Stability and Change in Adult Personality Over 6 Years: Findings From the Victoria Longitudinal Study

Brent J. Small,1 Christopher Hertzog,2 David F. Hultsch,3 and Roger A. Dixon4

1Department of Gerontology, University of South Florida, Tampa.  
2School of Psychology, Georgia Institute of Technology, Atlanta.  
3Department of Psychology, University of Victoria, British Columbia, Canada.  
4Department of Psychology, University of Alberta, Edmonton, Canada.

Data from the Victoria Longitudinal Study were used to examine the 6-year longitudinal stability of personality in older adults. Personality was measured with the NEO Personality Inventory. The longitudinal sample consisted of 223 adults initially ranging from 55 to 85 years of age. Longitudinal confirmatory factor analyses were used to examine the stability of individual differences in change over time, and the stability of the longitudinal factor structure. The results indicated both substantial stability at the level of individual differences in change, as well as significant individual differences in change that were related to age and gender. Finally, the factor structure of personality was invariant over time but did not approximate simple structure for the five dimensions of personality. Our study of 6-year personality development provided both (a) a confirmation of early significant stability findings and (b) unique evidence for significant individual differences in late adulthood.

During the past several decades, there has been a great deal of interest in the question of whether adult personality is characterized by stability or instability (e.g., Caspi & Roberts, 1999; Costa & McCrae, 1994; Heatherton & Weinberger, 1994; Helson, Jones, & Kwan, 2002). The debate is not simply whether significant changes exist across the life span; it also includes differences in how personality should be properly conceptualized, measured, and statistically analyzed. Some personality and life-span theorists emphasize the potential plasticity of personality as a function of contextual variables and compensatory behavioral changes to biological aging (Baltes, Staudinger, & Lindenberger, 1999; Caspi & Roberts, 1999) and advocate a change-oriented and person-oriented approach to personality in adulthood (Helson & Srivastava, 2001; Roberts, 1997). Others emphasize the predominant stability of personality in adulthood after ontogenetic change is completed in young adulthood (McCrae et al., 2000), even while they acknowledge that life events can alter personality in midlife (Costa, Herbst, McCrae, & Siegler, 2000).

For us to understand divergent views, we find it critical to define the multiple aspects of stability that are relevant to understanding personality consistency and change (Caspi & Roberts, 1999). One aspect is the stability of individual differences over time—or differential stability; that is, the rank-order consistency of individuals’ personality attributes across time (Ozer, 1986). A lack of differential stability implies that individuals vary in the trajectories of personality change over the life span. Differential stability is often indexed by cross-time correlation coefficients (stability coefficients); by any account, the stability coefficients of major personality traits tend to be substantial in magnitude (Costa & McCrae, 1994). Roberts and Del Vecchio’s (2000) meta-analysis reported a gradual increase in differential stability in personality across the life span. Childhood was characterized as the period of least stability, with trait consistency reaching a plateau between the ages of 50 and 70 and remaining constant thereafter. Specifically, they reported that across a 6- to 7-year follow-up interval, the average stability coefficient for the oldest age group was .74.

Alternatively, absolute stability involves stability of mean levels of behavioral dispositions across the life span. Recent arguments favoring personality change in midlife and old age have focused on evidence for instability in the level of personality in midlife and old age (e.g., Helson et al., 2002; Jones & Meredith, 1996). Helson and colleagues, using the California Personality Inventory (CPI), reported reliable curvilinear trends in mean personality change in adulthood, with the largest changes associated with the domain of norm adherence, including responsibility, socialization, self-control, and good impression. Flexibility of behavior manifested linear declines across the adult life span. Effect size estimates for longitudinal change were small to moderate in magnitude. In contrast, longitudinal research with the NEO Personality Inventory (NEO-PI; Costa & McCrae, 1985) has indicated only modest longitudinal changes in level of personality (e.g., Costa & McCrae, 1988; 1992). On the basis of existing evidence and interpretations of the literature, one might suspect that the NEO-PI is less likely to manifest personality change than the CPI.

A third aspect of stability concerns the invariance of the structure of personality over time. In this case, the question is whether the organization of behavioral dispositions shifts across the life course, leading to changes in the correlations of different attributes. The hypothesis of structural change in personality can be addressed in a variety of ways (Block, 2001), including comparative factor analysis. In the present study, we examined longitudinal evidence for these three
aspects of stability (differential, absolute, and structural) of personality among a group of older adults measured over a 6-year interval.

