



Published in final edited form as:

*J Pers Assess.* 2007 April ; 88(2): . doi:10.1080/00223890701268082.

## Bandwidth and Fidelity on the NEO-Five Factor Inventory: Replicability and Reliability of Saucier's (1998) Item Cluster Subcomponents

**Benjamin P. Chapman**

Department of Psychiatry, Laboratory of Personality and Development, University of Rochester  
Medical Center

### Abstract

Many users of the NEO-Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992) are unaware that Saucier (1998) developed item cluster subcomponents for each broad domain of the instrument similar to the facets of the Revised NEO Personality Inventory (Costa & McCrae, 1992). In this study, I examined the following: the replicability of the subcomponents in young adult university and middle-aged community samples; whether item keying accounted for additional covariance among items; subcomponent correlations with a measure of socially desirable responding; subcomponent reliabilities; and subcomponent discriminant validity with respect to age-relevant criterion items expected to reflect varying associations with broad and narrow traits. Confirmatory factor analyses revealed that all subcomponents were recoverable across samples and that the addition of method factors representing positive and negative item keying improved model fit. The subcomponents correlated no more with a measure of socially desirable responding than their parent domains and showed good average reliability. Correlations with criterion items suggested that subcomponents may prove useful in specifying which elements of NEO-FFI domains are more or less related to variables of interest. I discuss their use for enhancing the precision of findings obtained with NEO-FFI domain scores.

---

Most researchers have agreed that personality traits can be measured in hierarchical tiers that vary in generality (Cattell & Krug, 1986; Costa & McCrae, 1997; Eysenck, 1998; Goldberg, 1990). Traits at higher levels of multistratum taxonomies are merely composites of more fine-grained specific traits. For example, within the Five Factor Model of personality (Costa & McCrae, 1997; Digman, 1990; Goldberg, 1990; John & Srivastava, 1999), the broad trait of Conscientiousness is composed of narrow attributes such as self-discipline, achievement orientation, and orderliness. The commonly used Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) permits users to capture information about such higher order broad traits as well as the lower order, narrow traits that compose them (cf. Costa & McCrae, 1995).

The 60-item NEO-Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992) is a brief version of the NEO-PI-R designed to provide speedy and convenient measurement of the Five Factor Model domains; however, it does so by relinquishing information about the narrow traits that comprise each broad factor. To remedy this, Saucier (1998) derived item cluster subcomponents for the NEO-FFI that provided a more specific level of trait measurement. These subcomponents are somewhat similar, although not as specific, as the

facets of the full length NEO-PI-R and are listed in Table 1. In their development, these subcomponents evidenced strong average internal consistency comparable to NEO-PI-R facets (e.g., an average Cronbach's  $\alpha$  of .70 in a development sample and .66 in a cross-validation sample compared to average NEO-PI-R facet  $\alpha$  of .70) and captured the majority (although not all) of the content of the NEO-PI-R facets (cf. Saucier, 1998). Table 1 also reveals that the Neuroticism domain may be decomposed into either two or three item-cluster subcomponents; the former scoring scheme was empirically derived (as were all other sets of subcomponents), whereas the latter was rationally derived by Saucier to provide separate measures of anxiety and depression.

Although they provide a fruitful complement to the use of broad domain scores, these subcomponents have seen surprisingly little use despite Saucier's (1998) call for further investigation and refinement. PsycINFO and Social Science Citation Index revealed only three studies that have used one or more of the subcomponents in analyses (Rhodes & Courneya, 2003; Rhodes, Courneya, & Jones, 2002; Zelenski, Rusting, & Larsen, 2003) and only one additional study (Murray, Rawlings, Allen, & Trinder, 2003) was a direct effort to investigate their basic psychometric properties (i.e., means, standard deviations, test-retest reliability) in an Australian sample. One reason for this lack of use may be that the subcomponents are relatively unknown despite widespread use of the NEO-FFI. Another may be that in the absence of further development and cross-validation work, researchers are hesitant to utilize them. The goal of this study is therefore to provide a detailed investigation of the subcomponents' replicability and psychometric characteristics in two independent samples.

An appreciation of the subcomponent's importance may be gleaned from a brief review of the debate over the relative merits of scales capturing broad traits versus those capturing narrow traits (Ashton, 1998; Ashton, Jackson, Paunonen, Helmes, & Rothstein, 1995; Ones & Viswesvaran, 1996; Ones, Viswesvaran, & Schmidt, 1993; Paunonen, 1998, 2003; Paunonen & Ashton, 2001; Saucier & Ostendorf, 1999). In hierarchical personality inventories, the former correspond to higher order, general factors, whereas the latter correspond to lower order, specific factors. Ones and Viswesvaran (1996) and others (e.g., Ashton, 1998) couched the debate in the context of the bandwidth-fidelity dilemma originally laid out by Cronbach and Gleser (1965; see also Cronbach, 1970). The bandwidth of a scale refers to the breadth of its content or, in the case of a personality test, the scope of affective, behavioral, and cognitive tendencies it measures; the fidelity of the scale refers to its dependability or reliability (Cronbach & Gleser, 1965).

Rather than concerning themselves with the specific case of hierarchical personality inventories, Cronbach and Gleser (1965) discussed a general relationship between bandwidth and fidelity in which increasing bandwidth tends to reduce fidelity. For instance, if one uses 20 items to measure five specific constructs (e.g., 4 items for each construct), the reliability (fidelity) of each of those 4-item scales will be less than if all 20 items were expended to measure the construct of interest. Similarly, all other conditions being equal, the reliability of the broadband, factorially complex, 20-item composite should also be less than that of a narrowband, unifactorial, 20-item construct.

A slightly different relationship between bandwidth and fidelity may obtain in the case of personality tests measuring higher and lower order factors corresponding to broad and narrow traits (Ones & Viswesvaran, 1996). Broad traits corresponding to higher order factors are measured by wideband composites; these composites are themselves composed of narrowband scales for specific, lower order traits that load on the higher one. In this case, the broadband composite actually has higher reliability, or fidelity, compared to narrowband scales for specific traits because (a) the broadband scale contains far more items than

any of the narrower bandwidth scales and (b) all of the narrower bandwidth scales are highly intercorrelated. Thus, in hierarchical personality inventories, increasing bandwidth from a specific to a broad trait tends to increase fidelity as well, a situation slightly different than the inverse relationship between the two originally articulated by Cronbach and Gleser (1965).<sup>1</sup>

Despite the reliability advantage of broadband composites, the narrow traits in hierarchical personality inventories are important for at least three reasons. First, the higher reliability of broad trait scales means only that a general, multidimensional construct has been measured with greater precision. What do the resulting scores mean? They summarize people's standing on a multifaceted trait but may well obscure important individual differences across that trait's specific facets. Consider Neuroticism, which, on the NEO-PI-R, is measured by six facets: depression, anxiety, angry hostility, self-consciousness, impulsiveness, and vulnerability. Different individuals may all obtain Neuroticism scores at the 50th percentile through markedly different patterns of facet scores (Saucier & Ostendorf, 1999). One person may score high on angry hostility and impulsiveness and low on self-consciousness and depression. A second may score high on depression and self-consciousness but low on impulsiveness and angry hostility. A third may score in the 50th percentile on all facets of Neuroticism. Broadband Neuroticism scores mask such important phenotypic variation,<sup>2</sup> and may lead to accordingly imprecise results and interpretations.

A second reason lower order, narrow traits are important was highlighted by Ashton (1998). Ashton argued that the greater reliability of broadband trait scales does not always mean they have better predictive power. Although it is commonly held that reliability places an upper limit on validity (Anastasi & Urbina, 1997), this is a theoretical notion, and high reliability may be outweighed by another important consideration as Ashton (1998) noted:

Despite this increase in reliability, it does not follow that the broad composite scale must be a more effective predictor of a given criterion than are all of its constituent subscales. The question to be asked is whether or not the improved reliability derived by aggregating the subscales provides a gain in validity that outweighs the loss in validity due to the dilution of variance specific to certain subscales which relate to the criterion of interest. (p. 289)

In other words, imprecision in predictive analyses can arise because frequently only some but not all aspects of a broad trait are correlated with the criterion of interest. Figure 1 illustrates a set of personality items aggregated alternatively into a few broadband, multidimensional trait composites or several narrower bandwidth composites within each domain. In this diagram, only one specific dimension of the broad trait composite shares variance with the criterion measure. This is an oversimplified case because often other dimensions of the broadband composite will correlate to some degree with a criterion. However, in the simplified case of Figure 1, the broadband composite will correlate less with the criterion than the relevant specific composite because the broadband trait has a lower ratio of shared to unshared variance with the criterion. In other words, items or factors in the broadband composite that are unrelated to the outcome of interest (i.e., irrelevant information) will attenuate the correlation. Empirical results support this (Ashton, 1998; Ashton et al., 1995; Paunonen, 1998, 2003; Paunonen & Ashton, 2001; cf. also Carver, 1989).

<sup>1</sup>I would like to thank an anonymous reviewer for noting this important distinction.

<sup>2</sup>Jang, McCrae, Angleitner, Riemann, and Livesley (1998) also suggest that genetic bases underlie phenotypic variation at the level of specific, lower order traits as well as higher order trait factors.

