4-2019

# Validity of evaluative factors from Big Five and HEXACO questionnaires

Michael D. Biderman
*University of Tennessee, Chattanooga*

Samuel T. McAbee
*Bowling Green State University*

Nhung T. Hendy
*Towson University*

Zhuo Job Chen
*Clemson University*, zhuoc@clemson.edu

Journal of Research in Personality xxx (xxxx) xxx-xxx

Contents lists available at ScienceDirect

Journal of Research in Personality

journal homepage: www.elsevier.com

Full Length Article

# Validity of evaluative factors from Big Five and HEXACO questionnaires

Michael D. Biderman [a, *], Samuel T. McAbee [b], Nhung T. Hendy [c], Zhuo Job Chen [d]

[a] University of Tennessee at Chattanooga, United States
[b] Bowling Green State University, United States
[c] Towson University, United States
[d] Clemson University, United States

ARTICLE INFO



Article history:
Received 27 July 2018
Received in revised form 10 January 2019
Accepted 26 April 2019
Available online xxx



Personality researchers have long recognized that measures of the Big Five or HEXACO domains are moderately correlated. For example, Block (1995) noted that "…the empirical research findings indicate that the five factors are frequently and importantly correlated with each other, usually to reflect an overriding evaluative component" (p. 199). When Neuroticism is reverse scored, these correlations generally have been found to be consistently positive, known as positive manifold (e.g., Musek, 2007; Saucier & Goldberg, 2003).

The average correlation comprising the positive manifold of summated score measures of the Big Five domains has been reported to be about 0.20 (Saucier, 2002). Some have argued that this positive manifold is the natural result of a higher-order organization of the processes representing the Big Five traits, with one or more higher-order processes governing Big Five domains (e.g., Digman, 1997; Musek, 2007). Others have considered the possibility that positive manifold is the result of some characteristic of the methodology upon which much of the evidence for the Big Five depends; specifically, a self-report bias that affects responses to items from all domains (e.g., Chang, Connelly, & Geeza, 2015). In either case, the existence of this positive manifold creates collinearity among the predictors for researchers and practitioners seeking the relations of external criteria to the Big Five or HEXACO domains. Thus, methods for modeling and exploring the causes of this positive manifold among personality traits are greatly needed.

Recent research using bifactor factor analytic models (e.g., Reise, 2012) has implications for the intercorrelations representing the positive manifold as well as for the number and nature of the characteristics represented by typical personality questionnaire data. In the simplest of such models, a general factor indicated by all items in the questionnaire is estimated in addition to the typical domain factors. Such models have been applied to individual item responses or parcel scores of Big Five and HEXACO inventories by several researchers (e.g., Anglim, Morse, De Vries, MacCann, & Marty, 2017; Bäckström 2007; Bäckström, Björklund, & Larsson, 2009; Bäckström, Björklund, & Larsson, 2014; Biderman, Nguyen, Cunningham, & Ghorbani, 2011; Biderman, McAbee, Chen, & Hendy, 2018; Chen, Watson, Biderman, & Ghorbani, 2015). These researchers have addressed three general issues. The first is goodness-of-fit – whether adding such a factor significantly improves the ability of factor analytic models to represent Big Five and HEXACO data. The second is the nature of the general factor – whether it is a method or substantive factor and if substantive, what aspect of item content is important for the factor. The third, and final, is the issue of the characteristics of the persons represented by the factor.

## 1. General factors in personality questionnaires

### 1.1. Need for general factors

With respect to the issue of goodness-of-fit, there is an accumulation of evidence that for multidimensional questionnaires in general, bifactor models often provide better fit than correlated factors models (Brunner, Nagy, & Wilhelm, 2012; Chen, West, & Sousa, 2006; Gignac, 2016; Markon, 2009; Reise, 2012; Yung, Thissen, & McLeod, 1999). Bifactor models of Big Five and HEXACO questionnaires have also been found to exhibit better goodness-of-fit relative to models without a general factor. Using a chi-square difference test Bäckström (2007) found that confirmatory factor analysis (CFA) models of parcels from an International Personality Item Pool (IPIP; Goldberg, 1999) 100-item questionnaire fit better when a general factor was included. Biderman et al. (2011) found that CFA bifactor models of individual item responses to IPIP 50-item questionnaires

* Corresponding author. Psychology/2803, University of Tennessee at Chattanooga, Chattanooga, TN 37403, United States.

Email addresses: michael-biderman@utc.edu (M.D. Biderman); smcabee@bgsu.edu (S.T. McAbee); nhendy@towson.edu (N.T. Hendy); zhuoc@clemson.edu (Z.J. Chen)

https://doi.org/10.1016/j.jrp.2019.04.010
0092-6566/ © 2019.

across five samples and to the NEO-FFI questionnaire (Costa and McCrae, 1989) in an additional sample fit better than models without a general factor using chi-square differences tests. Anglim et al. (2017) found that the fit of CFA models of the HEXACO 200-item questionnaire was significantly improved when a general factor was added to the models.

Based on chi-square difference tests, Biderman et al. (2018) found fit of CFA models to individual item responses to the NEO-FFI-3 and HEXACO 100-item questionnaires was significantly better when a general factor was included and improved even more when two factors - one indicated by positively-keyed items and one indicated by negatively-keyed items - were added to the models. They also applied EFA models to the same data to address the possibility that improvement in CFA models was compensation for the lack of cross loadings in CFA models of multidimensional data. They found that the addition of three factors (rotated to one general factor and two keying factors) resulted in significant improvements in goodness-of-fit, thus providing a necessary condition for the existence of general and keying factors in the EFA models. The body of analyses subsequently conducted by Biderman et al. (2018) provided evidence for the utility of treating such factors as general and keying factors. These results all support the belief that goodness-of-fit of models of Big Five and HEXACO questionnaire data is improved when general factors are added to models of the data.

### *1.2. Item content*

The second issue – that of the relation of the general factor to item content - has two aspects. The first is whether the general factor represents a substantive characteristic of items to which respondents are sensitive or an artifactual measurement characteristic (e.g., Davies, Connelly, Ones, & Birkland, 2015). Researchers have often labeled factors affecting items across domains such as that described above as method factors (e.g., Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Specifically, such general factors have been called unmeasured latent method factors (Johnson, Rosen, and Djurdjevic, 2011; Richardson, Simmering, & Sturman, 2009; Williams & O'Boyle, 2015; Williams & McGonagle, 2016) presumably to reflect the fact that they influence responses that have been obtained using a specific method from the same respondents but with no variables uniquely indicating those factors. The distortion in correlations resulting from not accounting for such factors frequently has been called common method bias (e.g., Doty and Glick, 1998). In factor analytic models of individual items, method factors' indicators have been estimated by setting the unstandardized loadings to be equal (e.g., Arias, Jenaro, & Ponce, 2018; Danner, Aichholzer, & Rammstedte, 2015). Restricting the loadings to be equal makes the influence of a factor independent of item content. In contrast, substantive method factors' loadings are estimated freely. Their influence thus depends on item content. A bifactor general factor can be treated in either way. If loadings are restricted to be equal, it will be characterized as a method factor. If the loadings of individual items on the factor depend on item content, it will be characterized as a substantive factor.

In the case of Big Five and HEXACO questionnaires, researchers have typically treated the general factors in bifactor models as substantive. This raises the second issue involving item content. That is the issue of identification of the specific aspect of item content important for general factors. The consensus from studies that have addressed this issue is that an evaluative aspect of item content indicates the general factors. Bäckström et al. (2009) found that loadings on the general factor of item-parcels written to have neutral evaluative content were significantly closer to zero than item-parcels based on originally written items. Pettersson, Turkheimer, Horn, and Menatti (2012) performed a Schmid-Leiman transformation of a higher-order factor model of quadruples of personality items similar to those in Big Five questionnaires (Schmid & Leiman, 1957). This transformation resulting in a general factor like that of a bifactor model but with proportionality restrictions on the factor loadings (cf. Gignac, 2016). They found that loadings on the general factor correlated 0.86 with mean evaluation ratings of the items and that items of similar valence but opposite content were clustered. Kulas and Stachowski (2012) applied a single factor model of responses to personality items and found that loadings on the factor were positively related to independently obtained social desirability ratings of the items. Bäckström and Björklund (2016) extended the Bäckström et al. (2009) study and found the mean loadings on a general factor of items written to have extreme valence were farther from zero than that of loadings of neutrally framed items. Biderman et al. (2018) showed that loadings of individual items on a general factor, on a positive keying factor, and on a negative keying factor were positively correlated with independently obtained judgments of whether an item's characteristics would make a person look "good" or "bad" for items on both the NEO-FFI-3 and HEXACO-PI-R questionnaires. Positive correlations with item valence were found in EFA models both within and across keying categories. The findings of the above referenced studies suggest that general factors are dependent on item evaluative content – item valence or social desirability.

### *1.3. Nature of respondents*

If loadings of items on the general factor are related to item evaluative content, this suggests that respondent characteristics may reflect item content (cf. Kenny and West, 2010). The issue of what individual characteristic(s) is/are represented by the general factors in bifactor models of personality has also been addressed by researchers. Bäckström (2007) found that the general factor in a CFA of the Big Five demonstrated strong positive correlations with scales measuring social desirability (i.e., self-deception and impression management). Biderman et al. (2011) found that the general factor correlated positively with the PANAS Positive Affectivity (PA) subscale and negatively with the Negative Affectivity (NA) subscale (Watson, Clark, & Tellegen, 1988) suggesting that the general factor represents level of respondent affect. Bäckström et al. (2014) found that a factor like the general factor in a bifactor model but indicated by differences between originally written and neutrally framed items correlated with social desirability. Biderman et al. (2018) found that factor scores from the general factor and keying factors from EFA analyses were positively related to the Rosenberg (1965) Self-esteem Scale and the PANAS PA scale, and negatively related to the Costello and Comrey (1967) Depression Scale and PANAS NA scale. These studies all suggest that persons at extremes of the general factor are persons with extreme levels of either affect or tendency to describe themselves in a socially desirable fashion depending on how the factor is ultimately characterized.

### *1.4. The present study*

Because the relations of loadings to valence were similar across the general and keying factors and because the general and keying factors exhibited similar relations to self-esteem, depression, and affect variables, Biderman et al. (2018) computed the average of factor scores of the general and keying factors and treated that average of factor scores as an overall measure of participant affective response to the evaluative content of items. Thus, they focused on only one characteristic based on the three factors, which they called an affect composite.