We assessed stability and change in personality by using facets from the NEO-PI. It is based on the “Big Five” model of psychometric personality scales (Digman, 1990; John, 1989; Norman, 1963). Within this framework, personality is conceptualized as reflecting predominantly five higher-order factors: Neuroticism (N), Extroversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C). There is considerable evidence that the NEO-PI is stable at the level of individual differences, at least among older adults. For example, Costa and McCrae (1988) reported that the 6-year stability coefficients for N, E, and O were all above .80. In addition, when corrected for measurement error, the estimated stability coefficients were all above .90 for these three factors. Slightly lower stability coefficients were reported for A and C across a 3-year retest interval (.63 and .79, respectively).

Although the rank-order stability of the NEO-PI has been well established, much less attention has been focused on invariance in the longitudinal factor structure of the NEO-PI. Developmental psychologists (e.g., Baltes & Nesselroade, 1970; Labouvie, 1980) have long focused on this form of stability, referred to as measurement equivalence, as a necessary condition permitting unambiguous conclusions regarding absolute and differential stability of quantitative scales, such as the five NEO-PI scales (Maitland, Dixon, Hultsch, & Hertzog, 2001; Meredith & Horn, 2001). If there is change in the underlying factor structure, then quantitative changes in scales may be an artifact of changing measurement properties, or they may result from the fact that different constructs are being measured at each point in time (Horn & McArdle, 1992; Schaie & Hertzog, 1985).

Partly, the lack of attention to longitudinal factor invariance has been a function of the type of analytic techniques that have been traditionally applied to the analysis of this instrument (see Block, 1995). Although there have been some recent exceptions (e.g., Borkenau & Ostendorf, 1990; Church & Burke, 1994; Hofer, Horn, & Eber, 1997; Panter, Tanaka, & Hoyle, 1994), the majority of factor analytic studies on the NEO-PI have used exploratory techniques (e.g., Costa, McCrae, & Dye, 1991; Piedmont & Weinstein, 1993). McCrae and Costa (1989) demonstrated that the NEO-PI could be extracted into five orthogonal factors, each representing a specific dimension of personality. Although informative, these types of analyses pose problems when one attempts to draw firm conclusions regarding factorial invariance over time. Given the fundamental indeterminacy of exploratory factor analysis (EFA) solutions (Cunningham, 1978; Hertzog, 1990; Little, Lindenberger, & Nesselroade, 1999), it is difficult to accurately assess the extent to which the different solutions at different occasions of measurement correspond with one another (see Tisak & Meredith, 1990). This problem is made even more acute because longitudinal data typically pose their own unique specification problems, given that both factors and unique components can be expected to covary with themselves over time (Hertzog & Schaie, 1986). However, with the use of confirmatory factor analysis (CFA; Jöreskog & Sörbom, 1977), assessing the longitudinal invariance of the factor structure becomes a more straightforward procedure.

Another possible reason for the lack of confirmatory analytic-based research on the invariance of the structure of the NEO-PI is the relative difficulty in modeling this instrument with CFA. Several studies have reported that the facets from the NEO-PI do not adhere to a simple factor structure (e.g., Borkenau & Ostendorf, 1990; Church & Burke, 1994; Panter et al., 1994), whereby each facet corresponds to one and only one factor. One solution to this problem is the addition of cross-loadings based on modification indices (e.g., Church & Burke, 1994). Another solution is to reduce the number of facets being modeled. For the sake of conceptual clarity, in the present study we adopted the latter approach. Although this limits conclusions regarding the measurement properties of the entire NEO-PI, it allows us to formally examine the longitudinal stability of personality, which is the substantive focus of this paper.

Finally, although the available evidence indicates a substantial degree of differential stability in adult personality, it also suggests that the five factors measured by the NEO are not perfectly stable, even across relatively short follow-up periods. In one study, after correction for measurement error, between 10% and 20% of the reliable variance in N, E, and O across a 6-year follow-up interval was still left unaccounted for (Costa & McCrae, 1988). Just as a high score does not imply a perfect score, high stability is not synonymous with perfectly preserved rank orders of individuals over time (see Alwin, 1994; Hertzog & Nesselroade, 1987; Ozer, 1986). Given that there is less than perfect stability of individual differences in the five major personality factors, it stands to reason that individual differences in personality change during adulthood may exist and could be related to other variables, such as health status or life events (Jones & Meredith, 1996). However, there have been relatively few studies of individual differences in change for the five factors from the NEO-PI, let alone attempts to account for such changes with other variables. McCrae (1993) attempted to predict individual differences in change by the factor O, and three related psychological constructs. However, no consistent pattern of relationships emerged from the analyses. Moreover, some authors attributed observed individual differences in change largely to error of measurement (see also Costa & McCrae, 1988; Costa, McCrae, & Arenberg, 1980). Because we use CFA in the present study, we are able to attend to issues of measurement error when examining the extent of longitudinal stability in personality characteristics.

