	.71	—			
E				.47	—	.32	.38	—	—			
O				.53	—	.36	.48	—	—			
A				.45	.71	—	.40	.80	—			
C				.54	.38	—	.40	.26	—			

Note. Constraints were imposed on unstandardized coefficients. Standardized parameters vary slightly due to slight differences in variances in ratings of different Big Five dimensions. N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness; dashes represent paths that were constrained to zero.

Table 2
Cross-Informant Correlations (and 95% Confidence Intervals) of Halo, Alpha, and Beta Factors

Factor	Biesanz & West (2004)	DeYoung (2006) BFI	DeYoung (2006) Mini-Markers	Schimmack, Pinkus, & Lockwood (2004)
Halo	.08 (−.06, .22)	.03 (−.13, .18)	.10 (−.05, .24)	.05 (−.33, .43)
Alpha	.11 (−.25, .47)	.48 (.38, .58)	.38 (.29, .48)	.64 (.30, .97)
Beta	.90 (.34, 1.00)	.98 (.79, 1.00)		.74 (.29, 1.00)

Note. BFI = Big Five Inventory.

the CI indicates that this finding is not merely due to low statistical power, but due to an effect size close to zero.

Mini-Markers. Model fit of the standard HAB model was below standard criteria of model fit for CFI and at the cutoff point for RMSEA, $\chi^2(145, N = 487) = 471.31$, CFI = .903, RMSEA = .062, SRMR = .068. On the basis of DeYoung's (2006) findings, we relaxed the loading of on alpha. This minor modification lowered the loading of C on alpha and was sufficient to achieve acceptable model fit for all three fit indices, $\chi^2(141, N = 487) = 272.34$, CFI = .953, RMSEA = .044, SRMR = .050. Inspection of parameters revealed no evidence of beta in informant ratings. This finding is not surprising given the weak E–O correlations of informants that we discussed in the introduction. Thus, we dropped beta for informants and were unable to examine convergent validity of beta across raters. This model still had acceptable model fit, $\chi^2(145, N = 487) = 274.56$, CFI = .954, RMSEA = .043, SRMR = .051. This model revealed convergent validity for alpha ($r = .38$, $SE = .05$, 95% CI = .29, .48). Most important, the model showed a clear halo factor for all four raters with factor loadings ranging from .33 to .50, and the correlations among halo factors of different raters were again nonsignificant and close to zero ($r = .10$, $SE = .07$, 95% CI = −.05 to .24).

Schimmack, Pinkus, and Lockwood (2004)

The last data set is based on ratings provided by 113 spouse dyads. Each spouse provided self-ratings of personality and also rated the personality of her or his spouse. Because our analysis is not concerned with cross-partner relationships, we used all 226 individuals as the units of analysis. To account for the dependency of couple data, we used the cluster function in MPlus5. The cluster function applies corrections to standard errors and chi-square tests of model fit. With this function, standard errors are computed using a "sandwich" procedure, which assumes independence for couples but not for individuals within each couple. Moreover, low interclass correlations are not a threat to significance tests for interdependent data (Kenny, Kashy, & Cook, 2006), and previous studies have shown that personality similarity in dyads is low (Watson et al., 2004). For this reason, the standard errors were very similar in the model with and the model without a correction for dependence. The model fit of the standard HAB model was acceptable, $\chi^2(31, N = 226) = 38.00$, CFI = .984, RMSEA = .032, SRMR = .064. The model clearly identified halo, alpha, and beta in self-ratings and informant ratings. The model revealed convergent validity for alpha ($r = .64$, $SE = .17$, 95% CI = .30, .97) and for beta ($r = .74$, $SE = .23$, 95% CI = .29, 1.00). Most important, there was no agreement between spouses on the halo factor ($r = .05$, $SE = .20$, 95% CI = −.33, .43). Due to the small sample size and the use of only two raters, the CI in this study is large, but the halo results are consistent with those in our previous analyses.

Discussion

The main findings of all four MTMM analyses are summarized in Table 2. The halo factor was clearly identified in ratings by each rater in all four data sets, and correlations between halo factors of different raters were consistently close to zero. Because the already narrow CIs shrink even further when results are combined across studies, these findings provide strong support for the hypothesis that halo factors of individual raters predominantly reflect rater-specific biases. The results also provided evidence for alpha and beta as valid personality factors. Three of the four data sets revealed convergent validity for alpha and beta. The main problem was the lack of convergent validity for alpha in Biesanz and West's (2004) study. This could be a result of relatively low statistical power. We also failed to find evidence for beta in the Mini-Markers in DeYoung's (2006) study. This is likely due to unique characteristics of the Mini-Markers whose items were selected on the basis of producing a high degree of orthogonality among the Big Five dimensions (Saucier, 1994a) and may not generalize to other Big Five measures (DeYoung, 2006). The demonstration of some convergent validity for alpha and beta does not imply that these factors represent only valid personality traits. Some of the variance reflects additional biases that are unique to E and O as well as to N, A, and C. This finding is consistent with Paulhus and John's (1998) distinction of superheroes and saints who consider different personality traits desirable. In summary, the HAB provided reasonable fit to various data sets and integrates diverse findings in the literature.

Study 2: Halo as General Evaluative Bias

Study 1 provided evidence that the halo factor found in personality ratings does not converge across raters' perspectives, suggesting that this factor reflects rating biases. Study 2 aimed to provide convergent validity of our interpretation of the halo factor as a bias factor. For this aim, we included a newly developed measure of biases in self-ratings (Schimmack, 2008; Schimmack & Sidhu, 2007).

In Study 2, we were able to fit a full measurement model to the data at the item level because the data sets were substantially larger than those in Study 1. This method enables us to obtain more accurate estimates of the Big Five factors and is a more superior method to remove random measurement error than relying on Cronbach's alpha.

Additionally, modeling the Big Five at the item level also allowed us to separate another important bias factor: acquiescence bias. *Acquiescence bias* is the tendency to respond in the same manner on all questions (e.g., strongly agree with all items). Acquiescence bias is generally controlled by inclusion of positively and negatively worded questionnaire items. However, many scales do not have an even balance of these two types of items. Greater prevalence of one item type may further distort correlations. For example, Schimmack et al. (2008) surprisingly found lower loadings of Neuroticism on halo. One reason could be that acquiescence bias inflated the correlations among the other personality scales but attenuated the correlation with Neuroticism due to the unbalanced number of reverse-scored items.