A third reason narrow traits remain important is that the exact substantive meaning of a correlation between a broad trait and a criterion may be difficult to discern because the broad trait involves several components. Interpretations of such results are perforce overly general because it is impossible to know which aspects of the broad trait are more or less related to the outcome of interest. The understanding of an initial broadband trait effect can be enhanced by examining which of its specific dimensions are more or less related to the outcome. For instance, Weiss and Costa (2005) recently supplemented their overall finding that Conscientiousness was inversely related to mortality with facet-level analyses that indicated that it was primarily the self-discipline facet that confers longevity.

These arguments for the use of narrow trait composites suggest that Saucier's (1998) NEO-FFI subcomponents represent a potentially important tool for users of the NEO-FFI. However, their replicability and psychometric characteristics in cross-validation samples remain largely unknown. With the exception of Murray et al.'s (2003) examination of norms and test-retest reliability in an Australian sample, little work has built on Saucier's (1998) initial efforts. In this study, I pursued four primary aims in this respect. First, I assessed the degree to which the item clusters could be recovered in two independent samples of varying age and demographics via multiple-group confirmatory factor analysis (CFA). Saucier also expressed concern about the imbalance between positively and negatively keyed items within item clusters. Therefore, my second objective was to assess whether item keying explained additional covariance among items within each domain beyond that attributable to the subcomponents. This was supplemented by examining the subcomponents' differential correlations with a measure of socially desirable responding. Third, Saucier noted small reliability shrinkages in his cross-validation sample using coefficient alpha. Reliability assessment therefore constituted another aim of the study. Consistent with recent suggestions that alpha alone presents a limited picture of internal consistency (Zinbarg, Revelle, Yovelle, & Li, 2005), I used three different internal consistency estimates varying in their psychometric assumptions to describe the subcomponents: Cronbach's (1951) alpha ( $\alpha$ ); Raykov's (1997a) rho ( $\rho$ ); and Bentler's (1972, in press) dimension free, lower bound, reliability estimate, theta ( $\theta$ ). Finally, a fourth aim was to determine whether specific subcomponents either (a) correlated higher than broad domains with a small set of criterion items or (b) showed differential criterion correlations, thereby illuminating which elements of a broad domain drive criterion correlations.

I selected items to represent criteria for which narrow traits might vary in importance as well as criteria highly salient to each sample's life stage. Items for Sample 1, young adult university students, included the extent to which, over the last semester, participants had socialized with the other sex, felt depressed, anxious, or contemplated suicide. I selected these items to reflect Erikson's (1963) psychosocial conflict of intimacy versus isolation and well-known symptoms of college maladjustment (Zitzow, 1984a). I hypothesized the following subcomponent-criterion correlations to exceed those of other subcomponents within the domain and of the broad domain score: sociability and time spent socializing with the opposite sex, depression and amount of depressed mood reported over the semester, self-reproach and thoughts of suicide over the semester, and anxiety and the extent of anxious feelings reported over the prior semester.

Items for Sample 2, middle-age community adults, reflected satisfaction in domains salient to midlife from the perspective of Levinson's (1986; see also Harris, Elliot, & Holmes, 1986) midlife transitions: life satisfaction, marital satisfaction, job satisfaction, and engagement with creative or artistic pursuits. I hypothesized the following subcomponent-criterion correlations to exceed those of the criterion with other subcomponents within the domain or the broad domain score: positive affect and life satisfaction, goal striving and job satisfaction, prosocial orientation and marital satisfaction, and aesthetic interests and artistic

hobbies. Finally, the revised University of California, Los Angeles (UCLA) Loneliness Scale (Russell, Peplau, & Cutrona, 1980) served as a criterion measure for both samples. Social relationships remain important throughout the life course but serve different ends at different points in life. Because younger adults tend to cultivate a broad array of relationships in the service of knowledge acquisition goals (Carstensen, 1995), I hypothesized that intellectual interests and sociability would be stronger (negative) predictors of loneliness than other subcomponents and domain scores in Sample 1. By midlife, Carstensen's theory suggests that social relationships begin to function more as a means of emotion regulation; and thus, in Sample 2, depression and lower positive affect were hypothesized to predict loneliness better than other subcomponents or broad domains. I used both standard correlations and correlations corrected for attenuation due to measurement error (CAME; cf. Charles, 2005) in these analyses to gauge the effect on criterion correlations of probable decreased subcomponent reliability due to shorter length.

## METHOD

### Participants and Procedure

Sample 1 consisted of 308 young adults drawn from large introductory psychology classes (74% women; age  $M = 20$ ,  $SD = 2.2$ ; 66% White). Sample 2 consisted of 256 middle-age adults drawn from the community (62% women; age  $M = 49$ ,  $SD = 5.4$ ; 76% White). Approximately 70% of the sample was recruited through contacts in undergraduate classes using a snowball technique in which students provided surveys to their parents who then dispensed surveys to middle-age friends. The remainder of the sample was recruited through churches and a hospital, again using a snowball technique. The middle-age sample contained proportionally more women and Anglo-Americans and was also more educated than the college sample (15.23 years vs. 13.58 years). Participants confidentially completed a battery of measures for a large-scale project on personality, intelligence, emotional intelligence, and social adjustment in young adulthood and middle age (Chapman, 2005). A total of 305 young adults and 209 midlife adults provided usable NEO-FFI data. I reverse scored all negatively keyed items prior to analyses. In an earlier article on emotional intelligence, Chapman and Hayslip (2005) reported means and standard deviations of NEO-FFI domains and the Revised UCLA Loneliness Scale for 291 of the young adult sample.

### Other Measures

To assess social desirability, participants completed the 10-item version of the Marlowe-Crowne Social Desirability Scale (Strahan & Gerbasi, 1972). Criterion variables in each sample consisted of the UCLA Loneliness Scale, Revised (Russell et al., 1980), and single item measures from brief surveys. Three items asked middle-aged participants to rate their overall life satisfaction, marital satisfaction, and job satisfaction on a 5-point Likert scale ranging from 1 (*not at all satisfied*) to 5 (*very satisfied*). Another item asked them to report the degree to which they were active in artistic pursuits such as music, painting, or writing as either a producer or consumer on a scale ranging from 1 (*no involvement at all*) to 5 (*frequent involvement*). Four items were adapted for the young adults from the College Adjustment Rating Scales (CARS; Zitzow, 1984a, 1984b) to assess the frequency with which young adults had engaged in certain behaviors or experienced certain problems. One behavioral item asked how often in the past semester participants had socialized with the other sex; and three mood items asked how frequently over the past semester they had felt anxious, felt depressed, or contemplated suicide. Response format ranged from 1 (*not at all*) to 5 (*quite a bit*). The use of single-item criterion measures is common in survey research methodology and is consistent with prior bandwidth-fidelity research (cf. Ashton, 1998; Paunonen, 1998, 2003).



## Overview of Analyses

I included only participants with complete data in the analyses, which proceeded in several steps. CFAs modeled the item clusters as oblique factors within each NEO-FFI domain separately and fixed their variances to unity for scaling purposes and freely estimated all item loadings, residuals, and factor covariances (cf. Byrne, 2005). I estimated models from polychoric correlation matrices, which approximate the continuous distributions underlying response scales involving ordinal categories (Olsson, 1979) and were suggested for use with the ordinal response scales of personality instruments (Panter, Swygert, Dahlstrom, & Tanaka, 1997). I utilized maximum likelihood (ML) robust estimation (Satorra & Bentler, 1988) to produce a chi-square fit statistic scaled to correct for departures from multivariate normality (the Satorra-Bentler Chi-Square).

CFA models proceeded in the following order. First, I tested models including only latent trait factors for each domain separately in each sample. Then I tested a second set of models in which each item loaded on a corresponding trait factor and on a secondary method factor representing common variance among either the positively or negatively keyed items of each domain (cf. Saris & Aalberts, 2003). Figure 2 depicts the parameterization of this model for the Conscientiousness item clusters. I evaluated the extent to which these models fit the data better than models without method factors with the Akaike information criterion (AIC; Akaike, 1973), which can be used to compare nonnested factor models (lower values indicate better fit). Finally, in a third series of nested models, I assessed the subcomponents' measurement invariance across age groups. These models incrementally constrained trait factor loadings, method factor loadings, residual variances, and factor covariances (cf. Ployhart & Oswald, 2004; Vandenberg & Lance, 2000) to ascertain whether items functioned differently across age groups. I gauged fit degradations by use of the Satorra-Bentler (Satorra & Bentler, 2001) Chi-Square Difference Test only in cases in which baseline models had yielded nonsignificant chi-squares (Yuan & Bentler, 2004). In all other cases, I made nested comparisons on the basis of the comparative fit index (CFI) difference test suggested by Cheung and Rensvold (2002; e.g., CFI changes greater than .01).