However, other studies have found evidence for individual differences in personality change that may be related to characteristics of the participants (e.g., de Frias, Dixon, & Bäckman, 2003; Kohn, 1980; Mroczek & Spiro, 2003). Recent longitudinal evidence using the CPI indicates reliable individual differences in rates of personality change, although the magnitude of variance in change was small, relative to the total variance between individuals (Helson et al., 2002). Using some of the same data, Roberts, Helson, and Klohnen (2002) reported significant individual differences in change that were related to the occurrence of major life events such as divorce or work force participation. Similarly, Schooler (1999) reported that lifestyles and substantively complex environments were associated with personality and psychological change. Finally, Costa and colleagues (2000) reported that personality was relatively stable in participants from the UNC Alumni Heart Study, but that the modest instability that was observed could
be attributed to responses to various characteristics, such as coronary artery disease. Thus, these results suggest the possibility of both individual differences in change and predictors of such variability.

To examine individual differences in personality, we used a latent change model (e.g., McArdle & Nesselroade, 1994), shown in Figure 1, to determine whether there were significant individual differences in change and if these changes were related to age, gender, years of education, or self-rated health. In a standard longitudinal factor analysis (e.g., Hertzog & Schaie, 1986), changes are estimated indirectly through the longitudinal correlations of the latent variables with themselves. In a latent change analysis, the model is reformulated so that initial level and change are directly estimated as latent variables.

In a standard longitudinal factor analysis (e.g., Hertzog & Schaie, 1986), changes are estimated indirectly through the longitudinal correlations of the latent variables with themselves. In a latent change analysis, the model is reformulated so that initial level and change are directly estimated as latent variables for each personality factor. Using N from Time 1 (N_{T1}) and Time 2 (N_{T2}) as an example, the latent change model would specify two equations that map these occasion-specific factors into N_{level} and N_{change} latent variables:

\[ N_{T1} = N_{level}, \]
\[ N_{T2} = N_{level} + N_{change}. \]

Differencing these two equations makes the latent change specification more clear:

\[ N_{change} = N_{T2} - N_{T1}. \]

The full latent change model specifies the unobserved and change variables for each latent variable (N, E, O, A, C), and in the case of two occasions of measurement, these parameters are just identified in Equations (1) and (2). Instead of estimating variances and covariances of all occasion-specific factors, as in the standard longitudinal factor analysis, the model estimates variances and covariances of level and change latent variables (along with associated standard errors of estimate). With two occasions of measurement, the latent change model is a simple reparameterization of the longitudinal factor model, and it has an identical fit to the data (see McArdle & Nesselroade, 1994).

The present longitudinal data come from the Victoria Longitudinal Study (VLS; Hultsch, Hertzog, Dixon, & Small, 1998), in which a group of adults initially aged 55 to 85 years of age were administered the NEO-PI at baseline and again 6 years later.

**METHODS**

**Participants**

In the first sample (Sample 1), 484 adults, ranging in age from 55 to 85 years, were tested at baseline and reexamined 3 and 6 years after the original assessment. This sample was used to develop initial cross-sectional and longitudinal confirmatory factor models for the NEO-PI. The second sample was recruited from the same study population using similar recruitment procedures (Sample 2; for a description of this sample see Dixon, Wahlin, Maitland, Hultsch, Hertzog, & Bäckman, in press). This cross-sectional sample contains 520 adults, ranging in age from 55 to 92 years, and was used primarily to replicate the cross-sectional CFAs derived from Sample 1.