In contrast to Biderman et al. (2018), an alternative treatment of the general and keying factors is to estimate the general factor indicated by all items in the questionnaire after the negatively keyed items have been reverse scored and to estimate one other factor indicated positively by positively-keyed items and negatively by negatively-keyed items.[1] Estimated in this way, the general factor would reflect agreement with the positive aspect of whatever content was common across all items. The second factor would reflect the tendency to agree with items regardless of content. Such a tendency has been called acquiescence. Because a model consisting of a single general content factor and an acquiescence factor represents a more effective utilization of the information in the three factors than the single affect composite investigated in Biderman et al. (2018), we decided to use the two-factor model in this study. We called the general factor indicated without restriction by all items after reverse scoring the *evaluation* factor on the assumption that it reflects the evaluative aspect of item content. The factor indicated by differences in agreement with positively- vs negatively-keyed items was called the *acquiescence* factor.

In this study, we focused on the implications of the bifactor model described above of Big Five and HEXACO questionnaires data for high-stake selection situations. Previous conceptualizations of Big Five and HEXACO data have almost always represented personality data using correlated factors models, resulting in either five or six measures taken from the data - the typical summated scores of the Big Five or HEXACO domains. The application of bifactor models, on the other hand, results in estimating additional measures based on the general factors as well as the domain factors. We were interested in the relations of external variables to the new measures created by the application of the bifactor model – the evaluation factor and the acquiescence factor. We were also interested in whether measuring these additional factors would alter the relations of external variables to the domain measures.

We chose for this study the relation of Big Five and HEXACO data to university grade point average (GPA). Prediction of GPA is a topic of considerable research interest in the personality domain. There is considerable previous research on the Big Five (e.g., McAbee and Oswald, 2013; Poropat, 2009; Richardson, Abraham, & Bond, 2012), and some on the HEXACO questionnaires as predictors of GPA (e.g., De Vries, De Vries, & Born, 2011; McAbee, Oswald, & Connelly, 2014). Validity coefficients for Conscientiousness have been consistently found to be around 0.2, whereas validities for other domains have generally failed to meet acceptable levels of statistical significance (McAbee and Oswald, 2013). We applied bifactor models estimating the domain, evaluation, and acquiescence factors described above to data from three different questionnaires – the NEO-FFI-3 (Costa and McCrae, 1992), the HEXACO (Lee and Ashton, 2004), and the Big Five Inventory revised edition (BFI-2) (Soto & John, 2017). We compared simple and multiple-regression validities predicting GPA from summated scale scores – measures based on the correlated factors models of the questionnaires – with validities from factor scores computed from applying bifactor models to the same data. We included Structural Equation Model (SEM) estimates of validities in the comparisons.

Based on the previous research cited above, it was our belief that a measure based on the evaluation factor from bifactor models would represent overall affect levels of respondents. Although less commonly studied as predictors of GPA than measures of the Big Five or HEXACO domains, measures of affect have been found to demonstrate non-trivial relations with academic performance outcomes. For instance, in a *meta*-analysis of individual difference predictors of GPA, Richardson et al. (2012) reported that both optimism and self-esteem were positively related to GPA ($\bar{r}=0.11$ [95%CI=0.03, 0.17], and 0.09 [95%CI=0.05, 0.13], respectively). Moreover, the mean observed correlation for depression with GPA was −0.10 [95%CI=−0.17, 0.02], although this effect was not statistically significant ($k=17$, $N=6335$). In addition to the above research, there is evidence of the validity of respondent self-concept, particularly core self-evaluation, for prediction of academic performance (Judge, Erez, Bono, & Thoresen, 2002). Core self-evaluation has been found to positively correlate with both job satisfaction and life satisfaction (Judge, Erez, Bono, & Thoresen, 2003), which in turn are related to individual differences in affective states (e.g., Judge and Ilies, 2004).

One reason that positive affect might positively relate to academic performance outcomes is through the role of positive emotions in fostering engagement with the learning environment. According to Fredrickson (2004) Broaden and Build Theory, positive emotions broaden one's mindset, which increases the physical, social, and psychological resources available for people to improve their well-being and productivity. As an example of evidence supporting this positive emotion theory, in a comprehensive *meta*-analysis, Lyubomirsky, King, and Diener (2005) found that positive affect predicts a variety of desirable outcomes including job success, life happiness, and effective leadership. Other evidence suggests that core self-evaluation as measured by the Core Self-evaluation Scale (CSES; Bono and Judge, 2003; Chang, Ferris, Johnson, Rosen, & Tan, 2012; Judge et al., 2003) is positively related to job performance. These results all support the expectation that measures of affect—such as that represented by the evaluation factor estimated from bifactor models of the Big Five or HEXACO questionnaires – would be positively related to GPA.

Regarding the acquiescence factor, we are not aware of any research on direct associations between acquiescence in personality rating and GPA. There is, however, some research suggesting that cognitive ability is negatively related to acquiescence measured using a factor similar to the factor measured in this study (Lechner and Rammstedt, 2015; Zhou & McClendon, 1999). Based on these results we expect that GPA would be negatively related to the acquiescence factor.

Because the results discussed above suggest that the Big Five and HEXACO item responses are likely influenced by multiple respondent characteristics including domain, evaluative, and acquiescent tendencies, it follows that scale scores based on summing or averaging those responses may also be influenced by the same tendencies. Thus, it is possible that relations between GPA and the Big Five or HEXACO domain summated scales in previous studies have been contaminated by the evaluative and acquiescence influences on the summated scales thought to be primarily measuring domain-relevant variance. The consequences of controlling for those newly identified influences would depend on how each was independently related to GPA. If there were no relationship of GPA to the evaluative or acquiescence factors, then their presence as contaminating influences would simply act as "noise" to reduce the absolute magnitude of domain validities and controlling for these influences would result in positive validities becoming more positive and negative validities becoming more negative. On the other hand, if GPA were positively related to one or both extra factors, then controlling for these influences would result in validity of all measures of the domains becoming more negative than validities estimated without controlling for them.

[1] We would like to thank two anonymous reviewers for suggesting that the emphasis be on a general factor and an acquiescence factor.

## 2. Method

### *2.1. Samples*

Data for the NEO-FFI-3, HEXACO-PI-R, BFI-2, and official college GPAs were obtained from undergraduates at a medium-sized Southeastern public university in the U.S. The data were collected as three samples. The first sample was 73.3% female, and 69.0% White, 12.1% Black, and 18.9% other or declined to state. Mean age of Sample 1 was 19.91 years (*SD*=4.11). The second sample, from the same university, consisted of students who were not part of Sample 1, was 73.1% female, and 79.9% White, 12.2% Black, and 7.7% other or declined. Mean age for Sample 2 was 20.09 (*SD*=4.53). The third sample, again from the same university, contained students who were not part of Sample 1 nor of Sample 2. Sample 3 was 79.7% female, and 80.3% White, 11.1% Black and 8.6% other or declined to state. Mean age for Sample 3 was 24.05 (*SD*=3.81). Sample sizes were determined by availability of respondents.

We created four datasets from the three samples. In Sample 1, respondents took either the 60-item NEO-FFI-3, the 100-item HEXACO-PI-R, or both. Dataset 1A from Sample 1 contained data from 1377 respondents who took the NEO-FFI-3 questionnaire. Dataset 1B from the same sample consisted of 1597 respondents who took the 100-item HEXACO questionnaire. These datasets included 1307 respondents who took both questionnaires. Dataset 2 consisted of all 763 members of Sample 2, who took the 60-item HEXACO questionnaire. Dataset 3 consisted of all 916 members of Sample 3, who took the 60-item BFI-2 questionnaire.

Although there was considerable overlap between membership in Dataset 1A and 1B, Dataset 1B was used only for cross-validation of the results from Dataset 2 and was not treated as a primary dataset for evaluation of validities of the questionnaires. Thus, validities were evaluated separately on different respondents in Datasets 1A (NEO-FFI-3), 2 (HEXACO-60), and 3 (BFI-2), respectively.

### *2.2. Measures*

The questionnaires were online versions of the NEO-FFI-3 (Dataset 1A; Costa and McCrae, 1992), the 100-item HEXACO-PI-R (Dataset 1B; Lee and Ashton, 2004), the 60-item HEXACO-PI-R (Dataset 2; Lee and Ashton, 2004), and the BFI-2 (Dataset 3; Soto & John, 2017). Only the 60 items comprising the HEXACO-60 from the 100-item HEXACO-PI-R were analyzed. Participants were given no incentive or instructions to fake good or bad nor was any positive or negative outcome associated with responses to the questionnaires mentioned to participants. All items were answered on a 7-point scale for accuracy of self-description (1=*completely inaccurate*, 7=*completely accurate*). Items from the Neuroticism (NEO-FFI-3), Emotionality (HEXACO-PI-R), or Negative Emotionality (BFI-2) domains were reverse scored such that higher scores within each domain represented lower levels of these traits. Thus, rev-Neuroticism, rev-Emotionality, or rev-Negative Emotionality was the attribute represented by these items. For all domains, negatively-keyed items, i.e., items for which agreement would indicate less of the attribute, were reverse scored. Thus, for all domains, higher scores on items indicate greater amounts of the domain attribute and more positive self-evaluation. Internal consistency (α) reliabilities of summated scales are reported in Tables 1–4. Cumulative GPAs were obtained with students' permission from official college records at the end of the semester in which respondents took the personality questionnaire(s).

### *2.3. Analysis*

We used both EFA and CFA models in the analyses that follow. We performed EFA using the Exploratory Structural Equation Modeling (ESEM) procedure within Mplus in the belief that "… EFA usually results in more exact estimates of the true population values for the latent factor correlations than CFA" (Morin, Arens, & Marsh, 2016, p. 119). In addition to the reported superiority of EFA models in general, EFA was particularly important here because it could be argued that loadings on general factors in CFA models were simply substitutes for multiple cross-loadings that would be present in EFA models, i.e., that the general factor is simply capturing variance across domains that would otherwise be captured by cross loadings. Thus, we wanted to provide evidence for the existence and utility of general factors in models in which estimates of cross loadings were also obtained. At the same time, to evaluate the suggestion (cf. Asparouhov & Muthén, 2009) that CFA may force a researcher to specify a more parsimonious model than is suitable for the data, we were interested in the robustness of our findings with respect to the restrictions on loadings imposed in CFA. For these reasons CFA models were also applied. Results for all EFA and ESEM analyses are reported in the manuscript, and results of all CFA analyses are available in online supplementary tables.