Raykov (1998) recommended that the root mean square error of approximation (*RMSEA*), an absolute fit index, be given the heaviest weight in assessing models based on personality data because it is less affected by violations of multivariate normality typically seen in personality data. Similarly, after extensive simulation studies intended to reflect conditions typical of personality data (i.e., simple structure violations, non-normality, smaller sample sizes), Beauducell and Wittman (2005) also suggested the *RMSEA* as a trustworthy index of model fit. The *RMSEA* was supplemented with the CFI, a commonly used and well-supported incremental fit index. *RMSEA* values below .05 indicate good fit, .05 to .08 acceptable fit, and .10 or above poor fit (Browne & Cudeck, 1993); whereas CFI values greater than .90 are generally considered indicative of adequate fit (Tanaka, 1993).<sup>3</sup>

The reliability analyses also proceeded in multiple steps. Because Raykov (1997a, 1997b) demonstrated that Cronbach's  $\alpha$  underestimates reliability in composites that are not essentially tau equivalent (i.e., in which items differ in amount of true score variance, regardless of intercept), I first tested each subcomponent for essential tau equivalence. These tests consisted of specifying a simple one-factor model in which item loadings were free to vary and then a similar model in which I constrained all item loadings to equality (cf. Nunnally & Bernstein, 1994).<sup>4</sup> Significant decreases in fit of the more restrictive model, which I identified in the same manner as previously, identified scales that did and did not

<sup>3</sup>Although Hu and Bentler (1999) provided more stringent guidelines for evaluating model fit (i.e., CFI > .95, RMSEA < .05), Marsh and colleagues' (2004) recent simulation results suggest that these cut points for model rejection may be less reliable than previously thought.

meet the assumption of essential tau equivalence. Next, I compared three reliability coefficients for each subcomponent in each sample: Cronbach's  $\alpha$ ; Raykov's (1997a) reliability coefficient  $\rho$ , which is derived from a unidimensional latent-variable model but does not assume essential tau equivalence among items; and Bentler's (1972, in press) dimension-free, lower bound estimate of reliability  $\rho_{\text{LB}}$ , which does not even assume that a set of items is congeneric (i.e., multiple factors may characterize the item set).

Following reliability analyses, I calculated means, standard deviations, and subcomponent intercorrelations for each sample because adult and college norms for the NEO-FFI domains are different (Costa & McCrae, 1992; see also Terracciano, McCrae, Brand, & Costa, 2005). Finally, I tested criterion correlations of NEO-FFI broad domain scores in each sample. To reduce Type I error rate, I only considered correlations significant at .01. For each significant domain correlation, I then examined the pattern of subcomponent correlations. I then corrected all correlations for measurement error. I conducted analyses in SPSS 10.0 and EQS 6.1.

## RESULTS

### Subcomponent Replication

Table 2 depicts the fit indexes for factor models of each NEO-FFI domain including the final multigroup model in which I constrained all possible item loadings, residuals, and factor covariances to equality across groups. In general, the most restrictive models specifying only latent traits evidenced adequate fit, according to the *RMSEA* (e.g., .048 for the Conscientiousness subcomponents to .086 for the Extraversion subcomponents, both in the middle-aged sample). However, these values suggested that additional interitem covariance remained unexplained by models including only latent trait factors. When I added positive and negative item-keying method factors, *RMSEAs* decreased considerably. Substantial decreases in the AIC values also indicate that models including method factors provided a better fit. In general, the maximally constrained multigroup models with method factors provided the most general representation of the subcomponents as evidenced by their largely metric invariance, good absolute fit, and low AICs relative to less constrained multigroup models.

Of the 60 NEO-FFI items, 53 loaded primarily on their trait factor, whereas 7 loaded higher on their respective method factor. The series of nested multiple-group models revealed that 45 of the 60 items were invariant across samples with respect to trait and method factor loadings and residual variances when using the three-subcomponent scheme for Neuroticism, and 43 of 60 were invariant using the two-subcomponent Neuroticism scheme. Of these Neuroticism items, Item 46 loaded slightly (.46 vs. .42) and Item 41 moderately (.71 vs. .52) higher on depression in the young adults for the tripartite solution. In the bipartite alternative, Item 51 loaded marginally (.64 vs. .63) and Item 36 slightly (.64 vs. .58) higher on negative affect in the young adults, whereas item 21 loaded slightly (.57 vs. .56) and 56 moderately (.74 vs. .95) higher on negative affect in the middle-aged adults. Other loading discrepancies across samples were within .20 for all items except Items 34 and 54, which had loading differences across samples of .31 and .35, respectively. In both groups, loadings for the unconventionality subcomponent were quite low (i.e., all below .40), which

---

<sup>4</sup>Note that in baseline models of clusters of three items, the unconstrained model itself is just identified and always fits the data perfectly. In such cases, the constraints imposed by essential tau-equivalence overidentify the model and permit it to be tested. In these cases, the essential tau-equivalence of the item cluster was evaluated merely by the significance of the Satorra-Bentler chi square for this testable model. However, this approach may be less optimal than comparing the fit of an essentially tau-equivalent model nested within one in which loadings are permitted to vary.

suggested that it was not a well-defined factor despite the overall model fit for the Openness subcomponents.

Correlations among the trait factors tended to be above .60, with the Neuroticism subcomponents in the .90s; this was expected given that different elements of a multifaceted construct must by definition be strongly correlated; and when measurement error is removed in latent variable models, these correlations increase, sometimes substantially. When treated as unweighted linear composites of observed items scores rather than latent variables, subcomponent intercorrelations were lower (see following and Table 4).

Bivariate correlations between domain or subcomponent scores and the 10-item Marlowe-Crowne are presented in Table 3, which also includes reliability estimates (discussed following). Significant negative correlations with social desirability were observed in the young adults ( $Ns = 296-302$ ,  $ps < .005$ ) for negative affect and self-reproach in the two-subcomponent Neuroticism model and for anxiety and self-reproach in the three-subcomponent Neuroticism model. Within Agreeableness, the nonantagonistic orientation subcomponent correlated positively with social desirability. Among the middle-aged adults ( $Ns = 204-207$ ,  $ps < .005$ ), all subcomponents from either Neuroticism scheme correlated negatively with social desirability, whereas all subcomponents of Agreeableness and Conscientiousness correlated positively with social desirability. Table 3 shows that the parent domains of these subcomponents also correlated with social desirability in the same direction and to a comparable degree.

### Reliability Analyses

Table 3 also presents the results of tests for essential tau equivalence along with  $\omega$ ,  $\omega^2$ , and estimates of subcomponent internal consistency in each sample. For essentially tau-equivalent subclusters,  $\omega$  was nearly equal to  $\alpha$  as expected. However, for subcomponents in which the degree of true score variance differed across items,  $\omega$  was slightly higher than (i.e., .01-.02 rounding at the second decimal). These results are consistent with Raykov's (1997a) demonstration that  $\alpha$  may underestimate reliability when items contain differing amounts of true score variance. Bentler's (1972, in press)  $\omega$  provided far higher estimates of subscale reliability for some subcomponents. This coefficient makes no assumptions about the dimensionality of a composite and always exceeds alpha (Bentler, 1972, in press).

In general, it can be seen that the appropriate congeneric reliability estimates, bolded in Table 3 (i.e., either  $\omega$  or  $\omega^2$ , depending on tau equivalency) ranged from fair to good, and were slightly higher in the middle-aged sample. An obvious exception was the unconventionality subcomponent. Average reliability across the subcomponents in the middle-aged sample was .72 (.73 without unconventionality), slightly higher than that reported by Saucier (1998; .70 for his midlife sample) and the average NEO-FFI facet reliability (.70), both of which are based strictly on  $\alpha$ . For the young adults, the average subcomponent reliability was .66 (.68 without unconventionality), equal to what Saucier obtained in his college sample.

Finally, all three reliability estimates are provided in Table 3 for the domains themselves. Because the domains are multidimensional rather than congeneric by definition, I did not test essential tau equivalence. Whereas  $\alpha$  and  $\omega$  were intended for use with congeneric item sets,  $\omega$ , which entails no such assumption, may be regarded as a more appropriate reliability estimate for the domains. However, domain  $\alpha$ 's affirmed the well-known psychometric fact that  $\alpha$  may be high in multidimensional scales (Streiner, 2003a, 2003b) and thus should be not be regarded as a measure of "homogeneity" (Cortina, 1993).



### Subcomponent Intercorrelations, Means, and Standard Deviations

Table 4 presents the subscale intercorrelations, with the young adults below the diagonal, the middle-aged sample above, and correlations among subcomponents within each domain bolded in corresponding off-diagonal blocks. The pattern for both samples suggested moderate to strong correlation among subcomponents within a domain and much smaller correlations between item clusters from different domains. This is the precise pattern expected among the lower order dimensions of a set of personality domains.<sup>5</sup> Of interest, decomposing the Neuroticism domain into three rather than two item clusters resulted in slightly more independent subcomponents, particularly in the young adults. Means and standard deviations of the item clusters are presented in Table 5, which also shows a measure of effect size (Cohen's *d*) of the cross-sectional differences between these samples.

### Criterion Correlations

Domains correlated with criterion items are presented in Table 6 along with corresponding subcomponent criterion correlations. Correlations corrected for attenuation due to measurement error (using the appropriate reliability estimate for each subcomponent from Table 3) are bolded. In some cases, a set of subcomponents tended to be equally associated with a criterion, with their parent domain correlating slightly higher (e.g., Extraversion vs. its subcomponents and job satisfaction in the middle-age sample). In other instances, it was evident that broad domain correlations were driven primarily by one or two subcomponents (e.g., prosocial orientation and loneliness in the young adults). Some specific hypotheses were supported in the uncorrected correlations; and when measurement error was corrected, many were supported.