Table 1 displays the demographic characteristics and personality scale scores for both samples. Because of missing data from 10 persons in Sample 1 and 13 persons from Sample 2, 474 and 507 persons, respectively, were available for the analysis of the cross-sectional personality data. Multivariate analyses of variance computed on the demographic characteristics and personality scores revealed significant differences between the groups (Wilks \( \lambda = .924, F(4, 963) = 19.70, \) and \( p < .001; \) Wilks \( \lambda = .966, F(5, 971) = 6.74, \) and \( p < .001, \) respectively). Sample 2 tended to be younger \( [F(1, 966) = 10.81, \) \( p = .001, \) and \( \eta^2 = .01, \) had a higher proportion of female participants \( [\chi^2(1) = 8.81 \) and \( p < .01], \) and had more years of education \( [F(1, 966) = 58.89, \) \( p < .001, \) and \( \eta^2 = .06] \) than Sample 1. There were no differences in self-reported health status between the two samples. Among the personality characteristics, persons in Sample 2 had significantly higher ratings of O \( [F(1, 975) = 30.14, p < .001, \) and \( \eta^2 = .03]. \) There were no sample differences for N, E, A, or C.

**Measures**

**Personality.**—The NEO-PI (Costa & McCrae, 1985) consists of 181 statements (e.g., I really like most people I meet).
LONGITUDINAL STABILITY OF PERSONALITY

Table 1. Demographic Characteristics and Personality Scale Scores for Samples 1 and 2

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Sample 1</th>
<th>Sample 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( (n = 474) )</td>
<td>( (n = 507) )</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>69.10</td>
<td>67.83**</td>
</tr>
<tr>
<td>SD</td>
<td>5.80</td>
<td>7.46</td>
</tr>
<tr>
<td>Gender (% Female)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>13.16</td>
<td>14.78**</td>
</tr>
<tr>
<td>SD</td>
<td>3.14</td>
<td>3.39</td>
</tr>
<tr>
<td>Self-rated health**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>0.82</td>
<td>0.75</td>
</tr>
<tr>
<td>SD</td>
<td>0.76</td>
<td>0.72</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>76.28</td>
<td>76.64</td>
</tr>
<tr>
<td>SD</td>
<td>19.37</td>
<td>20.32</td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>98.51</td>
<td>101.96</td>
</tr>
<tr>
<td>SD</td>
<td>16.19</td>
<td>15.66</td>
</tr>
<tr>
<td>O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>108.55</td>
<td>114.67**</td>
</tr>
<tr>
<td>SD</td>
<td>18.20</td>
<td>17.59</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>49.95</td>
<td>50.72</td>
</tr>
<tr>
<td>SD</td>
<td>6.09</td>
<td>6.35</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>50.59</td>
<td>50.27</td>
</tr>
<tr>
<td>SD</td>
<td>7.25</td>
<td>8.06</td>
</tr>
</tbody>
</table>

*Relative to a perfect state of health (0 = very good to 4 = very poor). As a result of missing data, \( n = 465 \) in Sample 1.

*p < .01; **p < .001.

Participants indicate their agreement with each statement on a 5-point Likert scale (strongly agree to strongly disagree). The inventory is designed to measure five global domains of personality: N, E, O, A, and C. In addition, the N, E, and O domains consist of six specific facets, each measured by 8 items, that can be scored separately. The facets of N include Anxiety, Hostility, Depression, Self-Consciousness, Impulsiveness, and Vulnerability. The facets of E are Warmth, Gregariousness, Assertiveness, Activity, Excitement Seeking, and Positive Emotions. Finally, the facets of O tap Openness to Fantasy, Aesthetics, Feelings, Actions, Ideas, and Values. In the NEO-PI, Agreeableness and Conscientiousness are measured as global scales only, consisting of 18 items, unlike the newer version (NEO-PI-R; Costa & McCrae, 1992), which includes six facets for each of these two factors.

The NEO-PI has previously been demonstrated to have good reliability (Costa & McCrae, 1988), and it was found to have high internal consistency in the VLS as well. In Sample 1, Cronbach’s \( \alpha \) for N, E, O, A, and C was \( .92, .86, .89, .78, \) and \( .84 \), respectively. Equally high values were obtained in Sample 2 \( (.93, .86, .89, .79, \) and \( .86 \), respectively).

To examine the measurement properties of the instrument, we replicated the principal components analysis, with varimax rotation, that was reported by McCrae and Costa (1989), using information from Sample 1. The results of this analysis were in good agreement with those reported previously. For example, the amount of variance accounted for with the present sample was 62%, whereas 63% of variance was accounted for in the original paper. In addition, the average coefficient of congruence (Gorsuch, 1983) between the solution derived here and the one reported by McCrae and Costa was acceptable \( (M = .89) \). However, examination of the specific factor scales revealed that although some coefficients of congruence were quite high \( (N = .98; \ O = .98; \) and \( C = .97) \), others were much lower, suggesting an imperfect match between the VLS sample and the results reported by McCrae and Costa \( (E = .76; \ A = .76) \).