We employed bifactor modeling techniques similar to those described in Biderman et al. (2018) (see also Arias et al., 2018) for all analyses. Specifically, an ESEM was applied to the data of each questionnaire (Asparouhov & Muthén, 2009; Browne, 2001; Marsh et al., 2010). We note that since each item could differ from every other item in both domain and evaluative content as well as keying direction, it would be very difficult to create parcels that were homogenous with respect to both domain and evaluative content as well as keying. For this reason, we applied the model to individual item responses. We modeled distinct factors for each of the five or six domains of the Big Five or HEXACO models, respectively. In addition to these domain factors, we included two additional factors in each model, which were rotated to an evaluation factor and an acquiescence factor. Thus, seven factors were estimated from the Big Five NEO-FFI-3 and BFI-2 data, and eight factors were estimated from the HEXACO data, respectively. For each model, all factors were estimated to be orthogonal as is common practice in applying bifactor models to personality data (see Morin et al., 2016; Reise, 2012).

A targeted rotation was applied to each solution (Browne, 2001). For the domain factors, loadings of items written to represent a domain were estimated freely on the primary domain factor but targeted at zero for all other non-primary domain factors. For example, Extraversion items were allowed to load freely on the Extraversion factor but targeted at zero on the other domain factors. Note that targeting estimates at a specific value did not set those values as fixed; rather estimates were allowed to vary about the targeted value. For the evaluation factor, loadings of all items after reverse-scoring the negatively keyed items were estimated freely. For the acquiescence factor, loadings of all positively keyed items were estimated targeted at 0.20,

**Table 1**
Correlations between NEO-FFI-3 scale and factor scores from bifactor model estimating evaluation and acquiescence factors (Dataset 1A).

| | | *Mean* | *SD* | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. | 13. | 14. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Scale Scores** | | | | | | | | | | | | | | | | |
| 1. | Extraversion | 4.69 | 0.85 | *0.85* | **0.26** | **0.34** | **0.37** | 0.03 | **0.68** | **0.22** | **0.82** | 0.01 | −0.02 | **0.12** | −0.05 | **0.60** | **0.13** |
| 2. | Agreeableness | 4.95 | 0.78 | | *0.79* | **0.31** | **0.17** | **0.20** | **0.63** | **−0.33** | 0.02 | **0.83** | −0.01 | −0.05 | **0.12** | **0.54** | **−0.17** |
| 3. | Conscientiousness | 4.83 | 0.86 | | | *0.87* | **0.36** | 0.01 | **0.69** | **0.22** | −0.05 | 0.01 | **0.76** | **0.09** | **−0.06** | **0.71** | **0.07** |
| 4. | Rev. Neuroticism | 3.94 | 0.91 | | | | *0.83* | **−0.13** | **0.62** | **−0.22** | **0.12** | −0.03 | **0.09** | **0.87** | **−0.09** | **0.44** | **−0.17** |
| 5. | Openness | 4.71 | 0.81 | | | | | *0.78* | **0.35** | **0.11** | −0.04 | **0.06** | **−0.10** | **−0.16** | **0.95** | **0.19** | **0.08** |
| 6. | Evaluation[1] | 4.62 | 0.50 | | | | | | *0.88* | 0.00 | **0.30** | **0.27** | **0.25** | **0.32** | **0.27** | **0.83** | −0.02 |
| 7. | Acquiescence[2] | 0.28 | 0.65 | | | | | | | *0.70* | **0.12** | **−0.31** | **0.16** | **−0.22** | 0.04 | **0.12** | **0.91** |
| | **Factor Scores** | | | | | | | | | | | | | | | | |
| 8. | Extraversion | – | – | | | | | | | | *0.88* | −0.02 | **−0.10** | 0.05 | −0.04 | **0.09** | 0.04 |
| 9. | Agreeableness | – | – | | | | | | | | | *0.85* | **−0.06** | −0.05 | 0.02 | **0.08** | **−0.07** |
| 10. | Conscientiousness | – | – | | | | | | | | | | *0.85* | 0.04 | −0.05 | **0.13** | 0.01 |
| 11. | Rev. Neuroticism | – | – | | | | | | | | | | | *0.90* | −0.02 | 0.04 | **−0.08** |
| 12. | Openness | – | – | | | | | | | | | | | | *0.89* | 0.04 | 0.03 |
| 13. | Evaluation | – | – | | | | | | | | | | | | | *0.93* | 0.04 |
| 14. | Acquiescence | – | – | | | | | | | | | | | | | | *0.88* |

*Note*. $N = 1377$. **Boldfaced** estimates are significantly different from zero, $p < .05$. Rev. = reverse coded. Diagonal elements in *italics* are coefficient alphas for scale scores and factor determinacies for factor scores. All loadings estimated freely on Evaluation factor. Positively keyed item loadings targeted at +0.2 and reverse−scored negatively keyed item loadings targeted at −0.2 on Acquiescence factor.

[1] Average of all items after reverse coding of negatively keyed items.

[2] Average of all items prior to reverse coding.

while loadings of all reverse-scored negatively keyed items were estimated targeted at −0.20.[2]

The full-length versions of the HEXACO-PI-R and NEO-PI-R each contain four or six facets per domain, respectively. We identified items within facets in the 60-item versions of these scales using information from the longer scales. In order to better represent the potential for dependencies among items from the same facet within each shortened scale, we let the residuals of each pair of items from the same facet correlate. This process was identical to that employed by Marsh et al. (2010) for the NEO-FFI questionnaire. The models for the BFI-2 (Soto & John, 2017) which contains three facets for each domain allowed residuals of all items within the same facet within each domain to be correlated. No across-facets or across-domains residual correlations were estimated in any model.[3] The model without residuals is shown in Fig. 1.

Summated scale scores for each domain were computed. Factor scores from the ESEM models were computed using the regression method in Mplus Version 8 (Müthén and Müthén 1998-2017). In addition to domain summated scores, two additional summated scores were computed. The first was the mean of all items in each questionnaire after the above-mentioned reverse scoring of the negatively-keyed items. This score is analogous to the total score computed from multidimensional questionnaires designed to measure one general characteristics (e.g., Rodriguez, Reise, & Haviland, 2016). It is herein referred to as the summated evaluation score. The second summated score was the mean of all items prior to reverse-scoring of the negatively-keyed items. This variable has been used as a measure of acquiescence (e.g., Danner, Aichholzer, & Rammstedte, 2015). It is herein referred to as the summated acquiescence score. Both were computed to explore the extent to which they were related to the evaluation and acquiescence factor scores computed for this research.

We applied the bifactor model to each primary dataset, computed factor scores, and entered the factor scores for the domain, evaluation, and acquiescence factors into a statistical package. Summated score and factor score simple and multiple regression validity coefficients were computed using the statistical package. We applied final estimates obtained from application of the ESEM model to Dataset 2 to the data from Dataset 1B to compute factor scores from that dataset. These factor scores were then used to predict GPA, and the validity results were compared with the results from Dataset 2 for the purposes of cross-validation. For the multiple regression results, we compared multiple regression models using summated scores as predictors, models using factor scores as predictors, and ESEM models using Mplus.

## 3. Results

Tables 1–3 present correlations among summated scales and factor scores for Datasets 1A, 2, and 3, respectively, along with reliability estimates for summated scales and factor determinacy values for factor scores. Table 4 presents correlations of summated scales from Dataset 1B with cross-validated factor scores computed using final estimates obtained from application of the bifactor model to Dataset 2. All of the domain summated scale reliability estimates were larger than 0.76, and all of the factor score determinacies were larger than 0.79. All of the convergent validities of domain summated scale scores with their corresponding factor scores in Tables 1–4 were positive and larger than 0.70. The convergent validities of summated evaluation scores with the evaluation factor scores were larger than

[2] As a sensitivity analysis on the appropriateness of 0.2 and −0.2 targets, we investigated the robustness of the results presented here when target values on the acquiescence factor were ±0.1 through 1.0. The results in Tables 5 and 6 were essentially the same for all choices of targeting value for the acquiescence factor. We chose ±0.2 as the target values to be close to the mean of loadings of items on the general factor, a value that might be expected of secondary loadings of items on a factor.

[3] It is possible to account for correlations among items from the same facet with separate facet factors rather than with correlated residuals. While models with facet factors rather than correlated residuals create more parsimonious CFA models, the opposite is true for EFA models, since the loading of every questionnaire item on every factor must be estimated. Adding facet factors to the EFA models would have resulted in a very large number of parameters to be estimated. We acknowledge the importance of modeling the domain~facet relations in these models; however, we felt that the complexity of such models was beyond the scope of the present research.

**Table 2**
Correlations between HEXACO scale and factor scores from bifactor model estimating evaluation and acquiescence factors (Dataset 2).