Method

Participants

The analysis was based on raw data from two data sets collected at the University of Toronto at Mississauga over a period of 2 years (2006–2007). The data sets comprised introductory psychology students who voluntarily completed a questionnaire booklet at the beginning of each academic year. The questionnaire booklets included a personality inventory and an assessment of evaluative bias. Eight hundred twenty-two students in 2006 and 667 students in 2007 completed the questionnaire.

Measures

Personality. In 2006, students completed the Ten-Item Personality Inventory (TIPI; Gosling, Rentfrow, & Swann, 2003). TIPI asks participants to rate the extent to which each of the 10 pairs of traits (e.g., critical and quarrelsome) applies to them on a 7-point Likert scale ranging from 1 (*disagree strongly*) to 7 (*agree strongly*). Each of the Big Five dimensions is assessed with two trait pairs. Internal consistency of the subscales was quite low in the present sample, ranging from .20 to .60. The low consistency of this particular instrument has been observed in earlier literature and does not compromise the validity of the measure (Gosling et al., 2003). One explanation for the low internal consistency is that the correlation between the directly scored and the reverse-scored items are attenuated by acquiescence bias.

The 2007 questionnaire booklet included a short 15-item version of the BFI (BFI-15, Rammstedt, 2007; Schimmack et al., 2008). Participants were asked to respond to each item that started with the stem "I am somebody who . . ." on a 7-point Likert scale ranging from 1 (*disagree strongly*) to 7 (*agree strongly*). Neuroticism was assessed with descriptors "worries a lot," "gets nervous easily," and "is relaxed and handles stress well." For Extraversion, participants were asked to rate whether she or he is somebody who "is talkative," "is reserved," and "is outgoing." Openness to Experiences descriptors were "is original and comes up with ideas," "values artistic and aesthetic experiences," and "has an active imagination." Descriptors for Agreeableness were "has a forgiving nature," "is considerate and kind to almost everyone," and "is rude to others." Conscientiousness asked for ratings to items "is rather lazy," "does things efficiently," and "does a thorough job." Internal consistency for the five personality dimensions was low (range = .56–.74), but this has no practical implications for the present latent factor model that does not rely on Cronbach's alpha to estimate reliability.

Bias. Both data sets included the same bias measure. Participants provided self-ratings for four personal attributes: facial attractiveness, intelligence, athletic ability (i.e., long jump), and trivia knowledge. Ratings were made on a 7-point Likert scale ranging from 1 (*very bad*) to 7 (*very good*). The idea behind this bias measure is that the four attributes are relatively independent of each other. As a result, correlations among self-ratings of these four attributes reflect a general bias in self-ratings of desirable attributes (Campbell & Fiske, 1959). Using structural equation modeling, it is possible to estimate this latent disposition on the basis of the correlations among the four ratings. Schimmack and Sidhu (2007) validated the use of self-reported attribute ratings as a bias measure by demonstrating that the self-ratings were positively correlated, whereas the objective measures were independent. Furthermore, the composite score of the objective measures of the above attributes was not correlated with the Big Five (Schimmack, 2007). Thus, objective performance cannot account for the correlation between halo and the bias measure.

Modeling approach. The full measurement model was fitted for each data set at the item level. Thus, for the 2006 data set in which the TIPI was used as the personality measure, the Big Five were modeled as the latent variables, each defined by the two questionnaire items that assessed that particular dimension. Acquiescence bias was also modeled as a latent factor common to all 10 personality items. Factor loadings were constrained to be equal, under the assumption that acquiescence makes an equal contribution to all items. In the 2007 data set, which contained the 15-item BFI measure, each of the Big Five dimensions was modeled as a latent variable defined by three corresponding questionnaire items, and acquiescence bias was modeled as the latent factor defined by all 15 items. In each data set, an additional bias factor was modeled as a latent variable that represents the shared variance among the ratings of the four attributes (attractiveness, intelligence, athletic ability, and trivia knowledge). Consistent with previous findings, preliminary analyses revealed a relation between attractiveness and Extraversion (Borkenau & Liebler, 1992). For this reason, the final model also included a correlation between the residual variances of Extraversion and attractiveness. It should be noted, however, that correlated residuals between specific personality factors and specific attribute ratings mainly influence model fit but has negligible effects on theoretically important parameters (Green, Goldman, & Salovey, 1993).

For both data sets, the HAB model was fitted to the latent Big Five variables as in Study 1. In addition, the correlation between halo and bias was allowed to be freely estimated in order to demonstrate convergent validity between the two factors. Because it was assumed that alpha and beta represent valid personality dispositions, their correlations with the bias factor were constrained.

Results

The 2006 Data (TIPI)

Two fit indices (RMSEA, SRMR) suggested acceptable fit, whereas CFI suggested inadequate fit to the data, $\chi^2(72, N = 822) = 240.80$, CFI = .878, RMSEA = .053, SRMR = .049. However, the CFI value is influenced by the low correlations among the scale items (mean absolute $r = .14$). Indeed, comparison of observed and predicted cross-item correlations (see Table 3) did not reveal any substantial deviations.

Table 3
Actual (Lower Triangle) and Model-Predicted (Upper Triangle) Correlations Between the Items for the 10-Item TIPI Measure and Attribute Ratings in the 2006 Data Set (N = 822)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. E1	—	−.05	.20	−.09	.25	−.44	.21	−.05	.18	−.16	.26	.12	.09	.10
2. A1	.03	—	−.06	.12	−.06	.08	−.14	.09	−.06	.14	−.08	−.09	−.06	−.07
3. C1	.22	−.04	—	−.11	.23	−.04	.25	−.33	.21	−.11	.13	.14	.11	.12
4. N1	−.16	.27	−.08	—	−.10	.10	−.11	.11	−.37	.17	−.11	−.12	−.09	−.10
5. O1	.27	−.05	.22	−.17	—	−.08	.24	−.05	.20	−.23	.12	.14	.10	.11
6. E2	−.43	.04	−.05	.23	−.14	—	−.05	.08	−.04	.14	−.14	−.07	−.05	−.05
7. A2	.18	−.12	.25	.02	.21	.03	—	−.06	.22	−.12	.13	.15	.11	.12
8. C2	−.13	.07	−.34	.05	−.01	.07	−.13	—	−.05	.12	−.07	−.08	−.06	−.06
9. N2	.21	−.17	.26	−.38	.20	−.01	.17	−.15	—	−.10	.12	.13	.10	.11
10. O2	−.15	.11	−.11	.15	−.24	.16	−.16	.15	−.09	—	−.11	−.13	−.09	−.10
11. ATT	.29	−.02	.18	−.14	.16	−.12	.19	−.11	.23	−.16	—	.30	.22	.24
12. INT	.09	−.02	.14	−.11	.19	−.03	.05	.00	.18	−.14	.30	—	.24	.27
13. ATH	.14	−.07	.07	−.14	.10	−.09	.02	.02	.14	−.09	.23	.20	—	.20
14. TRI	.11	−.01	.08	−.12	.10	−.06	.05	.04	.16	−.11	.16	.32	.26	—