## DISCUSSION

The goal of this study was to determine if Saucier's (1998) NEO-FFI item cluster subcomponents were replicable in independent samples, to assess whether item keying accounted for additional item covariance within domains, the extent to which subcomponents were susceptible to socially desirable responding, the most appropriate reliability estimate for each subcomponent, and to provide a preliminary comparison of subcomponents versus domain correlations with a small set of criterion items. Results permit several conclusions about the NEO-FFI subcomponents.

First, Saucier's (1998) item clusters appear replicable in independent samples. Evaluation of trait-only CFA models for each domain on the basis of the *RMSEA* (Beauducell & Wittman, 2005; Raykov, 1998) suggested adequate fit in both the young and middle-aged adults. These trait-only models were fairly strict tests of cross-validation in that they posited no additional sources of covariation among items other than factors corresponding to narrow personality traits within each domain. Relative improvements in fit were noted with the addition of method factors reflecting item keying. Although additional factors usually improve model fit, many items loaded saliently on these method factors including a few that loaded higher on their corresponding method factor than on their trait factor. The fact that positive and negative item keying appear to contribute variance to items within each NEO-FFI domain was consistent with Raykov's (1998) characterization of personality data as capturing a complex set of processes that are not easily summarized by simple models. When individuals respond to questions about their typical behavior, attitudes, and feelings, they may be influenced by whether the item is worded in a way reflecting favorable or unfavorable characteristics.

---

<sup>5</sup>For instance, Watson, Clark, and Harkness (1994) characterized multistratum trait taxonomies as statistical variance-covariance hierarchies in which covariance among lower order dimensions constitute the variance of higher order factors.

The subcomponents were also structurally invariant across age groups, consistent with evidence that the basic configuration of higher and lower order factors characterize individuals of different ages (see, e.g., McCrae & Costa, 2003). Moreover, about three fourths of the items showed metric invariance with respect to trait and method factor loadings as well as equal residual variance across age groups. The remaining items evidenced some differences in one or more of these parameters. Exceptions included Item 34, which showed a loading difference of .31 in favor of the young adults. This item asks whether the respondent is well liked by the majority of people known to him or her and appeared to describe prosocial orientation much better in young adults. Item 54, which loaded .35 higher in middle-aged adults, asks whether respondents are forthright in expressing their antipathy toward individuals they dislike. Scores appeared to characterize nonantagonistic orientation much better in middle-aged than in young adults.

Finally, certain Neuroticism items may show variation according to whether the two or three subcomponent scheme is used. Under the tripartite arrangement, Item 41, which deals with feeling discouraged and giving up when hardship is encountered, characterizes trait depression better in young than midlife adults; decomposing Neuroticism into only two subcomponents, Item 56, which asks about shame, characterizes self-reproach better in middle-aged respondents. On the whole however, these analyses revealed that not only was the basic factor configuration of Saucier's (1998) lower tier NEO-FFI traits recoverable in somewhat different samples but that the vast majority of items functioned quite similarly across age groups.

Social desirability correlations were small but significant and were moderated by age group. In particular, middle-aged individuals may produce scores on all Neuroticism, Agreeableness, and Conscientiousness subcomponents more influenced by social desirability. Out of those domains' subcomponents, young adults by contrast may respond more honestly to questions tapping orderliness, goal striving, dependability, prosocial orientation, and depression. However, social desirability has itself been considered a personality trait reflecting positive characteristics or a sense of well-being rather than a dissembling response style per se (McCrae, 1986; McCrae & Costa, 1983). Some work has also indicated scores on the short Marlowe-Crowne are correlated with age on the order of .30, at least in women (Ray, 1988; Ray & Lovejoy, 2003). In this context, these correlations may also mean that whatever positive trait might be tapped by the Marlowe-Crowne is associated with all elements of Neuroticism, Agreeableness, and Conscientiousness in middle-aged adults but with only with a limited set of these domains' subcomponents in young adults. As one reviewer pointed out, whether social desirability is viewed as a substantively meaningful trait in and of itself or not, these results bear no necessary connection to the NEO-FFI validity question "Have you responded accurately and honestly" (Costa & McCrae, 1992). It would not be socially desirable to answer negatively, and individuals high on trait well-being probably have little to conceal. Naturally, dissent on this validity item should still constitute sufficient grounds for jettisoning a protocol, but NEO-FFI validity scales similar to those that have been developed for the NEO-PI-R (Schinka, Kinder, & Kremer, 1997) might still be considered in future research.

With respect to reliability, less than half the subcomponents in each sample met assumptions of essential tau equivalence required for Cronbach's  $\alpha$  to provide a strictly accurate estimate of reliability. Violations of essential tau equivalence are fairly common in practice (Raykov, 1997a, 1997b). Raykov's  $\omega$  (1997a), an internal consistency estimate unaffected by violation of this assumption, yielded reliability estimates slightly higher. Although the small magnitude of difference was consistent with prior work on these coefficients (Raykov, 1997a, 1997b; see also Zinbarg et al., 2005), these analyses do suggest it may slightly underestimate reliability for some subcomponents.

With shorter composites that violate essential tau equivalence, “rule-of-thumb” guidelines about acceptable values of  $\alpha$  that pervade applied psychological literature may be misleading (Cortina, 1993). Streiner (2003b), among others, emphasized the well-known fact that  $\alpha$  is strongly affected the number of items in a scale and, like Cortina (1993), cautioned against blind use of  $\alpha$  to evaluate an instrument’s reliability without consideration for factors that affect  $\alpha$  (Streiner, 2003a). For instance, even though  $\alpha$  is intended for unidimensional scales, these results clearly show that  $\alpha$  was high for the multifactorial NEO-FFI domains. It is well known in the psychometric literature that  $\alpha$  is not a measure of dimensionality or homogeneity (Cortina, 1993; Streiner, 2003a), and for this reason, alternative estimates of reliability provide a better estimate of true reliability when scales are not con-generic or essentially tau equivalent (Zinbarg et al., 2005; see also Bentler, in press).

Despite these caveats,  $\alpha$  may still often be a reasonable lower bound estimate of true score variance within a scale (Cortina, 1993);<sup>6</sup> and regardless of the internal consistency estimate used, these analyses generally revealed that the reduced length of the subcomponents brought expected reductions in reliability. Thus, moving to narrow bandwidth subcomponents on the NEO-FFI reduces fidelity somewhat. From Cronbach and Gleser’s (1965) original perspective, however, if one could use the same number of items to measure a specific trait and a broad, multidimensional trait, the narrower bandwidth measure would show higher fidelity. However, because the NEO-FFI subcomponents are located beneath domains in a factorial hierarchy, they cannot be measured with same number of items as the broad domains, and, all other things being equal, reductions in reliability are inevitable. This should not deter researchers interested in using the subcomponents given the availability of methods for the control of measurement error. In addition to correcting single and multiple correlations for attenuation due to measurement error, narrow traits can be modeled as latent variables in structural equation analyses. The use of factor scores from exploratory common factor analysis (not principal component analysis) rather than simple linear combinations of observed item scores is another alternative for reducing measurement error, as are regression models which account for errors in predictors.

Another interesting pattern with respect to the internal consistency analyses was that the subcomponents were more reliable in the middle-aged rather than the young adult sample, with the exceptions of intellectual interests, prosocial orientation, and activity. This is not surprising because the middle-aged sample was demographically quite similar to that in which Saucier (1998) derived the subcomponents. On the other hand, the college sample differed not only in age and gender proportion from Saucier’s middle-aged community sample but also in age from his college sample, the members of which were on average 32 years old. Nonetheless, the average internal consistency reliabilities observed in both this middle-aged sample (.72–.73, with and without unconventionality) and young sample (.66–.68) were nearly identical to the (  $\alpha$ -based) .70 and .66 reported by Saucier. As Saucier pointed out, these numbers compare favorably with the average (  $\alpha$ -based) reliability of .70 for the NEO-PI-R facets and, in this context, provide further evidence of the subcomponents’ stability.

The unconventionality item cluster was an obvious exception to this pattern. However, it also evidenced the lowest reliability and greatest cross-validation  $\alpha$  shrinkage in Saucier’s (1998) work. From a conceptual point of view, it is easy to understand why such a subcomponent was essentially unrecoverable in these samples. The item content tapped a dimension of personality defined by the rejection of intellectual and cultural norms; high scorers would be free thinking and unshackled by social and attitudinal conventions. By

---

<sup>6</sup>Note that even though internal consistency estimates may be high in multidimensional scales, the interpretation of true score variance is less clear because of factorial complexity (Cortina, 1993).

definition, however, the psychometric consistency of a trait in the population depends on shared convention in the interpretation of items. If a sample contains many unconventional participants who diverge in thought and attitude, their responses to items may be highly idiosyncratic. Rather than stable covariation, items measuring such a trait might be essentially uncorrelated as a result of diverging response tendencies. The effect on a four-item composite would be to substantially diminish or even eliminate the stable common variance on which internal consistency estimates are based (cf. Streiner, 2003b). One would also expect diminished internal consistency of the broadband scale within which such items are embedded. Not surprisingly, the Openness domain has routinely shown the lowest internal consistency (e.g., Costa & McCrae, 1992; Weiss et al., 2005). This raises an interesting paradox about the measurement of unconventionality itself as a lower order trait: Respondents need be at least somewhat conventional in their interpretations of items measuring unconventionality, and the items need to elicit a standard pattern of covariation across samples.