| | | *Mean* | *SD* | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. | 13. | 14. | 15. | 16. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Scale Scores** | | | | | | | | | | | | | | | | | | |
| 1. | Extraversion | 4.54 | 0.89 | *0.82* | **0.17** | **0.21** | **0.12** | 0.01 | 0.01 | **0.52** | **0.23** | **0.76** | −0.07 | 0.01 | **0.08** | −0.01 | **−0.16** | **0.66** | **0.13** |
| 2. | Agreeableness | 4.26 | 0.84 | | *0.77* | **0.09** | 0.01 | **0.09** | **0.31** | **0.55** | **0.10** | 0.00 | **0.89** | −0.05 | 0.00 | 0.06 | **0.16** | **0.40** | 0.04 |
| 3. | Conscientiousness | 4.77 | 0.83 | | | *0.79* | −0.05 | **0.12** | **0.28** | **0.54** | **−0.13** | 0.00 | −0.04 | **0.91** | **−0.07** | 0.06 | **0.08** | **0.42** | **−0.11** |
| 4. | Rev. Emotionality | 3.44 | 0.93 | | | | *0.81* | 0.06 | **−0.10** | **0.38** | **−0.25** | 0.04 | −0.01 | −0.06 | **0.96** | 0.03 | **−0.08** | **0.09** | **−0.15** |
| 5. | Openness | 4.31 | 0.98 | | | | | *0.80* | 0.05 | **0.49** | 0.00 | 0.00 | 0.05 | 0.09 | 0.05 | **0.97** | 0.03 | 0.05 | −0.02 |
| 6. | Honesty-Humility | 4.54 | 0.84 | | | | | | *0.72* | **0.50** | **−0.11** | **−0.16** | **0.17** | 0.14 | **−0.11** | 0.01 | **0.84** | **0.37** | **−0.08** |
| 7. | Evaluation[1] | 4.31 | 0.44 | | | | | | | *0.81* | −0.05 | **0.22** | **0.32** | 0.33 | **0.33** | **0.42** | **0.28** | **0.66** | −0.06 |
| 8. | Acquiescence[2] | 0.12 | 0.67 | | | | | | | | *0.67* | **0.18** | 0.04 | −0.07 | **−0.14** | 0.03 | −0.05 | 0.03 | **0.95** |
| | **Factor Scores** | | | | | | | | | | | | | | | | | | |
| 9. | Extraversion | – | – | | | | | | | | | *0.88* | 0.00 | −0.02 | −0.01 | 0.01 | **−0.09** | **0.08** | 0.04 |
| 10. | Agreeableness | – | – | | | | | | | | | | *0.84* | −0.06 | −0.02 | 0.02 | **0.10** | 0.04 | −0.02 |
| 11. | Conscientiousness | – | – | | | | | | | | | | | *0.91* | −0.04 | 0.03 | 0.05 | 0.07 | −0.02 |
| 12. | Rev. Emotionality | – | – | | | | | | | | | | | | *0.89* | 0.02 | −0.04 | 0.03 | −0.03 |
| 13. | Openness | – | – | | | | | | | | | | | | | 0.91 | 0.02 | −0.01 | 0.00 |
| 14. | Honesty−Humility | – | – | | | | | | | | | | | | | | *0.79* | 0.03 | −0.01 |
| 15. | Evaluation | – | – | | | | | | | | | | | | | | | *0.92* | 0.01 |
| 16. | Acquiescence | – | – | | | | | | | | | | | | | | | | *0.90* |

*Note*. $N = 763$. **Boldfaced** estimates are significantly different from zero, $p < .05$. Rev. = reverse coded. Diagonal elements in *italics* are coefficient alphas for scale scores and factor determinacies for factor scores. All loadings estimated freely on Evaluation factor. Positively keyed item loadings targeted at +0.2 and reverse-scored negatively keyed item loadings targeted at −0.2 on Acquiescence factor.

[1] Average of all items after reverse coding of negatively keyed items.

[2] Average of all items prior to reverse coding.

**Table 3**
Correlations between BFI-2 scale and factor scores from bifactor model estimating evaluation and acquiescence factors (Dataset 3).

| | | *Mean* | *SD* | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. | 13. | 14. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Scale Scores** | | | | | | | | | | | | | | | | |
| 1. | Extraversion | 4.49 | 0.88 | *0.84* | **0.18** | **0.27** | **0.33** | **0.22** | **0.64** | **0.17** | **0.85** | −0.06 | 0.01 | **0.10** | **0.07** | **0.48** | 0.04 |
| 2. | Agreeableness | 5.01 | 0.77 | | *0.81* | **0.39** | **0.24** | **0.25** | **0.63** | 0.04 | −0.05 | **0.79** | −0.02 | 0.01 | 0.06 | **0.62** | −0.05 |
| 3. | Conscientiousness | 4.82 | 0.89 | | | *0.88* | **0.33** | **0.15** | **0.69** | −0.01 | −0.06 | 0.01 | **0.71** | 0.04 | −0.06 | **0.74** | **−0.07** |
| 4. | Rev. Negative Emotionality | 4.12 | 0.95 | | | | *0.88* | 0.05 | **0.65** | **−0.10** | **0.10** | 0.00 | 0.03 | **0.89** | −0.06 | **0.45** | **−0.09** |
| 5. | Openness | 4.80 | 0.88 | | | | | *0.85* | **0.53** | **0.18** | 0.04 | 0.03 | −0.05 | −0.06 | **0.93** | **0.36** | 0.06 |
| 6. | Evaluation[1] | 4.65 | 0.55 | | | | | | *0.90* | **0.08** | **0.29** | **0.21** | **0.22** | **0.34** | **0.30** | **0.84** | −0.03 |
| 7. | Acquiescence[2] | 0.46 | 0.67 | | | | | | | *0.72* | **0.10** | −0.02 | **−0.15** | **−0.14** | 0.06 | **0.22** | **0.96** |
| | **Factor Scores** | | | | | | | | | | | | | | | | |
| 8. | Extraversion | – | – | | | | | | | | *0.90* | −0.03 | −0.03 | 0.03 | 0.01 | 0.04 | 0.02 |
| 9. | Agreeableness | – | – | | | | | | | | | *0.87* | **−0.07** | −0.04 | −0.01 | **0.10** | −0.02 |
| 10. | Conscientiousness | – | – | | | | | | | | | | *0.84* | −0.01 | −0.05 | **0.11** | −0.06 |
| 11. | Rev. Negative Emotionality | – | – | | | | | | | | | | | *0.90* | −0.04 | 0.06 | −0.05 |
| 12. | Openness | – | – | | | | | | | | | | | | *0.87* | 0.05 | −0.01 |
| 13. | Evaluation | – | – | | | | | | | | | | | | | *0.93* | 0.06 |
| 14. | Acquiescence | – | – | | | | | | | | | | | | | | *0.92* |

*Note*. $N = 916$. **Boldfaced** estimates are significantly different from zero, $p < .05$. Rev. = reverse coded. Diagonal elements in *italics* are coefficient alphas for scale scores and factor determinacies for factor scores. All loadings estimated freely on Evaluation factor. Positively keyed item loadings targeted at + 0.2 and reverse-scored negatively keyed item loadings targeted at −0.2 on Acquiescence factor.

[1] Average of all items after reverse coding of negatively keyed items.

[2] Average of all items prior to reverse coding.

0.66 across all datasets. Convergent validities of summated acquiescence scores with acquiescence factor scores were larger than 0.90.

Goodness-of-fit estimates and ESEM item factor loadings for all datasets are available in the online supplemental materials. ESEM goodness-of-fit values were uniformly acceptable with the largest RMSEA value equal 0.035 with 95% confidence limits of 0.033 and 0.037, the smallest CFI value equal 0.922, and the largest SRMR value equal 0.027. Chi-square difference tests indicated that beginning with a base model consisting of only domain factors, adding one or two additional factors resulted in significant improvements in goodness-of-fit.

Table 5 presents simple validity coefficients of summated scale scores and factor scores from the four datasets. Two general results are associated with the validities of the new factors estimated here – the evaluation factor and the acquiescence factor. First, the validity coefficient of the evaluation factor was positive and statistically significant for each dataset. For Datasets 1A, 2, and 3, the validity coefficients of the general factor were 0.20, 0.15, and 0.20, respectively, second to that of Conscientiousness for the Dataset 1A NEO-FFI-3 and Dataset 2 HEXACO 60 and largest for the Data 3 BFI-2. This result replicates previous research that variables representing affect are positive predictors of GPA (Richardson et al., 2012).

Second, the results of Table 5 show that the acquiescence factor by itself is a weak predictor of GPA. Although all validities were slightly negative, only the result from Dataset 3 met current standards of statistical significance. These results are not as strong as those of previous studies (e.g., Lechner and Rammstedt, 2015; Zhou & McClendon, 1999) who found negative relations between cognitive ability and acquiescence. One possible reason may be that since GPA is dependent on factors other than cognitive ability, the smaller relations observed here may be due to the influence of those other factors.

The results in Table 5 also show that validity estimates of the Big Five and HEXACO domains depend on whether the evaluation and acquiescence factors were estimated. In the table, 15 of the 16 factor scores domain simple validity coefficients from Datasets 1A, 2, and 3 were more negative than their summated scale counterparts. All 15 negative differences were statistically significant ($p < .05$) using Meng, Rosenthal, and Rubin (1992) test for dependent correlations. This result is consistent with what would be expected after controlling for the evaluation and acquiescence factors if the validity of the evaluation factor were positive and the validity of the acquiescence factor nearly zero.

Correlations from the evaluation and acquiescence factors computed from the ESEM model are in the last set of validities for each dataset. Those validities were quite similar to the factor score validities.

The last line of Table 5 presents simple validity coefficients of cross validated factor scores from data of Dataset 1B computed using final estimates obtained from applying bifactor model to Dataset 2. These follow the profile of validities of factor scores from Dataset 2. Moreover, as in the primary datasets, the simple validity of the cross-validated evaluation factor was positive ($p < .001$), second in size only to that of Conscientiousness. There was a small negative correlation between GPA and the cross validated acquiescence factor scores ($p < .01$).

The validities of the summated evaluation scores and the summated acquiescence scores were both comparable to those of the respective factor scores. For the summated evaluation scores, corresponding to the evaluation factor scores, the validities for the three samples were 0.15, 0.16, and 0.13 ($p < .001$ for each) for Datasets 1A, 2, and 3, respectively. For the summated acquiescence scores, corresponding to the acquiescence factor scores, the validities were 0.01, −0.02, and −0.05 ($p > .05$ for each), respectively.