Note. E = Extraversion; A = Agreeableness; C = Conscientiousness; N = Neuroticism; O = Openness. E1–O1 and E2–O2 represent individual questionnaire items. TIPI = Ten-Item Personality Inventory; ATT = attractiveness; INT = intelligence; ATH = athletic ability; TRI = trivia knowledge.

The full model results can be seen in Figure 2. The model showed a weak but significant acquiescence bias (~ 4% explained variance in observed measures). Factor loadings on the halo factor were strong in this data set, with loadings ranging from .38 to .77 in magnitude. Beta was also identified. However, the model did not identify alpha as a reliable factor. Our measurement model of the attribute ratings showed a bias factor that explained 18%–34% of the variance in ratings of single attributes. Most important, the halo factor and the bias factor were highly correlated ($r = .59$, $SE = .06$, 95% CI = .48, .69). The last finding supports our hypothesis that halo in Big Five ratings reflects a more general evaluative bias in self-ratings.

The 2007 Data (BFI-15)

Figure 3 shows the model and its estimates for this data set. As in the previous data set, RMSEA and SRMR indicated acceptable fit, but CFI values were below the common standard for acceptable fit, $\chi^2(147, N = 667) = 444.78$, CFI = .896, RMSEA = .055, SRMR = .063. Again, with the small size of the average correlation between the items ($r = .13$), we believe that the CFI value underestimates model fit. Table 4 shows no substantial differences between actual correlations between items and correlations predicted by the model.

As in the 2006 data set, acquiescence bias accounted for about 4% of the variance in individual item responses. As in the 2006 data set, the model identified beta but failed to find evidence for alpha. Once more, halo was clearly present. Standardized loadings ranged from .31 to .42. The general bias factor explained 18%–44% of the variance in attribute ratings. Most important, we replicated the finding that halo and bias are substantially correlated ($r = .76$, $SE = .08$, 95% CI = .60, .93).

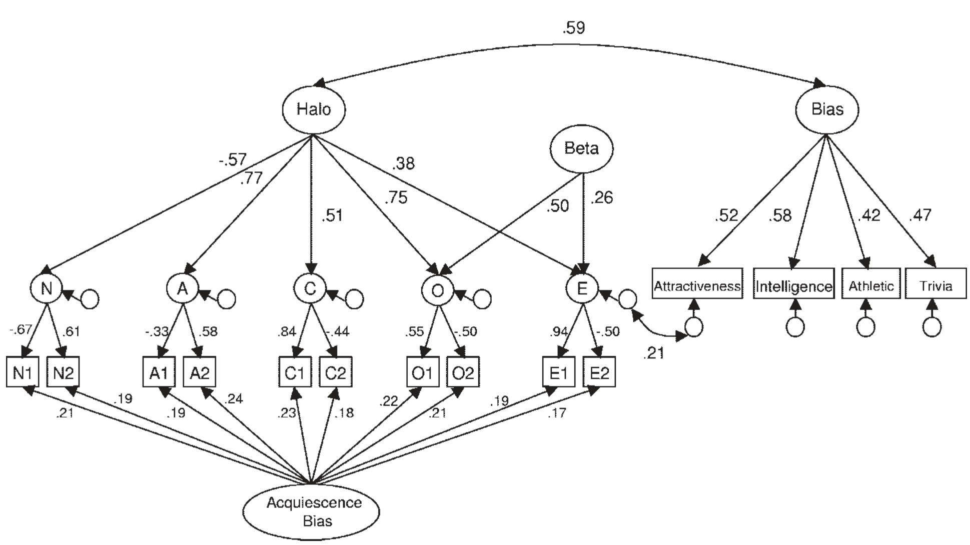

Figure 2. Measurement model for the 2006 data with the Ten-Item Personality Inventory. N = Neuroticism; A = Agreeableness; C = Conscientiousness; O = Openness; E = Extraversion. N1, N2, and so forth, represent individual questionnaire items.

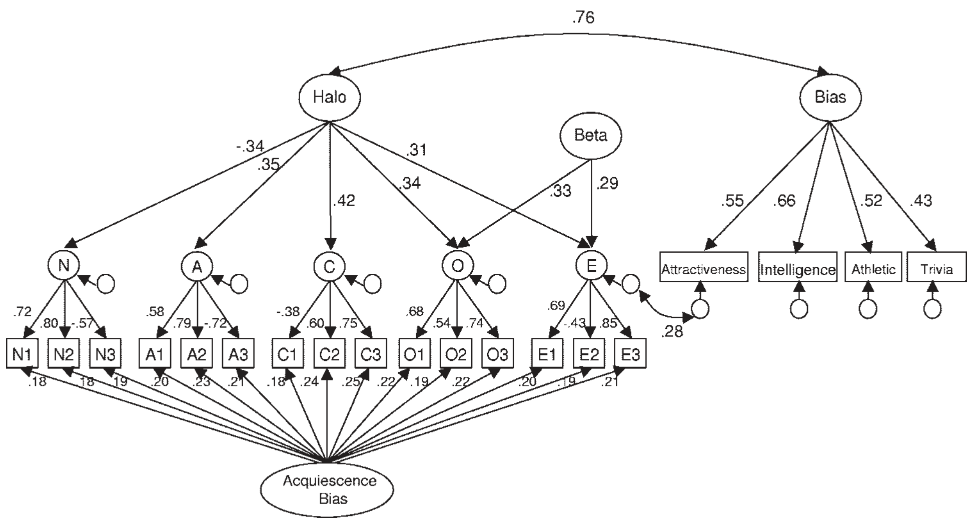

Figure 3. Measurement model for the 2007 data with the 15-item Big Five Inventory. N = Neuroticism; A = Agreeableness; C = Conscientiousness; O = Openness; E = Extraversion. N1, N2, N3, and so forth represent individual questionnaire items.