Finally, the criterion correlations suggested that NEO-FFI subcomponents hold promise for delineating which elements of broad domains are more and less related to criteria. Hypothesized relations obtained, in general. Exceptions involved goal striving, which did not predict job satisfaction in the middle-aged adults, and self-reproach, which did not predict suicidal ideation in the young adults. In the corrected correlations, sociability was the second most predictive subcomponent of loneliness in young adults, but intellectual interests was unrelated to loneliness. Sociability was more predictive than positive affect of loneliness among the middle-aged adults, and in both samples, broad Extraversion correlated more highly with loneliness than any subcomponent. Another interesting result I did not hypothesize was that in the young adults, it was the prosocial component of Agreeableness rather than nonantagonistic orientation that seemed to best predict loneliness. Prosocial orientation reflects proactive courtesy and consideration, which are likely to garner one friends; on the other hand, nonantagonistic orientation merely reflects a general absence of belligerence, Machiavellianism, and mistrust. Simply lacking these qualities may not be sufficient to protect against loneliness.

In other cases, criterion correlations were not markedly different across subcomponents. These may be instances in which all subcomponents were equally relevant to a given criterion, as when the criterion itself is multidimensional (Hogan & Roberts, 1996). The shared variance among “sibling” subcomponents captured by broadband composite scores appeared to outweigh the specific variance of each subcomponent in these cases, which resulted in higher criterion correlations for the domains rather than their subcomponents. This is consistent with Ones and Viswesvaran’s (1996) argument favoring broadband composites on the basis of their higher reliability, which is in part due to variance shared across constituent (highly correlated) domains. However, when reliability differences were equalized by correcting for measurement error, the pattern of differential correlations among the subcomponents and domains was magnified in some cases. For instance, aesthetic interests predicted the amount of weekly time middle-aged people devoted to artistic hobbies slightly better than broad Openness. Sociability predicted the extent to which college students socialized with the other sex better than broad Extraversion. The nature and magnitude of these differences depends, of course, on the population and criterion of interest. Nevertheless, these analyses suggest the NEO-FFI subcomponents may provide a way to examine the specific elements of monolithic personality factors most important for outcomes in the same way facets may be used to supplement broad domain results with the full-length NEO-PI-R (e.g., Costa & McCrae, 1995; Weiss & McCrae, 2005).

These results must be qualified by five limitations. First, the generalizability of the subcomponents to other populations requires further investigation. Second, in this study, I

relied on self-report methodology. These analyses suggest that either item keying and/or social desirability may have affected subcomponent scores, albeit minimally. Future work might investigate the agreement between self and informant ratings on the traits represented by these subcomponents, as has been done with NEO-PI-R facets (Costa & McCrae, 1992). Third, Saucier's (1998) development sample ( $N = 732$ ) was considerably larger than this middle-aged sample, and potential users looking for subcomponent norms may wish to use his means and standard deviations, although this study provides additional data on means and standard deviations in a fairly large traditional college sample. Fourth, the use of a small number of mostly single-item criterion measures was intended as only a supplementary investigation of predictive contrasts between the domains and subcomponents. The correction for attenuation due to measurement error may tend to favor shorter scales, other things being equal, and I did not subject differences between subcomponent and domain criterion correlations to significance testing. The use of the subcomponents in routine research will probably yield more robust tests of their predictive specificity and discriminant validity. Fifth, although strong correlations between narrow traits within the same domain are virtually axiomatic, modeling them as latent variables may inflate these correlations even more by removing measurement error. Thus, analyses implementing the subcomponents may wish to guard against collinearity problems by including subcomponents within the same domain one at a time, as Weiss and Costa (2005) recently did, or perhaps in some cases using a single linear combination of two subcomponents (if such an increase in bandwidth is deemed conceptually tolerable and interpretable given the research context).

In the final analysis, these findings suggest that with the exception of unconventionality, the narrow trait subcomponents embedded in the NEO-FFI were stable across independent samples, comparable in reliability to NEO-PI-R facets, and offer the potential to measure personality at higher specificity than is possible with only broadband domain scores. Saucier's (1998) robust item cluster subcomponents mean that greater speed in five-factor personality assessment no longer need come at the expense of measurement specificity.

## Acknowledgments

Preparation of this work was supported by National Institute of Mental Health Public Health Service Grant T32MH073452 to Jeffrey Lyness, University of Rochester Medical Center, Department of Psychiatry. I thank Lewis Goldberg and an anonymous reviewer for helpful comments on an earlier draft of the article. Additional thanks to Paul R. Duberstein who provided valuable input during the writing process, to Bert Hayslip Jr. who supervised the research on which the article is based, and to Gerard Saucier who offered impressions on the subcomponents that helped orient the article early on.

## References

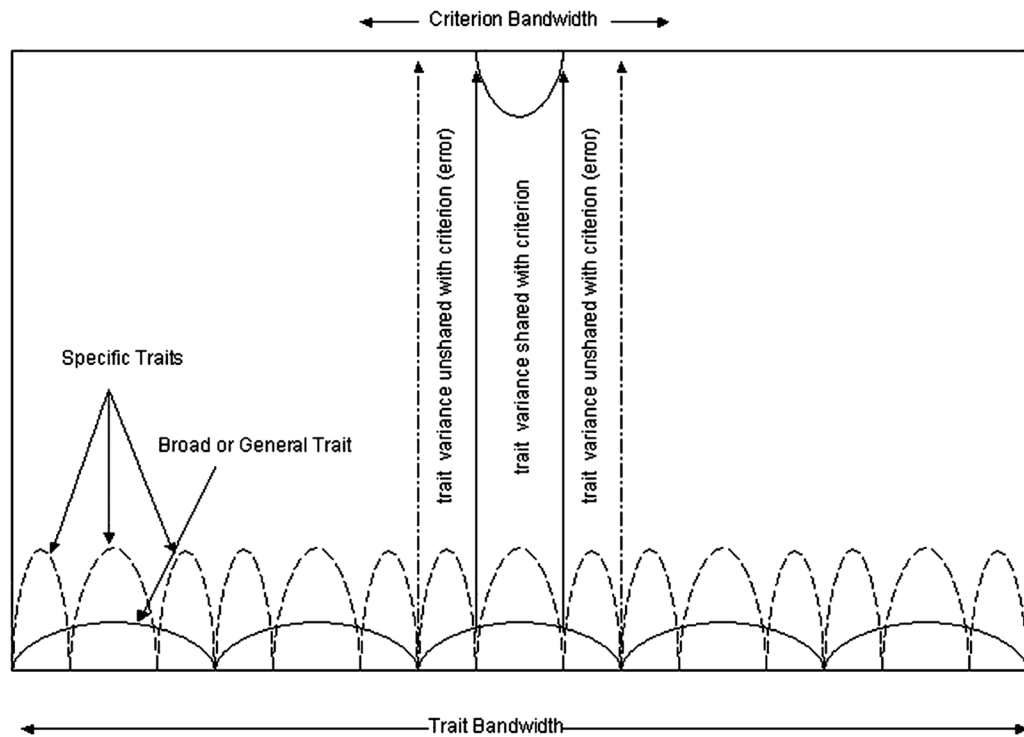
- Akaike, H. Information theory and an extension of the maximum likelihood principle. In: Petrov, BN.; Csáki, F., editors. *Proceedings of the Second International Symposium on Information Theory*. Budapest, Hungary: Akadémiai Kiadó; 1973. p. 267-281.
- Anastasi, A.; Urbina, S. *Psychological testing*. 7. Upper Saddle River, NJ: Prentice Hall; 1997.
- Ashton MC. Personality and job performance: The importance of narrow traits. *Journal of Organizational Behavior*. 1998; 19:289-303.
- Ashton MC, Jackson DG, Paunonen SV, Helmes E, Rothstein MG. The criterion validity of broad factor scales versus specific facet scales. *Journal of Research in Personality*. 1995; 29:432-442.
- Beauducel A, Whittmann WW. Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling*. 2005; 12:41-75.
- Bentler PM. A lower-bound method for the dimension-free measurement of internal consistency. *Social Science Research*. 1972; 1:343-357.
- Bentler, PM. Covariance structure models for maximal reliability of unit weighted composites. In: Lee, SY., editor. *Handbook of structural equation models*. Amsterdam: Elsevier; (in press)



- Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. Testing structural equation models. Newbury Park, CA: Sage; 1993. p. 136-162.
- Byrne BM. Factor analytic models: Viewing the structure of an assessment instrument from three perspectives. *Journal of Personality Assessment*. 2005; 85:17–32. [PubMed: 16083381]
- Carstensen LL. Evidence for a lifespan theory of socioemotional selectivity. *Current Directions in Psychological Science*. 1995; 4:151–156.
- Carver CS. How should multifaceted personality constructs be tested? Issues illustrated by self-monitoring, attributional style, and hardiness. *Journal of Personality and Social Psychology*. 1989; 56:577–585. [PubMed: 2651646]
- Cattell RB, Krug SE. The number of factors in the 16PF: A review of the evidence with special emphasis on methodological problems. *Educational and Psychological Measurement*. 1986; 46:509–522.
- Chapman, B. Unpublished doctoral dissertation. University of North Texas; CITY: 2005. Emotional intelligence at mid life: A cross sectional investigation of factorial variance, social correlates, and relationships to established personality and ability taxonomies.
- Chapman B, Hayslip B. Incremental validity of a measure of trait emotional intelligence. *Journal of Personality Assessment*. 2005; 85:154–169. [PubMed: 16171416]
- Charles EP. The correction for attenuation due to measurement error: Clarifying concepts and creating confidence sets. *Psychological Methods*. 2005; 10:206–226. [PubMed: 15998178]
- Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*. 2002; 9:233–255.
- Cortina JM. What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*. 1993; 78:98–104.
- Costa, PT.; McCrae, RR. Neo-PI, revised: Technical manual. Odessa, FL: Psychological Assessment Resources; 1992.
- Costa PT, McCrae RR. Domains and facets: Hierarchical personality assessment using the revised NEO Personality Inventory. *Journal of Personality Assessment*. 1995; 64:21–50. [PubMed: 16367732]
- Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951; 12:1–16.
- Cronbach, LJ. Essentials of psychological testing. 3. New York: Harper & Row; 1970.
- Cronbach, LJ.; Gleser, GC. Psychological tests and personnel decisions. Urbana, IL: University of Illinois Press; 1965.
- Digman JM. Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*. 1990; 41:417–440.
- Erikson, EH. Childhood and society. 2. New York: Norton; 1963.
- Eysenck, H. Dimensions of personality. New Brunswick, NJ: Transaction Publishers; 1998.
- Goldberg LR. An alternative “description of personality”: The Big-Five factor structure. *Journal of Personality and Social Psychology*. 1990; 59:1216–1229. [PubMed: 2283588]
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6:1–55.
- Harris RL, Elliot AM, Holmes DS. The timing of psychosocial transitions and changes in women’s lives: An examination of women aged 45–60. *Journal of Personality and Social Psychology*. 1986; 51:409–416. [PubMed: 3746621]
- Hogan J, Roberts B. Issues and non-issues in the fidelity-bandwidth trade-off. *Journal of Organizational Behavior*. 1996; 17:627–637.
- Jang KL, McCrae RR, Angleitner A, Riemann R, Livesley WJ. Heritability of facet-level traits in a cross-cultural twin sample: Support for a hierarchical model of personality. *Journal of Personality and Social Psychology*. 1998; 74:1556–1565. [PubMed: 9654759]
- John, OP.; Srivastava, S. The big five taxonomy: History, measurement, and theoretical perspectives. In: Pervin, LA.; John, OP., editors. Handbook of personality: Theory and research. 2. New York: Guilford; 1999. p. XXX-XXX.
- Levinson DJ. A conception of adult development. *American Psychologist*. 1986; 41:3–13.

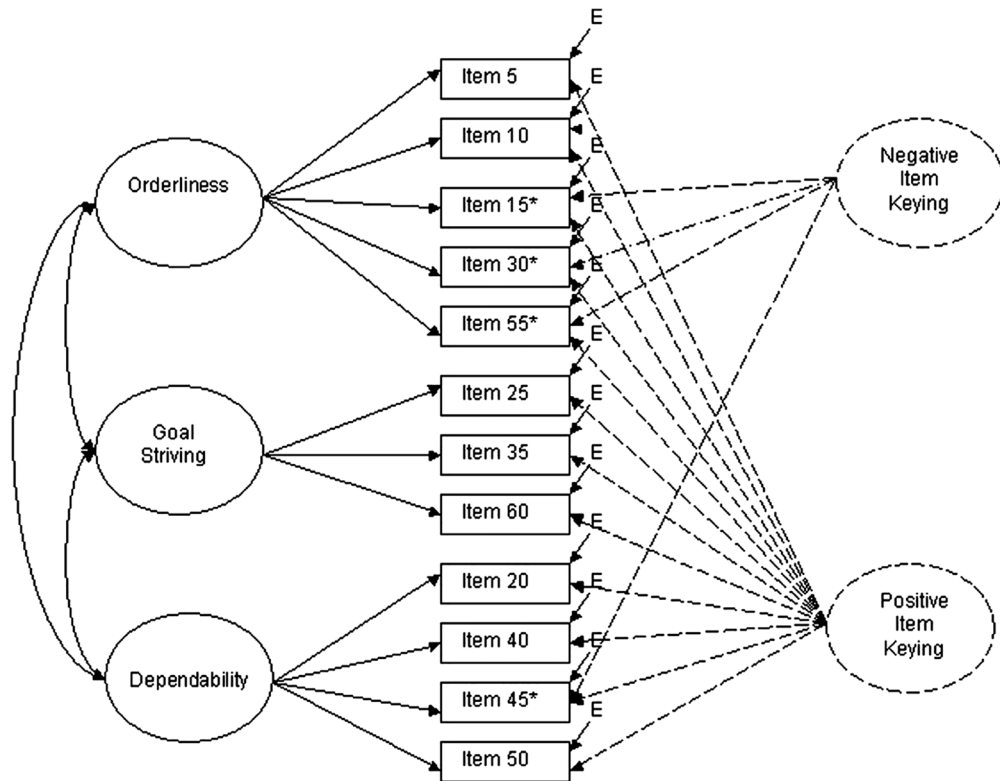
- Marsh HW, Hau KT, Wen Z. In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*. 2004; 11:320–341.
- McCrae RR. Well-being scales do not measure social desirability. *Journal of Gerontology*. 1986; 41:390–392.
- McCrae RR, Costa PT. Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*. 1983; 51:882–888.
- McCrae RR, Costa PT Jr. Personality trait structure as a human universal. *American Psychologist*. 1997; 52:509–516. [PubMed: 9145021]
- McCrae, RR.; Costa, PT. *Personality in adulthood*. 2. New York: Guilford; 2003.
- Murray G, Rawlings D, Allen NB, Trinder J. NEO–Five Factor Inventory scores: Psychometric properties in a community sample. *Measurement and Evaluation in Counseling and Development*. 2003; 36:140–149.
- Nunnally, JC.; Berstein, IH. *Psychometric theory*. 3. New York: McGraw-Hill; 1994.
- Olsson U. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*. 1979; 44:443–460.
- Ones DS, Viswesvaran C. Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior*. 1996; 17:609–626.
- Ones DS, Viswesvaran C, Schmidt FL. Meta-analysis of integrity test validities. *Journal of Applied Psychology*. 1993; 78:679–703.
- Paunonen SV. Hierarchical organization of personality and prediction of behavior. *Journal of Personality and Social Psychology*. 1998; 74:538–556.
- Paunonen SV. Big five factors of personality and replicated predictions of behavior. *Journal of Personality and Social Psychology*. 2003; 84:411–424. [PubMed: 12585813]
- Paunonen SV, Ashton MC. Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology*. 2001; 81:524–539. [PubMed: 11554651]
- Ployhart RE, Oswald FL. Applications of mean and covariance structure analysis: Integrating correlational and experimental approaches. *Organizational Research Methods*. 2004; 7:27–65.
- Ray JJ. Lie scales and the elderly. *Personality and Individual Differences*. 1988; 9:417–418.
- Ray JJ, Lovejoy FH. Age-related social desirability responding among Australian women. *The Journal of Social Psychology*. 2003; 143:669–671. [PubMed: 14609059]
- Raykov T. Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*. 1997a; 21:173–184.
- Raykov T. Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research*. 1997b; 32:329–353.
- Raykov T. On the use of confirmatory factor analysis in personality research. *Personality and Individual Differences*. 1998; 1998:291–293.
- Rhodes RE, Courneya KS. Relationships between personality, an extended theory of planned behaviour model and exercise behaviour. *British Journal of Health Psychology*. 2003; 8:19–36. [PubMed: 12643814]
- Rhodes RE, Courneya KS, Jones LW. Personality, the theory of planned behavior, and exercise: A unique role for extroversion's activity facet. *Journal of Applied Social Psychology*. 2002; 32:1721–1736.
- Russell D, Peplau L, Cutrona C. The revised UCLA loneliness scale: Concurrent and discriminant validity evidence. *Journal of Personality and Social Psychology*. 1980; 39:472–480. [PubMed: 7431205]
- Saucier G. Replicable item-cluster subcomponents in the NEO Five-Factor Inventory. *Journal of Personality Assessment*. 1998; 70:263–276. [PubMed: 9697330]
- Saucier G, Ostendorf F. Hierarchical subcomponents of the big five personality factors: A cross-language replication. *Journal of Personality and Social Psychology*. 1999; 76:613–627. [PubMed: 10234848]
- Saris WE, Aalberts C. Different explanations for correlated disturbance terms in MTMM studies. *Structural Equation Modeling*. 2003; 10:193–213.