Table 6 presents scale score, factor score, and ESEM multiple regression results for the three primary datasets. As was the case for simple validities, the partial regression coefficient for factor scores of the evaluation factor was statistically significant and positive for each questionnaire and was the largest factor score partial regression coefficient in Dataset 1 (NEO-FFI-3) and Dataset 3 (BFI-2). Thus, the evaluation factor exhibited not only simple validity but also provided incremental validity over all the domain factor scores in predicting GPA for each dataset. Comparison of individual domain partial regression coefficients of summated scales vs. those of factor scores

**Table 4**
*C*orrelations among HEXACO scale scores and cross-validated factor scores computed using Dataset 2 estimates for model estimating evaluation and acquiescence factors (*Dataset 1B*).

| | | *Mean* | *SD* | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. | 13. | 14. | 15. | 16. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Scale Scores** | | | | | | | | | | | | | | | | | | |
| 1. | Extraversion | 4.47 | 0.91 | *0.82* | **0.18** | **0.21** | **0.13** | 0.03 | **0.06** | **0.53** | **0.18** | **0.80** | −0.04 | 0.02 | **0.07** | 0.00 | **−0.12** | **0.68** | **0.14** |
| 2. | Agreeableness | 4.20 | 0.81 | | *0.77* | **0.16** | **0.12** | **0.15** | **0.32** | **0.61** | 0.01 | −0.02 | **0.89** | 0.03 | **0.09** | **0.12** | **0.16** | **0.44** | 0.05 |
| 3. | Conscientiousness | 4.69 | 0.85 | | | *0.81* | **−0.09** | **0.10** | **0.29** | **0.53** | **−0.15** | 0.04 | 0.03 | **0.92** | **−0.11** | **0.08** | **0.12** | **0.42** | **−0.05** |
| 4. | Rev. Emotionality | 3.44 | 0.90 | | | | *0.79* | **0.05** | **−0.05** | **0.39** | **−0.09** | 0.03 | **0.08** | **−0.11** | **0.96** | 0.00 | −0.04 | **0.13** | **−0.06** |
| 5. | Openness | 4.24 | 0.93 | | | | | *0.77* | **0.14** | **0.49** | −0.02 | 0.00 | **0.10** | 0.04 | 0.02 | **0.97** | **0.11** | **0.09** | 0.02 |
| 6. | Honesty-Humility | 4.44 | 0.86 | | | | | | *0.74* | **0.55** | **−0.21** | **−0.06** | **0.20** | **0.18** | **−0.09** | **0.12** | **0.85** | **0.35** | **−0.09** |
| 7. | Evaluation[1] | 4.25 | 0.45 | | | | | | | *0.83* | **−0.09** | **0.27** | **0.38** | **0.34** | **0.32** | **0.43** | **0.34** | **0.68** | 0.00 |
| 8. | Acquiescence[2] | −0.08 | 0.28 | | | | | | | | *0.51* | **0.16** | 0.00 | **−0.07** | 0.00 | −0.01 | **−0.15** | −0.04 | **0.95** |
| | **Factor Scores** | | | | | | | | | | | | | | | | | | |
| 9. | Extraversion | – | – | | | | | | | | | *0.88* | −0.04 | 0.00 | −0.03 | 0.00 | −0.05 | **0.18** | **0.07** |
| 10. | Agreeableness | – | – | | | | | | | | | | *0.84* | 0.00 | **0.06** | **0.07** | **0.10** | **0.10** | 0.00 |
| 11. | Conscientiousness | – | – | | | | | | | | | | | *0.87* | **−0.09** | 0.02 | **0.10** | **0.11** | 0.03 |
| 12. | Rev. Emotionality | – | – | | | | | | | | | | | | *0.89* | −0.04 | −0.03 | **0.07** | 0.03 |
| 13. | Openness | – | – | | | | | | | | | | | | | *0.91* | **0.11** | 0.04 | 0.02 |
| 14. | Honesty-Humility | – | – | | | | | | | | | | | | | | *0.79* | 0.04 | −0.04 |
| 15. | Evaluation | – | – | | | | | | | | | | | | | | | *0.92* | 0.03 |
| 16. | Acquiescence | – | – | | | | | | | | | | | | | | | | *0.90* |

*Note*. $N=1597$. **Boldfaced** estimates are significantly different from zero, $p<.05$. Rev.=reverse coded. Diagonal elements in *italics* are coefficient alphas for scale scores and factor determinacies for factor scores. To obtain cross-validated factor scores, the bifactor model was applied to Dataset 2. Final estimates from that application were then used to compute factor scores for Dataset 1B for the correlations shown in the table.

[1] Average of all items after reverse coding of negatively keyed items.

[2] Average of all items prior to reverse coding.

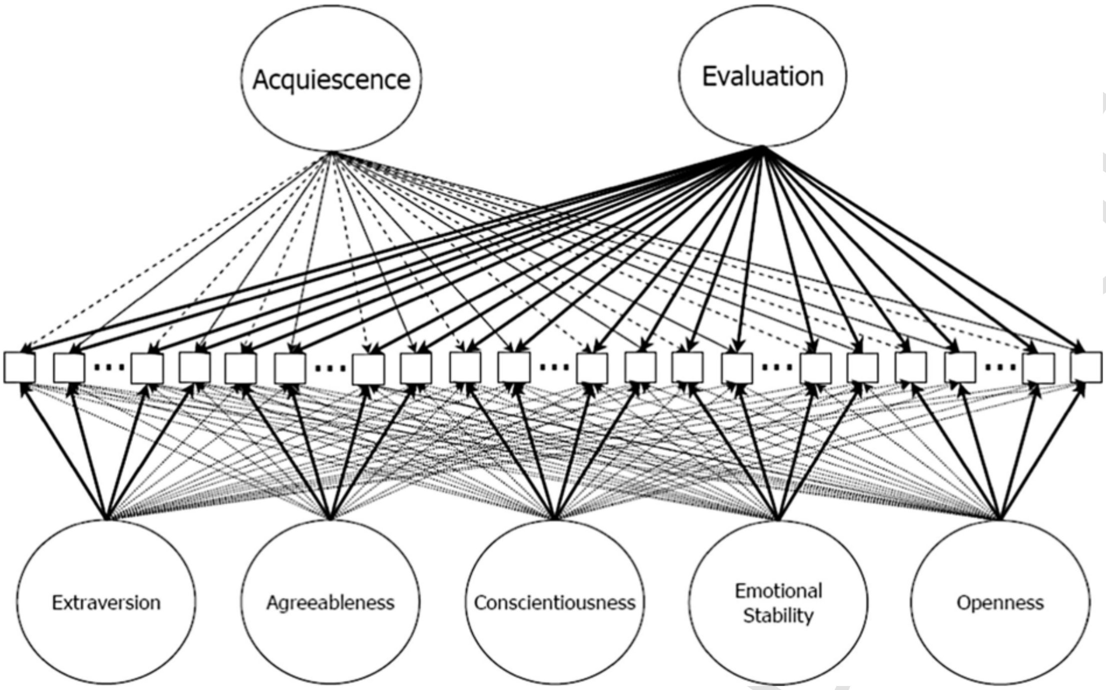


**Fig. 1.** Bifactor model applied to Big Five questionnaire.

**Table 5**
Validity coefficients predicting GPA from scale scores, factor scores, and cross-validated factor scores from the primary datasets.

| | Extraversion | Agreeableness | Conscientiousness | Rev. Neuroticism /Emotionality | Openness | Honesty-Humility | Evaluation | Acquiescence |
|---|---|---|---|---|---|---|---|---|
| **Dataset 1A** (*N* = 1377) | | | | | | | | |
| NEO Scale Scores | **0.03** | **0.14**[c] | **0.23**[c] | **0.02** | **0.02** | | | |
| NEO Factor Scores | **−0.09**[b] | **0.00** | **0.18**[c] | **−0.09**[b] | **0.00** | | 0.20[c] | −0.01 |
| NEO ESEM | −0.10[b] | −0.02 | 0.17[c] | −0.12[c] | −0.01 | | 0.21[c] | −0.03 |
| **Dataset 2** (*N* = 763) | | | | | | | | |
| HEXACO 60 Scale Scores | **0.08**[a] | **0.01** | **0.28**[c] | −0.04 | **0.03** | **0.12**[b] | | |
| HEXACO 60 Factor Scores | **−0.03** | **−0.05** | **0.24**[c] | −0.03 | **0.01** | **0.07** | 0.15[c] | −0.02 |
| HEXACO 60 ESEM | −0.03 | −0.06 | 0.26[c] | −0.03 | 0.01 | 0.08 | 0.15[c] | −0.02 |
| **Dataset 3** (*N* = 916) | | | | | | | | |
| BFI Scale Scores | **−0.01** | **0.10**[b] | **0.25**[c] | **0.02** | **0.05** | | | |
| BFI Factor Scores | **−0.08**[a] | **−0.01** | **0.17**[c] | **−0.07**[a] | **−0.01** | | 0.20[c] | −0.07[a] |
| BFI ESEM | −0.10[b] | −0.04 | 0.16[c] | −0.10[b] | −0.02 | | 0.22[c] | −0.08[a] |
| **Dataset 1B** (N = 1597) HEXACO 60 Cross-validated Factor Scores | −0.06[a] | −0.04 | 0.25[c] | −0.09[c] | 0.01 | 0.11[c] | 0.12[c] | −0.04[b] |

*Note*. **Boldfaced** estimates are significantly different from each other, $p < .05$. To obtain cross-validated factor scores for the last line of the table, the bifactor model was applied to Dataset 2. Final estimates from that application were then used to compute factor scores from the data of Dataset 1B.

[a] $p < .05$.

[b] $p < .01$.

[c] $p < .001$.

yielded the same pattern of results as did comparison of simple validities: Thirteen of 16 coefficients of the domain factor scores were more negative than their respective domain summated scales (two-sided $p < .05$ for a null hypothesis of probability = 0.5, $N = 16$).

Comparison of adjusted Multiple *Rs* of the scale score regressions and factor score regressions yielded small differences. The main difference between the two regressions was in the allocation of unique validity across the various predictors. The factor scores regressions showed that the characteristic represented by the evaluation factor is

**Table 6**
Multiple-regression partial regression coefficients from the primary datasets.

| | Extraversion | Agreeableness | Conscientiousness | Rev. Neuroticism/ Emotionality | Openness | Honesty-Humility | Evaluation | Acquiescence | *Adj. R* |
|---|---|---|---|---|---|---|---|---|---|
| **Dataset 1A** (*N*=1377) | | | | | | | | | |
| NEO Scale Scores | −0.05 | 0.08[b] | 0.25[c] | −0.07 | 0.00 | | | | 0.25 |
| NEO Factor Scores | −0.08[b] | −0.01 | 0.15[c] | −0.10[c] | −0.01 | | 0.19[c] | −0.02 | 0.27 |
| NEO ESEM | −0.10 | −0.02 | 0.17[c] | −0.12[c] | −0.01 | | 0.21[c] | −0.03 | 0.31 |
| **Dataset 2** (*N*=763) | | | | | 0.00 | 0.05 | | | 0.28 |
| HEX 60 Scale Scores | 0.03 | −0.04 | 0.26[c] | −0.03 | | | | | |
| HEX 60 Factor Scores | −0.03 | −0.05 | 0.23[c] | −0.03 | 0.01 | 0.04 | 0.14[c] | −0.02 | 0.27 |
| HEXACO ESEM | −0.03 | −0.06 | 0.27[c] | −0.03 | 0.01 | 0.08 | 0.15[b] | −0.02 | 0.31 |
| **Dataset 3** (*N*=916) BFI Scale Scores | −0.07[a] | 0.01 | 0.28[c] | −0.06 | 0.02 | | | | 0.26 |
| BFI Factor Scores | −0.08[b] | −0.03 | 0.14[c] | −0.09[b] | − 0.02 | | 0.20[c] | −0.07[a] | 0.27 |
| BFI ESEM | −0.10[b] | −0.04 | 0.17[c] | −0.10[b] | −0.02 | | 0.21[c] | −0.08[a] | 0.31 |

*Note.*
[a] $p<.05$.
[b] $p<.01$.
[c] $p<.001$.

uniquely related to GPA and that estimating this characteristic changed the estimates of how measures of the domains are related to the criterion.