Discussion

The two data sets of Study 2 provided convincing evidence of association between the halo factor observed in the Big Five ratings and general evaluative bias seen in other self-ratings. Thus, people who tend to rate themselves as low in N, and high in E, O, A, and C, also rate themselves as more attractive, more intelligent, more athletic, and more knowledgeable in trivia. These findings suggest a global evaluative bias in self-ratings that influences ratings on personality questionnaires as well as ratings of other personal attributes. Any alternative theory that assumes that halo reflects a higher order personality trait would have to assume either (a) that this trait produces biases in perceptions of attractiveness and intelligence or (b) that meta-traits of personality are also related to other desirable dispositions.

Another important finding is that halo in personality ratings persists even when we statistically control for acquiescence bias. This finding further supports our interpretation of halo as a factor that reflects the content of items rather than a mere response style that reflects the use of rating scales. The analyses also provided

Table 4
Actual (Lower Triangle) and Model-Predicted (Upper Triangle) Correlations Between the Items for the 15-Item BFI Measure and the Attribute Ratings in the 2007 Data Set (N = 667)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. E1	—	−.26	.63	−.02	−.02	.08	.14	.11	.15	.08	.10	−.01	.00	.10	.12	.23	.11	.08	.07
2. E2	−.27	—	−.32	.07	.07	.01	−.02	−.01	−.02	.01	.01	.07	.06	.01	.01	−.15	−.07	−.05	−.04
3. E3	.62	−.31	—	−.03	−.03	.09	.16	.13	.17	.09	.12	−.02	.00	.11	.13	.29	.13	.10	.09
4. N1	.02	.20	−.07	—	.60	−.37	−.02	−.01	−.02	−.02	−.03	.10	.07	−.02	−.03	−.10	−.12	−.10	−.08
5. N2	−.05	.21	−.14	.62	—	−.42	−.02	−.02	−.03	−.02	−.03	.11	.08	−.02	−.04	−.11	−.14	−.11	−.09
6. N3	.04	−.04	.09	−.38	−.41	—	.09	.07	.09	.08	.10	−.01	.00	.09	.11	.08	.10	.08	.06
7. O1	.13	−.03	.20	.04	−.13	.21	—	.41	.56	.09	.11	−.01	.00	.11	.13	.10	.12	.09	.08
8. O2	.10	.04	.10	.12	.07	.03	.42	—	.45	.08	.09	−.01	.01	.09	.11	.08	.09	.07	.06
9. O3	.13	.04	.19	.14	.05	.05	.57	.48	—	.09	.12	−.02	.00	.12	.14	.11	.13	.10	.08
10. A1	.05	.09	.09	.09	.06	.10	.14	.15	.20	—	.51	−.38	.00	.10	.11	.09	.10	.08	.07
11. A2	.11	.12	.14	.12	.08	.02	.09	.18	.15	.48	—	−.53	.00	.12	.14	.12	.14	.11	.09
12. A3	.06	−.06	−.03	.00	−.04	−.02	−.05	−.11	−.03	−.35	−.54	—	.08	−.01	−.03	−.11	−.13	−.10	−.08
13. C1	−.01	.07	−.06	.10	.07	−.01	−.08	.00	.03	−.03	−.06	.16	—	−.18	−.24	−.07	−.08	−.06	−.05
14. C2	.10	.02	.17	−.01	−.04	.15	.18	.06	.11	.00	.04	.00	−.23	—	.51	.11	.13	.10	.08
15. C3	.15	−.03	.19	.15	.03	.07	.29	.19	.20	.11	.15	−.13	−.23	.53	—	.13	.16	.13	.11
16. ATT	.21	−.08	.27	−.06	−.14	.14	.14	.04	.09	.04	.05	−.01	−.15	.19	.24	—	.37	.29	.24
17. INT	.03	.02	.02	−.12	−.21	.17	.22	.08	.13	.02	.04	−.02	−.02	.17	.22	.38	—	.35	.29
18. ATH	.04	−.07	.14	−.17	−.22	.18	.17	.04	.13	.01	.07	.04	−.14	.18	.15	.32	.29	—	.23
19. TRI	.03	−.05	.00	−.13	−.16	.12	.18	.06	.13	.02	.03	−.03	−.01	.11	.11	.10	.34	.27	—

Note. BFI = Big Five Inventory; E = Extraversion; N = Neuroticism; O = Openness; A = Agreeableness; C = Conscientiousness. E1–E3, N1–N3, O1–O3, A1–A3, and C1–C3 represent individual questionnaire items. BFI = Big Five Inventory; ATT = attractiveness; INT = intelligence; ATH = athletic ability; TRI = trivia knowledge.

support for the presence of beta in self-ratings of personality traits, whereas the presence of alpha was not supported. This is probably a limitation of using brief measures of the Big Five that can only cover a narrow range of behavior. Even though our second measure was a short-form of the BFI, it may be necessary to use all BFI items to demonstrate alpha (DeYoung, 2006). The main finding of Study 2 was that we found clear evidence for the halo factor of the HAB model and demonstrated convergent validity with a second bias measure. Future research needs to examine whether this finding can be replicated with longer Big Five measures.

Study 3: Biases in Ratings or Self-Perception

The previous studies provided consistent evidence that halo largely represents evaluatively consistent biases in personality ratings and ratings of other attributes. In Study 3, we further examined the construct validity of halo by relating it to self-esteem ratings. Self-esteem ratings can be related to halo in two ways. First, self-esteem is desirable and should elicit the same rating biases as other desirable attributes (e.g., Agreeableness, attractiveness). Second, Taylor and Brown (1988) proposed that self-esteem is related to biased self-perceptions. Thus, self-esteem ratings should be more strongly related to halo than ratings on more specific desirable attributes.