- Satorra A, Bentler PM. Scaling corrections for chi-square statistics in covariance structure analysis. *Proceedings of the American Statistical Association*. 1988:308–313.
- Satorra A, Bentler PM. A scaled chi-square test statistic for moment structure analysis. *Psychometrika*. 2001; 66:507–514.
- Schinka JA, Kinder B, Kremer T. Research validity scales for the NEO–PI–R: Development and initial validation. *Journal of Personality Assessment*. 1997; 7:395–408.
- Strahan R, Gerbasi KC. Short, homogenous versions of the Marlowe–Crown Social Desirability Scale. *Journal of Clinical Psychology*. 1972; 28:191–193.
- Streiner DL. Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of Personality Assessment*. 2003a; 80:217–222. [PubMed: 12763696]
- Streiner DL. Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*. 2003b; 80:99–103. [PubMed: 12584072]
- Tanaka, JS. Multifaceted conceptions of fit in structural equation models. In: Bollen, KA.; Long, JS., editors. *Testing structural equation models*. Newbury Park, CA: Sage; 1993. p. 10-39.
- Terraciano A, McCrae RR, Brand LJ, Costa PT. Hierarchical linear modeling analyses of the NEO–PI–R scales in the Baltimore Longitudinal Study of Aging. *Psychology and Aging*. 2005; 20:493–506. [PubMed: 16248708]
- Watson D, Clark LA, Harkness AR. Structures of personality and their relevance to psychopathology. *Journal of Abnormal Psychology*. 1994; 103:18–31. [PubMed: 8040477]
- Weiss A, Costa PT. Domain and facet personality predictors of all-cause mortality among medicare patients aged 65 to 100. *Psychosomatic Medicine*. 2005; 67:724–733. [PubMed: 16204430]
- Weiss A, Costa PT, Karuza J, Duberstein PR, Friedman B, McCrae RR. Cross sectional age differences in personality among Medicare patients aged 65 to 100. *Psychology and Aging*. 2005; 20:182–185. [PubMed: 15769223]
- Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*. 2000; 3:4–70.
- Yuan KH, Bentler PM. On chi-square difference and z tests in mean and covariance structure analysis when the base model is misspecified. *Educational and Psychological Measurement*. 2004; 64:737–757.
- Zelenski JM, Rusting CL, Larsen RJ. Consistency in the time of experiment participation and personality correlates: A methodological note. *Personality and Individual Differences*. 2003; 34:547–558.
- Zinbarg RE, Revelle W, Yovel I, Li W. Cronbach's alpha, Revelle's beta, and McDonald's omega<sub>H</sub>: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*. 2005; 70:1–11.
- Zitzow D. The college adjustment rating scale. *Journal of College Student Personnel*. 1984a; 25:160–164.
- Zitzow, D. The college adjustment rating scale (CARS) manual. Detroit Lakes, MN: Personal Development Consultants; 1984b.



**FIGURE 1.**

The same set of personality items may be aggregated into a smaller number of broad, general traits or a greater number of more specific, narrow traits. Often only portions of variance within broad traits (i.e., a specific narrow trait subcomponent) may be relevant to a criterion of interest.



**FIGURE 2.** In this trait-method model, variance in the 12 items of the NEO-FFI Conscientious domain is partitioned into three latent trait factors, two method factors corresponding to the direction of item keying, and residual or item-specific variance.



TABLE 1

## Saucier's (1998) NEO–Five Factor Inventory Item Cluster Subcomponents

Domain and Subcomponents	Items	Possible Description of High Scorers
Neuroticism (alternative 1)		
Self Reproach	6, 21, 26, 36, 41, 51, 56	Feels inferior, worthless, helpless, reactive, tense, ashamed
Negative Affect	1, 11, 16, 31, 46	Worried, stressed, anxious, blue, depressed
Neuroticism (alternative 2)		
Self Reproach	6, 11, 26, 51	Feels inferior, worthless, helpless, stressed
Anxiety	1, 21, 31	Worried, fearful, tense, anxious
Depression	16, 41, 46	Blue, discouraged, sad, depressed
Extraversion		
Positive affect	7, 12, 37, 42	Light-hearted, cheerful, optimistic
Sociability	2, 17, 27, 57	Gregarious, enjoys others, prefers company
Activity	22, 32, 47, 52	Energetic, active, fast paced, action seeking
Openness		
Aesthetic interests	13, 23, 43	Artistic, poetic, aesthetically sensitive
Intellectual interests	48, 53, 58	Abstract, philosophical, intellectual
Unconventionality	3, 8, 18, 38	Nonconforming, free thinking, whimsical
Agreeableness		
Nonantagonistic orientation	9, 14, 19, 24, 29, 44, 54, 59	Cooperative, trusting, amiable, conflict avoidant
Prosocial orientation	4, 34, 39, 49	Actively courteous and considerate, well-liked
Conscientiousness		
Orderliness	5, 10, 15, 30, 55	Methodical, neat, organized, efficient
Goal-striving	25, 35, 60	Goal-driven, hard working, motivated to excel
Dependability	20, 40, 45, 50	Reliable, consistent, dependable

*Note.* Saucier (1998) offered two alternative scoring schemes for the Neuroticism items: The first was empirically derived, the second rationally derived to disentangle anxiety and depression. Possible descriptions of high scorers based on item wording.

Reverse keyed item.

**TABLE 2**  
Fit Indexes for Confirmatory Factor Models of NEO-Five Factor Inventory (NEO-FFI) Subcomponents by Domain

NEO-FFI Domain/ Factor Model	SB <sup>2</sup>	df	p	CFI	RMSEA	AIC
Neuroticism (scoring alternative 1)						
Self-reproach, negative affect						
Young adults, trait factors only	98.74	53	<.001	.94	.053	-7.26
Middle-age adults, trait factors only	95.18	53	<.001	.95	.062	-10.82
Young adults, trait + method factors	54.44	41	.077	.98	.038	-27.56
Middle-age adults, trait + method factors	55.81	41	.060	.98	.042	-26.19
Multiple group, trait + method factors	173.07	113	<.001	.97	.032	-52.93
Neuroticism (scoring alternative 2)						
Self-reproach, anxiety, depression						
Young adults, trait factors only	67.15	32	<.001	.92	.060	3.15
Middle-age adults, trait factors only	76.29	32	<.001	.91	.082	12.29
Young adults, trait + method factors	30.96	22	.097	.98	.037	-13.04
Middle-age adults, trait + method factors	26.05	22	.250	.99	.030	-17.95
Multiple group, trait + method factors	100.70	73	.018	.97	.027	-45.30
Extraversion						
Sociability, positive affect, activity						
Young adults, trait factors only	111.19	51	<.001	.93	.063	9.19
Middle-age adults, trait factors only	128.61	51	<.001	.89	.086	26.61
Young adults, trait + method factors	62.78	39	<.009	.97	.045	-15.22
Middle-age adults, trait + method factors	76.32	39	<.001	.95	.068	-1.68
Multiple group, trait + method factors	122.98	74	<.001	.93	.036	-25.02
Openness						
Aesthetic interests, intellectual interests, unconventionality						
Young adults, trait factors only	70.98	32	<.001	.88	.064	6.98
Middle-age adults, trait factors only	57.05	32	.004	.91	.062	-6.95
Young adults, trait + method factors	48.21	22	<.001	.92	.063	4.21
Middle-age adults, trait + method factors	21.33	22	.500	1.00	.000	-22.67
Multiple group, trait + method factors	122.98	74	<.001	.925	.036	-25.02
Agreeableness						

NEO-FFI Domain/ Factor Model	SB <sup>2</sup>	df	p	CFI	RMSEA	AIC
Nonantagonistic orientation, prosocial orientation						
Young adults, trait factors only	108.78	53	<.001	.89	.059	2.78
Middle-age adults, trait factors only	114.47	53	<.001	.87	.075	8.47
Young adults, trait + method factors	60.08	41	.027	.96	.040	-21.92
Middle-age adults, trait + method factors	68.00	41	.005	.94	.056	-14.00
Multiple group, trait + method factors	172.01	108	<.001	.94	.034	-43.99
Conscientiousness						
Orderliness, goal striving, dependability						
Young adults, trait factors only	140.67	51	<.000	.85	.077	38.67
Middle-age adults, trait factors only	74.68	51	.017	.96	.048	-27.32
Young adults, trait + method factors	63.98	39	.007	.96	.032	-30.76
Middle-age adults, trait + method factors	47.24	39	.171	.99	.047	-14.02
Multiple group, trait + method factors	175.32	109	<.001	.95	.035	-42.68

Note. SB<sup>2</sup> = Satorra-Bentler Scale corrected chi-square; CFI = comparative fit index; RMSEA = root mean square error of approximation; AIC = akaike information criteria. Multiple-group models represent the final factor model after nested model tests gradually constraining the most parameters possible to equality across groups without deterioration of model fit.