Table 6 also presents results of ESEM regressions conducted within Mplus. The profiles of results of these regressions were quite similar to the factor score regressions. Not surprisingly, all coefficients in the ESEM regressions were more extreme – positive coefficients were larger and negative coefficients were more negative – than their factor score counterparts, consistent with the corrections for measurement error inherent to SEM. Finally, multiple R values for the ESEM analyses were larger than the factor score analyses. These differences are expected given that a small amount of indeterminacy is associated with factor scores that reduces the strengths of relations with other variables.

To test the limits of the extent to which estimates from Dataset 2 cross-validated to predictions based on multiple regressions, factor score multiple regression equation coefficients from the Dataset 2 analysis were used with the cross-validated factor scores computed from the data of Dataset 1B using Dataset 2 final estimates in a multiple regression analysis. This resulted in a two-step cross-validation – two-step because cross-validated regression equation coefficients were used with cross-validated factor scores. Since the cross-validated prediction equation was not optimized for the data to which it was applied, unadjusted Multiple *R*s were compared. The unadjusted factor score Multiple *R* for Dataset 2 was 0.29. The unadjusted two-step cross-validated factor score Multiple *R* for Dataset 1B was 0.30. These results suggest that the parameter estimates and regression equations from application of the bifactor model to one dataset can be used to create factor scores and regression equations in another holdout dataset that are essentially as predictive of a criterion as the factor scores in the original dataset, further increasing our confidence in the models applied.

*3.0.1. CFA Analyses.*

As mentioned above, as a check on the robustness of the results with respect to type of factor analytic model applied to the data, all of the above analyses were replicated using CFA models. Factor scores from the CFA models were analyzed in the same way as factor scores from the ESEM models. Traditional SEM analyses based on the CFA models were also conducted. The results of these analyses virtually mirror those in Tables 5 and 6, thus no elaboration of these analyses will be presented here (see supplemental tables).

## 4. Discussion

In this study, we investigated the implications of the use of factors from bifactor models of Big Five and HEXACO data in prediction of GPA. We estimated bifactor models which included an evaluation factor representing tendency to agree with all items in a questionnaire after the negatively-keyed had been reverse-scored and an acquiescence factor representing the difference in agreement with positively keyed items and reverse-scored negatively keyed items. We found that the evaluation factor positively predicted GPA. Moreover, the evaluation factor exhibited incremental validity over factors representing all of the Big Five or HEXACO domains and acquiescence. The acquiescence factor, on the other hand, exhibited weak simple and incremental validity relations with GPA. The results were replicated across three different samples and questionnaires suggesting that the evaluation factor in these models is a valid predictor of academic performance as represented by GPA.

We note that the characteristic represented the evaluation factor is a hidden aspect of the Big Five and HEXACO questionnaires. It is indicated by content found in varying degrees in all of the items rather than by a separate, identifiable set of items. Thus, the evaluation factor represents a relatively unobtrusive aspect of item content, one that respondents may not recognize as they respond to Big Five or HEXACO items. A summated score analog of the evaluation factor – the mean of items after reverse-scoring the negatively-keyed items – was correlated highly with the evaluation factor and also a modest predictor of the criterion used here. We argue, though, that analyses based on the evaluation factor from the bifactor model – factor scores or ESEM – are ultimately more useful. First, the usefulness of the summated evaluation variable as an estimate of the evaluation factor depends on the questionnaire items partitioning in approximately equal numbers into orthogonal domains. Without equal numbers of items in orthogonal domains these score values would be confounded with item domain content. Second, the summated evaluation variable would not be usable in a multiple regression analysis involving the summated domain scores since it is strongly correlated with each domain summated scale and is linearly dependent on the sum of those scores. The evaluation factor, on the other hand, is estimated as orthogonal to the domain factors and it can be included along with domain factors in multiple regression analyses predicting an external criterion. Finally, we believe that the bifactor model's evaluation factor represents a specific type of item content in a way that corresponds closely to the factor analytic theory underlying domain responses to questionnaires such as those we investigated. Thus, the evaluation factor estimated from the bifactor model is a more useful measure than the summated evaluation score from both a practical and theoretical standpoint.

A byproduct of the estimation of the evaluation factor and acquiescence factor is that their estimation accounted for systematic variation in items that would otherwise have been attributed to the domain factors in the analyses. This changed the estimates of the domain factors, resulting in those factors representing a relatively greater proportion of domain content of items. The result was that the bifactor model measures of the domains exhibited validities different from validities of measures of the same domains using summated scale scores. Because the contaminating influence on each response was positively related to GPA, removing that influence resulted in domain factor scores from the bifactor model being more negatively correlated with the GPA criterion than the original summated domain scores in 15 of 16 comparisons across three questionnaires.

Although the validity of Conscientiousness became less positive after controlling for the evaluation and acquiescence factors across three different questionnaires and samples, our results suggest that GPA remains positively and uniquely related to Conscientiousness. Our expectation is that Conscientiousness will continue to be a significant positive predictor of other, non-academic criteria even after controlling for the evaluative content of summated scale measures, although the effect of our modeling procedures on the validity of facets of Conscientiousness remains to be examined. (e.g., Dudley, Orvis, Lebiecki, & Cortina, 2006). While affirming the validity of Conscientiousness, our results yielded a different conclusion regarding the estimated validity of the other Big Five domains. After partialling the effects of the evaluation and acquiescence factors, all domain validities remained so close to zero that based on these results, none would be strongly recommended for prediction of GPA.

The fact that the evaluation factor itself was a valid positive predictor of GPA for all three questionnaires has implications for differences in validity of summated scales across questionnaires. Specifically, variation in the amount of evaluative content of items in different questionnaires may explain some of the variation in validity for predicting GPA of domains measured using summated scales (McAbee and Oswald, 2013). Based on our results, validities of domain summated scales consisting of items with greater evaluative content would be expected to be more positive than those in whose items evaluative content was minimal.

In the present study, the model applied to the data was similar to the model applied by Biderman et al. (2018) but different in specifics of the general factors. Biderman et al. (2018) estimated three factors – a general factor, a factor indicated freely by positively-keyed items and a factor indicated freely by negatively-keyed items. They averaged factor scores of the three factors into a variable that they labelled an affect composite. In the present study we estimated a model including a general factor as in Biderman et al. (2018) but estimated only one additional factor representing differences in agreement with positively-keyed vs negatively-keyed items. To provide an indication that the model applied here yielded results that were essentially the same as results that would have been obtained if the model of Biderman et al. (2018) had been applied, we computed the affect composite of the Biderman et al. (2018) models. That composite was very highly correlated with the evaluation factor in this study, with correlations of 0.96, 0.91, and 0.69 for Datasets 1A, 2, and 3, respectively. We also estimated a factor indicated freely by positively-keyed items and one indicated freely by negatively-keyed items and computed a difference variable that was the difference between the positive keying and negative keying factor scores. That difference variable was highly correlated with the acquiescence factor scores estimated here. The correlations across samples 1A, 2, and 3 were 0.97, 0.99, and 0.95, respectively. Finally, the analyses reported in Tables 5 and 6 were replicated using the Biderman et al. (2018) model. Those results, in the supplementary materials, are essentially identical to those in Tables 5 and 6. Thus, we are confident that the evaluation factor estimated in this study represents the same characteristic as the affect composite studied by Biderman et al. (2018), and that the acquiescence factor estimated here is essentially the difference between the two keying factors estimated in the previous study.[4]

We used the final estimates from application of the model to Dataset 2 to create factor scores from Dataset 1B. Those cross-validated factor scores yielded a pattern of simple validities with the criterion that was quite similar to the pattern shown in the original Dataset 2. In addition, we performed a multiple regression of the criterion onto the factor scores from the original Dataset 2 and used the coefficients from that regression analysis in a separate multiple regression of the criterion onto the cross-validated factors scores of Dataset 1B. This two-step cross-validation yielded predictive validity for dataset 1B that was slightly larger than the original validity from dataset 2. These results suggest that estimates from the bifactor model to one dataset are applicable to another dataset. This raises the possibility of publishing model estimates obtained from one application of the model and allowing others to use those estimates to generate useful predictions in other settings. Not only would predictions of GPA be useful, but scores on the factors such as the evaluation factor might be of use. For example, if subsequent research suggests that the evaluation factor is a measure of affect, scores of incoming students on that factor might be an indicator of college success. Students who score high on the evaluation factor would not only be better performing students, they would also be more likely to be effective leaders based on results of previous studies showing that people having a

[4] In application of the general + positive keying factor + negative keying factor model we encountered a condition called factor collapse (Geiser et al., 2015; Mansolf & Reise, 2016) in Dataset 1A and in Dataset 2. That condition was circumvented by targeting general factor loadings at 0.1 for those two datasets. No factor collapse was found in the evaluation factor + acquiescence factor models whose results are reported here.

positive affect tend to have others more likely to want to work for them (e.g., Bono and Ilies, 2006).

### 4.1. Limitations and considerations for future research

#### 4.1.1. Indeterminacy of factor size[5]

Although we believe there is strong evidence of the existence and importance of a factor indicated by evaluative content of all items within Big Five and HEXACO questionnaires, we acknowledge that we are less sanguine about the precision with which the evaluation factor can be estimated. A competition between the general factor and group factors for item variance is part of the process of estimation of loadings in bifactor models. It appears that certain profiles of item content within specific domains have the potential to result in a phenomenon called factor collapse (Geiser, Bishop, & Lockhart, 2015; Mansolf & Reise, 2016). In such instances, loadings of all items from a domain on the general factor are quite large at the expense of loadings of those items on the factor representing the domain from which the items came –a collapse of the domain loadings into the general factor. In such instances, relations of the general factor with external variables would mimic those expected for the domain factor whose loadings had been co-opted by the general factor. Although in some instances, factor collapse is quite noticeable, it is clearly possible that factor loadings on the general factor may be overestimated to an extent that may not be noticed.