In order to address this issue, we reanalyzed the data of an Internet survey with 326,641 respondents (Robins, Tracy, Trzesniewski, Potter, & Gosling, 2001). In this study, we assessed the Big Five with the complete BFI. Thus, Study 3 also provides another test of the fit of the HAB model to BFI ratings. We performed this analysis on the covariance matrix made available to us by Richard W. Robins. Cronbach's alphas were .83 for N, .85 for E, .78 for O, .79 for A, and .82 for C. As self-esteem was assessed with a single item, reliability of the self-esteem measure was assessed using retest correlations (reliability = .75; Robins et al., 2001). As in Study 1, we used these reliability estimates to adjust observed variances of the Big Five and self-esteem.

In addition, self-esteem was allowed to have unique relations with some specific personality dimensions on the basis of the regression results in the original article. Robins et al. (2001) found strong unique relations of self-esteem with Extraversion and Neuroticism. Thus, we allowed for these unique relations in our model. In addition, we allowed for an additional relation between Agreeableness and self-esteem because Agreeableness produced a theoretically important finding. Whereas Agreeableness ratings were positively correlated with self-esteem ratings, Agreeableness ratings were a negative predictor of self-esteem in a regression analysis with all Big Five scales as predictors. The HAB model has a simple explanation for this sign reversal. Agreeableness is negatively related to self-esteem. However, individuals high in self-esteem provide inflated ratings of Agreeableness because it is a desirable characteristic. To test this prediction, we also allowed for Agreeableness to be correlated with the halo-free, residual variance in Agreeableness, and predicted a negative relation.

The first model constrained the standardized loading of self-esteem on the halo factor to be the same as the standardized loadings of the Big Five dimensions. This model did not meet standard criteria of fit for two of the three criteria, $\chi^2(9, N = 326{,}641) = 18{,}853.161$, CFI = .918, RMSEA = .080, SRMR = .050. A second model relaxed the assumption of equal loadings on halo for self-esteem. Model fit increased significantly, $\chi^2(8, N = 326{,}641) = 10{,}274.325$, CFI = .955, RMSEA = .063, SRMR = .039. As predicted, the standardized loading of self-esteem on halo was much stronger (.77) than the loading of the Big Five dimensions (.38). The difference has a moderate effect size ($q = .62$). A model that reverses the path from self-esteem to halo fitted the data slightly better, $\chi^2(8, N = 326{,}641) = 9{,}066.48$, CFI = .960, RMSEA = .059, SRMR = .035. However, the data are insufficient to determine whether self-esteem ratings are more strongly influenced by halo or whether self-esteem produces halo biases in Big Five ratings.

In short, in Study 3, we extended the support for the HAB model. First, the HAB fitted yet another data set in which the BFI was used to measure the Big Five, suggesting that it is a robust model of the structural relations among the BFI scales. Second, Study 3 revealed that halo is strongly correlated with self-esteem. This finding suggests that the halo factor is not simply a rating bias, but also reflects people's self-evaluations. However, we acknowledge that alternative explanations are possible. For example, alpha and beta could both increase self-esteem. To test these competing hypotheses empirically, it is necessary to conduct MTMM studies, in which alpha and beta should produce cross-rater correlations of all Big Five scales with self-esteem, whereas halo should only produce correlations within raters.

Study 4: Stability of Halo

The previous studies presented evidence for the validity of the halo factor as a factor that reflects biases in self-perceptions of personality traits. In Study 4, we further examined the nature of halo by examining its temporal stability. Halo could be a momentary bias, or it could itself be a more stable personality characteristic. Study 3 suggests that halo is a stable personality characteristic because halo is strongly related to self-esteem, and self-esteem is a highly stable personality characteristic (Trzesniewski, Donnellan, & Robins, 2003). Moreover, Schimmack and Sidhu (2007) demonstrated stability of the bias measure that was highly correlated with halo in Study 2, over a 3-month retest interval. Thus, we predicted that a large portion of the variance in halo is stable over time. However, it is also possible that halo variance is partially influenced by occasion-specific fluctuations in self-perceptions or by more transient rating biases. Thus, we predicted that halo variance is both stable and variable but that a larger portion of the variance is stable, at least in the short term. We tested this hypothesis by applying the HAB model to a longitudinal study of the Big Five.

Method

To test whether halo is better conceptualized as occasion specific or stable over time, Biesanz and West's (2004) longitudinal data were analyzed with self-reports of the Big Five on three occasions. Biesanz and West (2004) asked participants to complete a 97-item set of Goldberg's (1992) unipolar adjectives three times, with up to 1 week between assessments.

The model was fitted to the covariance matrix on the basis of the correlations and standard deviations reported in Biesanz and West's (2004) article. No change in valid variance was expected over the three assessments because personality ratings are highly stable over time (e.g., Conley, 1984). Thus, the Big Five was modeled as stable factors across assessment occasions. As the HAB model assumes that alpha and beta reflect valid personality variance, alpha and beta were also modeled as factors that influence stable variance, but not occasion-specific variance. Self-esteem also tends to be highly stable over time (Trzesniewski et al., 2003). Thus, halo should also show some stability over the three assessments of Big Five. In addition, halo also reflects rating biases that may be more transient and specific to a single occasion. For this reason, additional halo factors were modeled at each occasion. The occasion-specific halo factors were independent of each other and independent of alpha, beta, and the stable halo factor. Moreover, it was assumed that the proportion of variance that may be attributed to the occasion-specific halo should not vary across measurement occasions. Thus, the loadings of the Big Five were constrained at each assessment to be equal (e.g., loading of Time 1 N on Time 1 occasion-specific halo = loading of Time 2 N on Time 2 occasion-specific halo).

Results and Discussion

Fit of the model with trait alpha, beta, and halo factors, and three independent occasion-specific halo factors, was acceptable for CFI and SRMR, but not the RMSEA, $\chi^2(86, N = 339) = 249.32$, CFI = .959, RMSEA = .075, SRMR = .080. Despite the less than ideal fit, inspection of reproduced correlations revealed no major discrepancies with the original data (see Table 5). Figure 4 shows the model results. The stable halo factor accounted for 18%–34% of the variance in observed Big Five ratings, whereas occasion-specific halo factors accounted for 5%–10% of that variance. This finding supports our hypothesis that the majority of halo variance is stable over time. The model also provided clear evidence for alpha and beta in this data set. The alpha factor accounted for 5%–8% of observed variance in personality ratings, and beta accounted for 9%–15%. In summary, the most important finding was that a substantial portion of halo variance is stable over short retest intervals. This finding shows that halo is not simply a transient rating bias. Moreover, Study 4 provided some additional evidence that the HAB model is a robust model that can fit various data sets reasonably well.