TABLE 3

Internal Consistency Reliability Estimates of Subcomponents in Each Sample

NEO-FFI Domain/Sub Components	Middle-Age Adults		Young Adults		Social Desirability r
	Essentially Equivalent?	Social Desirability r	Essentially Equivalent?	Social Desirability r	
Neuroticism (alternative 1)	—	.86 .87 <u>.92</u>	—	.84 .84 <u>.89</u>	-.25*
Negative affect	No	.74 <u>.74</u> .81	Yes	.66 .66 <u>.73</u>	-.17*
Self-reproach	No	.87 <u>.88</u> .90	No	.84 <u>.85</u> .90	-.25*
Neuroticism (alternative 2)	—	—	—	—	—
Anxiety	Yes	.62 .62 <u>.62</u>	Yes	.59 .59 <u>.59</u>	-.21*
Depression	Yes	.66 .69 <u>.69</u>	Yes	.56 .61 <u>.61</u>	-.10
Self-reproach	No	.72 <u>.73</u> .75	No	.70 <u>.71</u> .71	-.26*
Extraversion	—	.85 .85 <u>.93</u>	—	.85 .85 <u>.91</u>	-.03
Positive affect	No	.79 <u>.80</u> .84	No	.71 <u>.73</u> .77	.01
Sociability	No	.67 <u>.68</u> .71	Yes	.63 .64 <u>.68</u>	.02
Activity	Yes	.72 <u>.72</u> .78	No	.74 <u>.74</u> .77	-.09
Openness	—	.72 <u>.73</u> <u>.85</u>	—	.83 .83 <u>.91</u>	.05
Aesthetic interests	Yes	.81 .81 <u>.81</u>	Yes	.67 .69 <u>.69</u>	.05
Intellectual interests	Yes	.52 .53 <u>.53</u>	Yes	.64 .65 <u>.65</u>	.09
Unconventionality	No	.27 <u>.47</u> .48	No	.01 <u>.34</u> .24	-.05
Agreeableness	—	.83 .83 <u>.91</u>	—	.77 <u>.77</u> <u>.87</u>	.18*
Nonantagonistic orientation	No	.76 <u>.77</u> .86	No	.67 <u>.67</u> .75	.26*
Prosocial orientation	No	.73 <u>.74</u> .79	Yes	.78 .78 <u>.78</u>	.03
Conscientiousness	—	.88 .88 <u>.93</u>	—	.79 .79 <u>.89</u>	.02
Orderliness	No	.74 <u>.75</u> .80	No	.55 <u>.57</u> .66	.07
Goal striving	No	.78 <u>.80</u> .80	No	.72 <u>.75</u> .75	.01
Dependability	Yes	.75 .75 <u>.78</u>	No	.61 <u>.63</u> .72	-.03

Note. NEO-FFI = NEO-Five Factor Inventory; = Cronbach's (1951) alpha; = Raykov's (1997b) congeneric reliability estimate rho; = Bentler (1972) lower bound, dimension free, reliability estimate theta; Social desirability r = correlations with 10-item Marlowe-Crowne Social Desirability Scale. The most appropriate internal consistency estimate is underlined, depending on whether the subscale meets the assumption of essential tau-equivalence. Because domains were considered multidimensional rather than congeneric, their essential tau-equivalence was not tested and may be considered an appropriate reliability estimate.

\*  
 $p < .01$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



TABLE 4

Self-reproach 2 (N)	Anxiety (N)	Depression (N)	Positive affect (E)	Sociability (E)	Activity (E)	Aesthetic interests (O)	Intellectual interests (O)	Unconventionality (O)	Nonantagonistic Orientation (A)	Orderliness (C)	Goal striving (C)	Depend- ability
			-.36	-.35	-.24	-.05	-.15	-.01	-.25	-.22	-.35	-.28
			-.31	-.24	-.23	-.02	-.09	.05	-.16	-.11	-.17	-.15
	.58	.53	-.32	-.28	-.21	.02	-.13	-.01	-.17	-.14	-.30	-.21
	.57	.46	-.27	-.25	-.07	-.10	-.09	-.06	-.30	-.16	-.16	-.13
	.59	-.37	-.37	-.32	-.36	-.05	-.12	.06	-.14	-.21	-.30	-.26
-.43	-.49	-.48	.57	.55	.55	.07	.09	.02	.26	.56	.21	.34
-.33	-.23	-.27	.50	.53	.53	.04	-.03	-.07	.11	.30	.21	.30
-.33	-.17	-.31	.43	.50	.50	.08	.18	-.10	-.10	.30	.24	.35
-.10	-.01	-.02	.11	.15	.16		.47	.22	.09	.16	.03	.07
-.30	-.00	-.08	.05	.08	.27	.46		.21	.02	.19	.03	.14
.03	-.02	.09	-.18	-.16	-.14	.24	.23		-.05	.21	-.13	.03
-.29	-.04	-.28	.32	.11	-.01	.20	.20	-.10		.43	.13	.12
-.26	-.11	-.25	.49	.28	.23	.13	.13	-.17	.50		.12	.39
-.35	-.18	-.29	.29	.32	.20	-.01	.03	-.17	.29	.37	.34	.37
-.37	-.08	-.37	.38	.44	.41	.06	.06	-.22	.29	.49	.51	.53
-.38	-.13	-.20	.32	.25	.19	.12	.11	-.15	.41	.55	.54	.54

above diagonal. The two and three-subscale scoring alternatives for the Neuroticism domain are mutually exclusive and therefore correlations between them are not reported. Correlations between subscales

O = Openness; C = Conscientious.

**TABLE 5**  
 Comparison of Domain and Subcomponents Means Between Midlife and Young Adults

Domain and Subcomponent	Middle-Aged Adults		Young Adults		t	d
	M	SD	M	SD		
<b>Domains</b>						
Neuroticism	20.00	8.41	24.14	8.09	-5.60****	-.50
Extraversion	27.35	7.16	27.74	7.37	-.59	-.05
Openness	24.75	5.97	27.78	5.64	-5.81****	-.53
Agreeableness	31.52	5.94	29.01	5.65	4.80****	.44
Conscientiousness	33.71	6.43	28.29	6.22	9.46****	.86
<b>Subcomponent</b>						
Negative affect	10.07	3.86	11.61	3.29	-4.85****	-.44
Self-reproach	9.93	5.53	12.53	5.43	-5.29****	-.48
Anxiety	5.72	2.49	6.38	2.30	-3.09**	-.28
Depression	5.09	2.50	6.09	5.68	-2.39**	-.22
Self-reproach (Alternate)	5.73	3.09	7.38	3.26	-5.75****	-.52
Positive affect	10.34	9.48	10.49	2.94	-0.26	-.02
Sociability	7.80	2.87	8.06	2.78	-1.03	-.09
Activity	9.20	2.91	9.14	3.11	0.22	.02
Aesthetic interests	6.40	2.91	7.02	2.60	-2.52****	-.23
Intellectual interests	5.99	2.03	6.31	2.15	-1.70*	-.15
Unconventionality	7.80	2.26	9.94	2.11	-10.93****	-.99
Nonantagonistic orientation	19.30	4.73	17.69	4.46	3.91****	.35
Prosocial orientation	12.57	4.91	11.80	2.73	2.27**	.20
Orderliness	13.85	3.40	11.14	3.11	9.28****	.84
Goal striving	7.76	1.85	6.76	2.23	5.34****	.48
Dependability	12.08	2.39	10.37	2.50	7.72****	.70

Note. *d* = Cohen's *d* measure of effect size for difference between means.

\*  $p < .10$ .  
\*\*  $p < .05$ .  
\*\*\*  $p < .01$ .  
\*\*\*\*  $p < .001$ .

TABLE 6

## Domain and Subcomponents Criterion Correlations

NEO-FFI Domain and Subcomponent	Mid Life Adults					Young Adults				
	Life Satisfaction	Job Satisfaction	Marital Satisfaction	Artistic Hobbies	Loneliness	Opposite Sex Socializing	Feeling Anxious	Feeling Depressed	Contemplating Suicide	Loneliness
Neuroticism	-.39	<u>.41</u>	-.24	-.23	-.24	.51	.50	.52	.38	.49
Anxiety	-.27	-.34	-.19	-.14	-.18	.43	.41	.53	.32	.46
Depression	-.37	-.46	-.30	-.24	-.30	.52	.36	.48	.29	.44
Self-reproach	-.34	-.40	-.27	-.18	-.21	.56	.44	.52	.33	.53
Extraversion	.27	<u>.29</u>	.28	.30	.30	-.64	.28	-.31	-.31	-.65
Positive affect	.31	<u>.35</u>	.21	.23	.23	-.63	.29	-.29	-.34	-.57
Sociability	.19	<u>.23</u>	.25	.30	.30	-.62	.29	-.24	-.28	-.63
Activity	.15	<u>.17</u>	.21	.24	.24	-.38	.17	-.24	-.20	-.49
Openness				.44	.49					
Aesthetic interests				.48	.53					
Intellectual interests				.24	.33					
Unconventionality				.16	.23					
Agreeableness	.27	<u>.29</u>	.20	.21	.21	.49				-.42
Nonantagonistic orientation	.28	<u>.32</u>	.21	.24	.24	-.44				-.29
Prosocial orientation	.17	<u>.20</u>	.09	.11	.11	-.53				-.45
Conscientiousness						-.43	.19	.21	-.28	-.31
Orderliness						-.43	.14	.16	-.19	-.41
Goal striving						-.33	.15	.17	-.26	-.39
Dependability						-.33	.18	.21	-.22	-.33
						-.46		.28		-.44

Note. Domain criterion correlations significant at  $p < .01$ , followed by subcomponent criterion correlations. Underlined numbers are correlations corrected for attenuation due to measurement error.