Factor collapse can be controlled to a certain extent by targeting loadings on the general factor to a value deemed appropriate for loadings of items on a secondary factor. For example, targeting loadings of all items on the general factor at a value appropriate for a secondary factor, say, 0.20, would very likely prevent situations in which the loadings of items from one domain collapsed onto the general factor. But the manipulability of loadings thorough targeting means that targets for loadings on the general factor should be used with caution and that the researcher should be prepared to defend the targets chosen. Since factor size is dependent on loading values, manipulability of loading targets on the general factor raises means that the proportion of variance accounted for by the general factor can be manipulated. Such a possibility is worthy of further study but is beyond the scope of the present study. We note that no such loading targets were used in estimating the evaluation factor in the present study.

#### 4.1.2. Characterizing the evaluation factor

Although the evidence presented above suggests that item evaluative content is the key aspect indicating the evaluation factor, we note that factors whose indicators belong to multiple domains will reflect whatever content is common to those items and salient to respondents. The aspect of common item content that is salient likely depends on the context in which the measure is taken including any instructions or scenarios provided to respondents when completing personality questionnaires. For example, Anglim et al. (2017) compared CFA models estimating a single general factor estimated from data of the HEXACO questionnaire gathered under conditions in which there was a strong incentive to fake good with the same models applied to data in which there was no such incentive. These authors found that the difference between the condition in which there was an incentive to fake good and the condition in which there was little reason to do so was best represented by mean differences in the general factor between conditions as opposed to mean differences in domain factors. This result suggests that the context in which the questionnaire is administered influences all item responses, and these influences are reflected in the general factor in bifactor models. Research by Klehe et al. (2012) also provided evidence from both person- and item-level analyses that the general factor estimated from the NEO-FFI Big Five questionnaire reflects response inflation in faking conditions. Thus, any aspect of the context that would affect all item responses would affect the nature of the general factor. Others have suggested that information on what outcomes are to be measured might also influence results (Kleinmann et al., 2011; König, Melchers, Kleinmann, Richter, and Klehe, 2007). It is for these reasons that there was no incentive to fake nor was there mention of any specific overarching context and no mention of any criterion outcome in the present study.

As a check on the nature of the evaluation factor for the three datasets studied here, we correlated our evaluation factor scores with scores on the Rosenberg Self-esteem scale, one of the variables used by Biderman et al. (2018) to decide the nature of the affect composite, and which we collected alongside the personality questionnaires reported for all three samples tested in the present study. The correlations of evaluation factor scores in the present data with RSE scores were 0.64, 0.79, and 0.52 for Datasets 1A, 2, and 3, respectively. These provide evidence supporting the belief that the evaluation factor is strongly related to respondent affect. In contrast, the correlations of acquiescence factor scores with the RSE were 0.02, −0.02, and −0.05 ($p > .05$ for all) suggesting that the acquiescence factor represents a construct different from affect.

An additional issue regarding the nature of the characteristic represented by the evaluation factor is the issue of whether the respondent characteristic it represents is affect or a tendency to engage in socially desirable responding. Some investigators (e.g., Bäckström et al., 2009; Kulas and Stachowski 2012; Kuncel and Tellegen 2009), have labeled the evaluative content as “social desirability” implying that the respondent characteristic represented by the general factor is a tendency to respond in a socially desirable fashion. Other investigators (e.g., Biderman et al., 2018), have considered the same content as indicating not a tendency to emit socially desirable responses but affective levels of the respondents. As noted above, our position, based on the high correlations of the evaluation factor with measures traditionally labeled as measures of affect is that evaluation factor is, in fact, a measure of affective level of the respondent. Regrettably, however, we cannot clearly resolve these differences in conception in this study alone. The finding of significant relations between the evaluative content of personality items and consequential outcomes (e.g., GPA) makes such clarity an important endeavor for future research. Clearly, whatever the characteristic captured within respondents’ reactions to the evaluative content of personality items is called, this is an area worthy of further study.

#### 4.1.3. GPA as a criterion

It might be argued that the criterion employed in this research, GPA, is simply one of convenience, unrepresentative of selection criteria in organizational settings, diminishing the usefulness of results from a study such as this investigating only the prediction of GPA. We suggest that there is evidence of the value of GPA, for example, predicting future earnings of MBA students (Harrell and Harrell 1974). Moreover, there is evidence that GPA is relevant for job performance based on a large-scale *meta*-analysis (Roth, BeVier, Switzer, & Schippmann, 1996). For that reason, we believe that discovering the variables to which GPA is related is important and useful. With respect to the usefulness or generality of results involving the prediction of GPA, *meta*-analyses of the validity of Big Five questionnaires for job performance have shown patterns of validities similar to those found in this study. For these reasons, we believe that the parsing of validity of Big Five or HEXACO domain scores used

[5] We thank an anonymous reviewer for bringing up this issue.

here is a technique that would be useful in all high-stake selection situations.

## 5. Conclusion

The results of this study provide evidence that factors sometimes considered to be unwanted consequences of common method variance may in fact represent substantive personality characteristics worthy of study. Our results suggest that the evaluative and acquiescence factors represent unique characteristics of the respondents that are potentially useful predictors of valuable outcomes.

## Uncited reference

Yik et al. (2011).

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jrp.2019.04.010.

## References


Anglim, J., Morse, G., De Vries, R.E., MacCann, C., Marty, A., 2017. Comparing job applicants to non-applicants using an item-level bifactor model on the HEXACO Personality Inventory. European Journal of Personality 31, 669–684. https://doi.org/10.1002/per.2120.

Arias, V.B., Jenaro, C., Ponce, F.P., 2018. Testing the generality of the general factor of personality: An exploratory bifactor approach. Personality and Individual Differences 129, 17–23. https://doi.org/10.1016/j.paid.2018.02.042.

Asparouhov, T., Muthén, B., 2009. Exploratory structural equation modeling. Structural Equation Modeling 16, 397–438. https://doi.org/10.1080/10705510903008204.

Bäckström, M., 2007. Higher-order factors in a five-factor personality inventory and its relation to social desirability. European Journal of Psychological Assessment 23, 63–70. https://doi.org/10.1027/1015-5759.23.2.63.

Bäckström, M., Björklund, F., Larsson, M., 2009. Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. Journal of Research in Personality 43, 335–344, 0.1016/j.jrp.2008.12.013.

Bäckström, M., Björklund, F., Larsson, M., 2014. Criterion validity is maintained when items are evaluatively neutralized: Evidence from a full-scale five-factor model inventory. European Journal of Personality 28, 620–633. https://doi.org/10.1002/per.1960.

Bäckström, M., Björklund, F., 2016. Is the general factor of personality based on evaluative responding? Experimental manipulation of item-popularity in personality inventories. Personality and Individual Differences 96, 31–35. https://doi.org/10.1016/j.paid.2016.02.058.

Biderman, M.D., Nguyen, N.T., Cunningham, C.J.L., Ghorbani, N., 2011. The ubiquity of common method variance: The case of the Big Five. Journal of Research in Personality 45, 417–429. https://doi.org/10.1016/j.jrp.2011.05.001.

Biderman, M.D., McAbee, S.T., Chen, Z.T., Hendy, N.T., 2018. Assessing the evaluative content of personality questionnaires using bifactor models. Journal of Personality Assessment 100, 375–388. https://doi.org/10.1080/00223891.2017.1406362.

Block, J., 1995. A contrarian view of the five-factor approach to personality description. Psychological Bulletin 117, 187–215. https://doi.org/10.1037/0033-2909.117.2.187.

Bono, J.E., Judge, T.A., 2003. Core self-evaluations: A review of the trait and its role in job satisfaction and performance. European Journal of Personality 17, 5–18. https://doi.org/10.1002/per.481.

Bono, J.E., Ilies, R., 2006. Charisma, positive emotions and mood contagion. The Leadership Quarterly 17, 317–334. https://doi.org/10.1016/j.leaqua.2006.04.008.

Browne, M.W., 2001. An overview of analytic rotation in exploratory factor analysis. Multivariate Behavioral Research 36, 111–150. https://doi.org/10.1207/S15327906MBR3601_05.

Brunner, M., Nagy, G., Wilhelm, O., 2012. A tutorial on hierarchically structured constructs. Journal of Personality 80, 796–846. https://doi.org/10.1111/j.1467-6494.2011.00749.x.

Chang, L., Connelly, B.S., Geeza, A.A., 2015. Separating method factors and higher order traits of the Big Five: A meta-analytic multitrait-multimethod approach. Journal of Personality and Social Psychology 102, 408–426. https://doi.org/10.1037/a0025559.

Chang, C., Ferris, D.L., Johnson, R.E., Rosen, C.C., Tan, J.A., 2012. Core self-evaluations: A review and evaluation of the literature. Journal of Management 38, 81–128. https://doi.org/10.1177/0149206311419661.

Chen, F.F., West, S.G., Sousa, K.H., 2006. A comparison of bifactor and second-order models of quality of life. Multivariate Behavioral Research 41, 189–225. https://doi.org/10.1207/s15327906mbr4102_5.

Chen, Z., Watson, P.J., Biderman, M.D., Ghorbani, N., 2015. Investigating the properties of the general factor (M) in bifactor models applied to Big Five or HEXACO data in terms of method or meaning. Imagination, Cognition, and Personality: Consciousness in Theory, Research, and Clinical Practice 35, 216–243. https://doi.org/10.1177/0276236615590587.

Costa, P.T., McCrae, R.R., 1989. The NEO PI/FFI manual supplement. Psychological Assessment Resources, Odessa, FL.

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

Costello, C.G., Comrey, A.L., 1967. Scales for measuring depression and anxiety. The Journal of Psychology: Interdisciplinary and Applied 66, 303–313. https://doi.org/10.1080/00223980.1967.10544910.

Danner, D., Aichholzer, J., Rammstedte, B., 2015. Acquiescence in personality questionnaires: Relevance, domain specificity, and stability. Journal of Research in Personality 57, 119–130. https://doi.org/10.1016/j.jrp.2015.05.004.

Davies, S.E., Connelly, B.S., Ones, D.S., Birkland, A.S., 2015. The General Factor of Personality: The "Big One", a self-evaluative trait, or a methodological gnat that won't go away?. Personality and Individual Differences 81, 13–22. https://doi.org/10.1016/j.paid.2015.01.006.