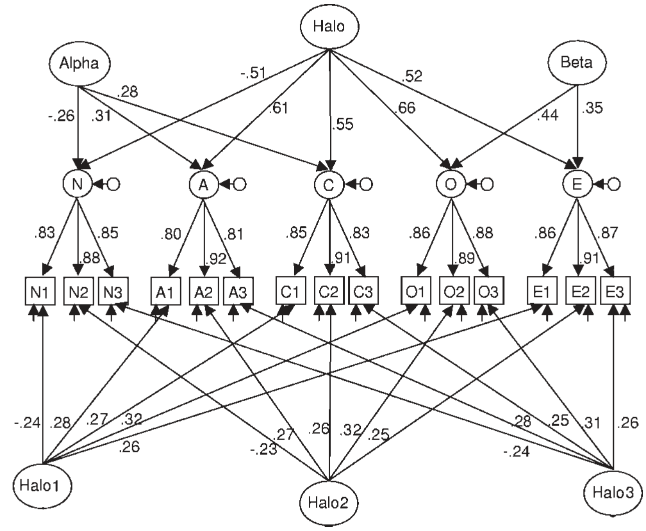

Figure 4. Stability model for the Biesanz and West's (2004) data. N = Neuroticism; A = Agreeableness; C = Conscientiousness; O = Openness; E = Extraversion. N1, N2, N3, and so forth represent observed scores at Time 1, Time 2, and Time 3, respectively.

General Discussion

In the present article, we introduced the HAB model as an integrative model that accounts for major findings in studies of the correlations among Big Five ratings. The prevailing theory of the Big Five correlations to date has been that they are caused by two independent higher order factors, alpha and beta. However, the AB model fails to provide adequate fit to mono-method studies of Big

Table 5
Actual (Lower Triangle) and Model-Predicted (Upper Triangle) Correlations Between the Big Five Measured at Three Occasions (From Biesanz & West, 2004; N = 339)

Big Five	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. A1	—	.36	.29	−.33	.37	.73	.31	.23	−.28	.28	.65	.28	.22	−.26	.28
2. C1	.40	—	.28	−.31	.35	.33	.77	.22	−.26	.27	.29	.70	.21	−.25	.27
3. E1	.22	.21	—	−.25	.45	.25	.22	.79	−.20	.38	.22	.20	.75	−.19	.38
4. N1	−.33	−.20	−.20	—	−.32	−.30	−.27	−.20	.73	−.25	−.26	−.24	−.19	.70	−.25
5. O1	.33	.44	.42	−.17	—	.32	.28	.39	−.26	.77	.28	.26	.37	−.25	.76
6. A2	.72	.33	.21	−.25	.31	—	.42	.33	−.38	.41	.74	.32	.25	−.30	.32
7. C2	.32	.77	.19	−.14	.40	.48	—	.30	−.34	.37	.31	.75	.23	−.27	.29
8. E2	.20	.20	.80	−.20	.40	.35	.28	—	−.27	.48	.23	.22	.80	−.20	.40
9. N2	−.26	−.14	−.16	.68	−.15	−.47	−.27	−.29	—	−.34	−.28	−.26	−.20	.75	−.26
10. O2	.28	.40	.38	−.10	.80	.43	.51	.46	−.16	—	.29	.27	.38	−.25	.78
11. A3	.64	.30	.21	−.14	.32	.73	.41	.26	−.28	.34	—	.35	.30	−.33	.37
12. C3	.24	.69	.13	−.10	.35	.34	.76	.16	−.16	.43	.44	—	.27	−.31	.34
13. E3	.20	.21	.75	−.13	.38	.28	.26	.80	−.20	.41	.34	.26	—	−.26	.46
14. N3	−.23	−.15	−.13	.66	−.13	−.26	−.16	−.16	.70	−.06	−.33	−.24	−.26	—	−.32
15. O3	.25	.34	.38	−.11	.78	.34	.41	.41	−.14	.82	.41	.46	.47	−.15	—

Note. A1–N1, A2–O2, and A3–O3 represent observed scores at Time 1, Time 2, and Time 3, respectively. A = Agreeableness; C = Conscientiousness; E = Extraversion; N = Neuroticism; O = Openness.

Five correlations. DeYoung et al. (2002) proposed a model with correlated alpha and beta factors as an alternative. However, no theory of the correlation between alpha and beta was offered. Moreover, the correlated AB model cannot explain why alpha and beta are correlated in monomethod studies but not across raters in multimethod studies. We tested the hypothesis that the reason for the positive correlation between alpha and beta in ratings by a single rater is an evaluative bias in personality perception. We created the HAB model to test this assumption.

In four studies, we examined the fit of the HAB model to various data sets in different samples using different measures of the Big Five. In Study 1, our MTMM analyses provided strong evidence for halo biases in Big Five ratings across four data sets. Moreover, the results generally provided evidence for alpha and beta as additional factors that produce convergent validity across raters. Although the goodness-of-fit indices were sometimes below the standard criteria of model fit, the consistency of the findings was more remarkable, especially given the discrepancies in conclusions reached by Biesanz and West (2004) and by DeYoung (2006). The discrepancies appear to be mainly due to a focus on significance testing. A focus on effect sizes suggests that all data sets are consistent with the hypothesis that alpha and beta are meta-traits that produce small to moderate covariations among the Big Five. Most important, however, all studies replicated the finding of halo factors in ratings of individual raters that were essentially unrelated to halo factors of other raters.

We extended the findings of Study 1 in Study 2 by demonstrating convergent validity of the rater-specific halo factor in Big Five ratings with another measure of evaluative bias in self-ratings. This finding provided further evidence for our interpretation of halo variance in Big Five ratings as evaluative bias variance. The correlations between halo and evaluative bias in self-ratings were moderate to high, indicating that a large proportion of the variance in the Big Five halo factor is due to a more global tendency to rate oneself in an evaluatively consistent manner. Moreover, in Study 2 we distinguished evaluative bias from acquiescence bias and found clear evidence that halo variance reflects evaluative content of personality items, although acquiescence had a significant, yet small effect on personality ratings.