De Vries, A., De Vries, R.E., Born, M.Ph., 2011. Broad versus narrow traits: Conscientiousness and Honesty-Humility as predictors of academic criteria. European Journal of Personality 25, 336–348. https://doi.org/10.1002/per.795.

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

Doty, D.H., Glick, W.H., 1998. Common methods bias: Does common methods variance really bias results?. Organizational Research Methods 1, 374–406. https://doi.org/10.1177/109442819814002.

Dudley, N.M., Orvis, K.A., Lebiecki, J.E., Cortina, J.M., 2006. A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. Journal of Applied Psychology 91, 40–57. https://doi.org/10.1037/0021-9010.91.1.40.

Fredrickson, B.L., 2004. The broaden-and-build theory of positive emotions. Philosophical Transactions of the Royal Society B: Biological Sciences 359 (1449), 1367–1377. https://doi.org/10.1098/rstb.2004.1512.

Geiser, C., Bishop, J., Lockhart, G., 2015. Collapsing factors in multitrait-multimethod models: Examining consequences of a mismatch between measurement design and model. Frontiers in Psychology 6, 1–15. https://doi.org/10.3389/fpsyg.2015.00946.

Gignac, G.E., 2016. The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. Intelligence 55, 57–68. https://doi.org/10.1016/j.intell.2016.01.006.

Goldberg, L.R., 1999. A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In: In: Mervielde, I., Deary, I., De Fruyt, F., Ostendorf, F. (Eds.), Personality Psychology in Europe Vol. 7, Tilburg University Press, Tilburg, The Netherlands, pp. 1–28.

Harrell, T.W., Harrell, M.S., 1974. Relations of second year MBA grades to business earnings. Personnel Psychology 27, 487–491. https://doi.org/10.1111/j.1744-6570.1974.tb01170.x.

Johnson, R.E., Rosen, C.C., Djurdjevic, E., 2011. Assessing the impact of common method variance on higher order constructs. Journal of Applied Psychology 96, 744–761. https://doi.org/10.1037/a0021504.

Judge, T.A., Erez, A., Bono, J.E., Thoresen, C.J., 2002. Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct. Journal of Personality and Social Psychology 83, 693–710. https://doi.org/10.1037/0022-3514.83.3.693.

Judge, T.A., Erez, A., Bono, J.E., Thoresen, C.J., 2003. The core self-evaluations scale: Development of a measure. Personnel Psychology 56, 303–331. https://doi.org/10.1111/j.1744-6570.2003.tb00152.x.

Judge, T.A., Ilies, R., 2004. Affect and job satisfaction: A study of their relationship at work and at home. Journal of Applied Psychology 89, 661–673. https://doi.org/10.1037/0021-9010.89.4.661.

Kenny, D.A., West, T.V., 2010. Similarity and agreement in self- and other perception: A meta-analysis. Personality and Social Psychology Review 14, 196–213. https://doi.org/10.1177/1088868309353414.

Klehe, U., Kleinmann, M., Hartstein, T., Melchers, K.G., König, C.J., Heslin, P.A., Lievens, F., 2012. Responding to personality tests in a selection context: The role of the ability to identify criteria and the ideal-employee factor. Human Performance 25, 273–302. https://doi.org/10.1080/08959285.2012.703733.

Kleinmann, M., Ingold, P.V., Lievens, F., Jansen, A., Melchers, K.G., König, C.J., 2011. A different look at why selection procedures work: The role of candidates'

ability to identify criteria. Organizational Psychology Review 1, 128–146. https://doi.org/10.1177/2041386610387000.
König, C.J., Melchers, K.G., Kleinmann, M., Richter, G.M., Klehe, U., 2007. Candidates' ability to identify criteria in nontransparent selection procedures: Evidence from an assessment center and a structured interview. International Journal of Selection and Assessment 15, 283–292.
Kulas, J.T., Stachowski, A.A., 2012. Social desirability in personality assessment: A variable item contamination perspective. The International Journal of Educational and Psychological Assessment 11, 23–42.
Kuncel, N.R., Tellegen, A., 2009. A conceptual and empirical reexamination of the measurement of the social desirability of items: Implications for detecting desirable response style and scale development. Personnel Psychology 62, 201–228. https://doi.org/10.1111/j.1744-6570.2009.01136.x.
Lechner, C.M., Rammstedt, B., 2015. Cognitive ability, acquiescence, and the structure of personality in a sample of older adults. Psychological Assessment 27, 1301–1311. https://doi.org/10.1037/pas0000151.
Lee, K., Ashton, M.C., 2004. Psychometric properties of the HEXACO Personality Inventory. Multivariate Behavioral Research 39, 329–358. https://doi.org/10.1207/s15327906mbr3902_8.
Lyubomirsky, S., King, L., Diener, E., 2005. The benefits of frequent positive affect: Does happiness lead to success?. Psychological Bulletin 131, 803–855. https://doi.org/10.1037/0033-2909.131.6.803.
Mansolf, M., Reise, S.P., 2016. Exploratory bifactor analysis: The Schmid-Leiman orthogonalization and Jennrich-Bentler analytic rotations. Multivariate Behavioral Research 51, 698–717. https://doi.org/10.1080/00273171.2016.1215898.
Markon, K.E., 2009. Hierarchies in the structure of personality traits. Social and Personality Psychology Compass 315, 812–826. https://doi.org/10.1111/j.1751-9004.2009.00213.x.
Marsh, H.W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A., Trautwein, U., Nagengast, B., 2010. A new look at the Big Five factor structure through exploratory structural equation modeling. Psychological Assessment 22, 471–491. https://doi.org/10.1037/a0019227.
McAbee, S.T., Oswald, F.L., 2013. The criterion-related validity of personality measures for predicting GPA: A meta-analytic validity competition. Psychological Assessment 25, 532–544. https://doi.org/10.1037/a0031748.
McAbee, S.T., Oswald, F.L., Connelly, B.S., 2014. Bifactor models of personality and college student performance: A broad versus narrow view. European Journal of Personality 28, 604–619. https://doi.org/10.1002/per.1975.
Meng, X., Rosenthal, R., Rubin, D.B., 1992. Comparing correlated correlation coefficients. Psychological Bulletin 111, 172–175. https://doi.org/10.1037/0033-2909.111.1.172.
Morin, A.J.S., Arens, A.K., Marsh, H.W., 2016. A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. Structural Equation Modeling: A Multidisciplinary Journal 23, 116–139. https://doi.org/10.1080/10705511.2014.961800.
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. https://doi.org/10.1016/j.jrp.2007.02.003.
Müthén, L.K., Müthén, B.O., 1998-2017. Mplus user's guide, (8th ed.). Muthén & Muthén, Los Angeles.
Pettersson, E., Turkheimer, E., Horn, E.E., Menatti, A.R., 2012. The general factor of personality and evaluation. European Journal of Personality 26, 292–302. https://doi.org/10.1002/per.839.
Podsakoff, P.M., MacKenzie, S.B., Lee, J., Podsakoff, N.P., 2003. Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of Applied Psychology 88, 879–903. https://doi.org/10.1037/0021-9010.88.5.879.
Poropat, A.E., 2009. A meta-analysis of the five-factor model of personality and academic performance. Psychological Bulletin 135, 322–338. https://doi.org/10.1037/a0014996.
Reise, S.P., 2012. The rediscovery of bifactor measurement models. Multivariate Behavioral Research 47, 667–696. https://doi.org/10.1080/00273171.2012.715555.
Richardson, H.A., Simmering, M.J., Sturman, M.C., 2009. A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. Organizational Research Methods 12, 762–800. https://doi.org/10.1177/1094428109332834.
Richardson, M., Abraham, C., Bond, R., 2012. Psychological correlates of university students' academic performance: A systematic review and meta-analysis. Psychological Bulletin 138, 353–387. https://doi.org/10.1037/a0026838.
Rodriguez, A., Reise, S.P., Haviland, M.G., 2016. Evaluating bifactor models: Calculating and interpreting statistical indices. Psychological Methods 21, 137–150. https://doi.org/10.1037/met0000045.
Rosenberg, M., 1965. Society and the adolescent self-image. Princeton University Press, Princeton, NJ.
Roth, P.L., BeVier, C.A., Switzer, F.S., Schippmann, J.S., 1996. Meta-analyzing the relationship between grades and job performance. Journal of Applied Psychology 81, 548–556. https://doi.org/10.1037/0021-9010.81.5.548.
Saucier, G., 2002. Orthogonal markers for orthogonal factors: The case of the Big Five. Journal of Research in Personality 36, 1–31. https://doi.org/10.1006/jrpe.2001.2335.
Saucier, G., Goldberg, L.R., 2003. The structure of personality attributes. In: Barrick, M.R., Ryan, A.M. (Eds.), Personality and work. Jossey-Bass, San Francisco, pp. 1–29.
Schmid, J., Leiman, J.M., 1957. The development of hierarchical factor solutions. Psychometrika 22, 53–61.
Soto, C.J., John, O.P., 2017. The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. Journal of Personality and Social Psychology 113, 117–143. https://doi.org/10.1037/pspp0000096.
Watson, D., Clark, L.A., Tellegen, A., 1988. Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology 54, 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063.
Williams, L.J., McGonagle, A.K., 2016. Four research designs and a comprehensive analysis strategy for investigating common method variance with self-report measures using latent variables. Journal of Business and Psychology 31, 339–359. https://doi.org/10.1007/s10869-015-9422-9.
Williams, L.J., O'Boyle, E.H., 2015. Ideal, nonideal, and no-marker variables: Confirmatory factor analysis (CFA) marker technique works when it matters. Journal of Applied Psychology 100, 1579–1602. https://doi.org/10.1037/a0038855.
Yik, M., Russell, J.A., Steiger, J.H., 2011. A 12-point circumplex structure of core affect. Emotion 11, 705–731. https://doi.org/10.1037/a0023980.
Yung, Y., Thissen, D., McLeod, L.D., 1999. On the relationship between the higher-order factor model and the hierarchical factor model. Psychometrika 64, 113–128. https://doi.org/10.1007/BF02294531.
Zhou, B., & McClendon, M. J. (1999). Cognitive ability and acquiescence. In American Statistical Association (Eds.), Proceedings of the survey research methods section (pp. 1003–1012), Alexandria.