In Study 3, we examined the relation between halo and self-esteem. As predicted, self-esteem was strongly correlated with halo. This finding suggests that self-esteem is partially based on overly positive self-perceptions (Taylor & Brown, 1988), but it is also possible that self-esteem ratings are more susceptible to evaluative rating biases. Future research needs to test these competing hypotheses using multiple methods. In Study 4, we demonstrated that halo biases are quite stable over short retest intervals. Studies 3 and 4 also provided further evidence that the HAB model is a robust structural model of correlations among Big Five ratings.

In summary, the HAB model provides a coherent theory of correlations among Big Five ratings that is consistent with diverse data sets and previous findings in the literature. The most important contribution of the HAB model is to explicitly recognize and model the influence of evaluative biases on self-ratings—and informant ratings—of Big Five personality domains. In addition, our studies provided further evidence about the nature of these biases. These biases generalize to ratings of other personality attributes, they reflect biases in self-perceptions, and they are relatively stable over time. In this regard, halo has all the properties of a personality trait. However, it is not a meta-trait that produces actual covariations among the Big Five personality domains. That is, it does not produce correlations in the disposition to be extraverted and conscientious. Rather, it produces covariations in self-ratings of Extraversion and Conscientiousness. In contrast, alpha and beta can be identified in ratings of personality traits and provide some evidence for convergent validity across raters. Thus, consistent with Digman's (1997) original theory, these factors represent meta-traits that produce actual covariations among Big Five factors. However, the effect size is relatively weak, which is consistent with other studies that control for measurement artifacts (Ashton et al., 2009). None of the present results would suggest that the Big Five can be reduced to a smaller number of basic personality dispositions without a loss of substantial variance. In this regard, factor loadings of the CAB model can be misleading because they combine the variance due to halo and the variance due to alpha and beta. These inflated factor loadings have led some researchers to propose that a smaller number of basic personality factors may be sufficient (Digman, 1997). Once the shared method variance is removed, however, most of the valid variation in one Big Five dimension is not shared with other Big Five dimensions. Thus, our results support the general practice in personality psychology to treat the Big Five as separate personality factors. In addition, we show how personality researchers can control for rating biases even in monomethod studies by including halo and acquiescence factors in the measurement models. This can be especially desirable in studies that rely on self-ratings to measure Big Five and outcome variables (Schimmack et al., 2008).

Limitations

It is important to emphasize that our results do not provide conclusive evidence that the HAB model is the best model to explain correlations among Big Five ratings. Model fit was not always acceptable, and even when it was acceptable, it was not perfect. It is possible that other models provide better fit to the data. Thus, an important avenue for future research is to test the HAB model against potentially better models. The main contribution of this article is to propose a strong and parsimonious theoretical model that provides a coherent account of all findings in the literature. Future research can use the HAB model as a benchmark to evaluate model fit. Alternative models should fit the data better than the HAB model, after adjusting for the number of degrees of freedom. Moreover, researchers should examine whether the improvement in fit generalizes to existing data sets because the structural information is not fully consistent across studies. The main advantage of the HAB model is that it is a robust model with few parameters and still fits surprisingly well to various data sets. However, model fit alone is insufficient to test structural theories. A main advantage of our studies was the inclusion of additional variables (self-informant ratings, a bias measure, a self-esteem measure) to test the HAB model.

Although we demonstrate that halo is related to self-esteem in self-reports of personality, the nature of halo in peer reports is still open to interpretation. It is likely that halo in peer reports also reflects informants' evaluations of the targets or relationship quality. For example, the shared variance among self-ratings of the Big Five as well as partner ratings of the Big Five accounted for over

50% of the explained variance in relationship quality (Holland & Roisman, 2008). Future research should further examine the nature of halo in self-ratings and informant ratings. It is important to use multiple methods and to include additional measures in these studies even if the main focus is to examine the structure of personality in order to separate effects of personality traits from effects of biased person perceptions.

Some of our findings also require replication and extension. The relation between halo and self-esteem needs to be examined in studies that assess self-esteem with multiple methods. The stability of halo needs to be examined over longer time periods. Finally, the fit of the HAB model, and the contribution of halo, alpha, and beta factors, in a wider variety of Big Five measures needs to be examined. The two data sets in Study 2 that examined halo's relation to another bias measure used very short measures of the Big Five. This correlation should be replicated using more conventional personality measures. Despite these limitations, the HAB model provides a solid foundation for future research on correlations among Big Five ratings because it integrates the main findings in the literature (Biesanz & West, 2004; DeYoung, 2006; Digman, 1997; Paulhus & John, 1998).

The Measurement of the Big Five

Another attractive feature of the HAB model is that it provides a measurement model for studies that aim to relate personality traits to potential causes or consequences of personality traits. The bulk of past studies and most future studies will likely rely on single-informant ratings (e.g., self or peer) to measure personality traits. The HAB model recognizes that this approach is sensible because single-informant ratings contain valid variance, but it also recognizes that single-informant ratings are not identical with personality traits. Most important, it allows researchers to separate halo biases in single-informant ratings of personality from variance in Big Five personality factors. Thus, we encourage researchers to use the HAB model as a measurement model when they examine the relation between single-informant ratings of the Big Five and outcome measures, especially if outcome measures are also rated by the same informant (Schimmack et al., 2008).

We recognize that some readers may be reluctant to follow our advice because the HAB model makes some explicit assumptions that may be false (e.g., equal loadings on halo, independence of alpha and beta). However, the common practice to rely on observed correlations and regression analysis provides only an illusory sense of certainty. The reason is that these methods also make assumptions about the relation between self-ratings of personality and the underlying personality traits, but these assumptions are hidden and often forgotten. One implicit assumption of simple regression analyses is that predictor variables (e.g., the Big Five) are measured without error. Another assumption is that the shared variance among the Big Five reflects true personality variance. Observed correlations and regression coefficients provide meaningful results only if this assumption is true, which is unlikely to be the case. Moreover, multiple R^2 values provide misleading information about the amount of explained variance because a large portion of explained variance is often due to the unspecified shared variance among the Big Five (e.g., Holland & Roisman, 2008). However, our results suggest that a large portion of this variance reflects rating biases (i.e., halo). Thus, the multiple R^2 values likely overestimate the contribution of true personality traits in explaining variance in the outcome variable.

Fortunately, the HAB model also implies that results of previous regression analyses provide fairly accurate estimates of the true relation between personality traits and self-report measures, as long as one focuses on regression coefficients. The reason is that regression coefficients reveal only the unique contribution of a Big Five dimension to an outcome variable. Thus, regression analyses automatically remove halo variance, which is shared across ratings of the Big Five from the regression coefficients. However, shared variance due to rating biases often leads to inflated estimates of explained variance (Schimmack et al., 2008).

Although regression analyses do provide meaningful results, the HAB model has several advantages over regression analyses. One advantage is that it is possible to examine whether alpha and beta are unique predictors of an outcome variable. For example, being adventurous could be more highly correlated with beta than with E and O. To test these hypotheses, it is necessary to explicitly model the shared variance in E and O as a beta factor. More elaborate structural models are also needed to separate the contribution of acquiescence bias and evaluative bias. Recent advances in software development have made it easier for personality psychologists to conduct confirmatory analyses and to specify measurement models (Muthén & Muthén, 1998–2007). We hope that personality researchers will take advantage of these advances.

References

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11,* 150–166.

Ashton, M. C., Lee, K., Goldberg, L. R., & de Vries, R. E. (2009). Higher order factors of personality: Do they exist? *Personality and Social Psychology Review, 13,* 79–91.

Biesanz, J. C., & West, S. G. (2004). Towards understanding assessment of the Big Five: Multitrait-multimethod analyses of convergent and discriminant validity across measurement occasion and type of observer. *Journal of Personality, 72,* 845–876.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin, 117,* 187–215.

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology, 62,* 645–657.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Conely, J. J. (1984). The hierarchy of consistency: A review and model of longitudinal findings on adult individual differences in intelligence, personality, and self-opinion. *Personality and Individual Differences, 5,* 11–25.

Costa, P. T., Jr., & McCrae, R. R. (1985). *The NEO Personality Inventory Manual.* Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1995). Solid ground in the wetlands of personality: A reply to Block. *Psychological Bulletin, 117,* 216–220.

De Young, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology, 91,* 1138–1151.

De Young, 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.

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology, 73,* 1246–1256.

Edwards, A. L. (1970). *The measurement of personality traits by scales and inventories.* New York: Holt.

Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (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.

Finch, J. F., & West, S. G. (1997). The investigation of personality structure: Statistical models. *Journal of Research in Personality, 31,* 439–485.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 2, pp. 141–165). Beverly Hills, CA: Sage.

Goldberg, L. R. (1992). The development of markers for the Big Five factor structure. *Psychological Assessment, 4,* 26–42.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48,* 26–34.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. de Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg, the Netherlands: Tilburg University Press.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big Five personality domains. *Journal of Research in Personality, 37,* 504–528.

Green, D. P., Goldman, S. L., & Salovey, P. (1993). Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology, 64,* 1029–1041.

Holland, A. S., & Roisman, G. I. (2008). Big Five personality traits and relationship quality: Self-reported, observational, and physiological evidence. *Journal of Social and Personal Relationships, 25,* 811–829.

Horst, P. (1968). *Personality: Measurement of dimensions.* San Francisco: Jossey-Bass.

John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality, 2,* 171–205.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality, 61,* 521–551.

John, O. P., & Srivastava, S. (1999). The Big Five taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102–138). New York: Guilford Press.

Kenny, D. A. (2008). *Measuring model fit*. Retrieved June 1, 2008, from http://davidakenny.net/cm/fit.htm

Kenny, D. A., & Berman, J. S. (1980). Statistical approaches to the correction of correlational bias. *Psychological Bulletin, 88,* 288–295.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis.* New York: Guilford Press.

Konstabel, K., Aavik, T., & Allik, J. (2006). Social desirability and consensual validity of personality traits. *European Journal of Personality, 20,* 549–566.

Musek, J. (2007). A general factor of personality: Evidence for the Big One in the five-factor model. *Journal of Research in Personality, 41,* 1213–1233.

Mutch, C. (2005). Higher-order factors of the Big Five model of personality: A reanalysis of Digman (1997). *Psychological Reports, 96,* 167–177.

Muthén, L. K., & Muthén, B. O. (1998–2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Author.

Paulhus, D. L. (1981). Control of social desirability in personality inventories: Principal-factor deletion. *Journal of Research in Personality, 15,* 383–388.

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66,* 1025–1060.

Rammstedt, B. (2007). Who worries and who is happy? Explaining individual differences in worries and satisfaction by personality. *Personality and Individual Differences, 43,* 1626–1634.

Robins, R. W., Tracy, J. L., Trzesniewski, K., Potter, J., & Gosling, S. D. (2001). Personality correlates of self-esteem. *Journal of Research in Personality, 35,* 463–482.

Saucier, G. (1994a). Mini-Markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment, 63,* 506–516.

Saucier, G. (1994b). Separating description and evaluation in the structure of personality attributes. *Journal of Personality and Social Psychology, 66,* 141–154.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research, 8,* 23–74.

Schimmack, U. (2007). [Objective measures of personal attributes is unrelated to evaluative bias]. Unpublished raw data.

Schimmack, U. (2008). The structure of subjective well-being. In M. Eid & R. J. Larsen (Eds.), *The science of subjective well-being* (pp. 97–123). New York: Guilford Press.

Schimmack, U., Oishi, S., Furr, R. M., & Funder, D. C. (2004). Personality and life satisfaction: A facet-level analysis. *Personality and Social Psychology Bulletin, 30,* 1062–1075.

Schimmack, U., Pinkus, R. T., & Lockwood, P. (2004). *Intrapersonal and interpersonal determinants of life satisfaction in married couples.* Unpublished manuscript.

Schimmack, U., Schupp, J., & Wagner, G. G. (2008). The influence of environment and personality on the affective and cognitive component of subjective well-being. *Social Indicators Research, 89,* 41–60.

Schimmack, U., & Sidhu, R. (2007). *Positive illusions and well-being: A new measure of evaluative bias in self-ratings.* Unpublished manuscript.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspectives on mental health. *Psychological Bulletin, 103,* 193–210.

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology, 4,* 25–29.

Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2003). Stability of self-esteem across the life span. *Journal of Personality and Social Psychology, 84,* 205–220.

Watson, D., Klohnen, E. C., Casillas, A., Nus Simms, E., Haig, J., & Berry, D. S. (2004). Match makers and deal breakers: Analyses of assortative mating in newlywed couples. *Journal of Personality, 72,* 1029–1068.

Received September 14, 2007
Revision received June 3, 2009
Accepted July 2, 2009 ■