**Self-reported health.**—This domain was indexed by three measures: chronic illness, instrumental health, and self-rated health. For chronic illness the presence and severity of 26 chronic conditions were assessed. Participants were asked to indicate whether they had experienced each of the illnesses during the past 2 years, and, if so, whether it was “fairly serious” or “not too serious.” These reports were summed to yield an overall chronic illness score. Instrumental health required participants to rate the extent to which their health had caused them to change their level or pattern of daily activity in eight domains over the past 2 years. Participants were asked to rate each domain on a 5-point scale (no change to give up activity). Scores were summed across the eight items to yield an overall instrumental health score. Finally, self-rated health consisted of two items that asked participants to rate their own health on a 5-point scale (very good, good, fair, poor, or very poor) compared with a perfect state of health and compared with other people their own age. Scores on the two items were summed to yield an overall score.

**Procedure**

The test battery was administered across multiple sessions scheduled over a period of approximately 1 to 1.5 months. For Sample 1, there were three testing sessions (two group and one individual) at the first and second occasions of measurement and four testing sessions (two group and two individual) to accommodate an expanded cognitive battery at the third occasion of measurement. Each session was approximately 2 hr in length, with a rest break in the middle during which participants were served refreshments. In the group sessions, small groups ranging from 2 to 10 individuals were tested at a time. In this sample the NEO-PI was administered at only the first and second occasions of testing. Because data were available at these occasions only, they will be referred to as Time 1 and Time 2 throughout this report.

The average retest interval between Time 1 and Time 2 was 5.91 years (range = 5.27 to 6.77 years). The 6-year longitudinal sample was considerably smaller \( (n = 223) \) than the cross-sectional sample as a result of attrition (see Hultsch et al., 1998). This sample was more select in terms of the baseline demographic characteristics, compared with individuals who were not retested again [Wilks \( \lambda = .945, F(4, 458) = 6.80 \), and \( p < .001 \)]. Specifically, the longitudinal sample was younger at first test \( [M_s = 67.86 \text{ vs. } 70.11 \text{ years}; F(1, 461) = 18.04 \text{ and } p < .001 \] had more years of education \( [M_s = 13.58 \text{ vs. } 12.79; F(1, 461) = 7.52 \text{ and } p < .001 \] and rated their health more poorly \( [M_s = .73 \text{ vs. } .89; F(1, 461) = 5.20 \text{ and } p = .023] \), respectively. There were no significant differences in terms of numbers of chronic diseases, report of instrumental health,
or gender composition of the two groups. More important, both groups showed similar levels on the five NEO-PI personality scales at baseline [Wilks $\lambda = .983$, $F(4, 458) = 1.59$, and $p > .05$].

Given the 6-year retest plan for the NEO-PI, Sample 2 has only cross-sectional data. The test battery was administered over a 1- to 1.5-month period and participants attended four testing sessions (two group and two individual). The NEO-PI was administered at the final individual testing session.

**Statistical Analyses**

We used Bentler’s EQS program (version 5.7, Bentler, 1995) to estimate the cross-sectional and longitudinal confirmatory factor models for the NEO-PI facets. The confirmatory factor models use covariance matrices as input. However, it is much easier to evaluate standardized solutions where factors and variables are rescaled to z-score form after the covariance structure model is estimated. In the standardized solution, factor loadings are expressed as standardized regressions of variables on factors, and relationships among factors are expressed as factor correlations. Where important, the original parameter estimates, expressed in covariance metric, are reported.

Two features of the analyses should be outlined to aid in interpreting the statistical analyses. First, the standard practice in evaluating models involves attending to multiple goodness-of-fit indices. The $\chi^2$ test is almost always significant, indicating less than perfect fit to the data, even when the model does a good job of reproducing the sample covariance matrix. This problem with $\chi^2$ is especially pronounced when the sample size is large ($N > 200$), as it is in the present case (Bollen & Long, 1993). Hence, two alternative fit indexes are also reported. The first is the comparative fit index (CFI; Bentler, 1990), which is an index of the proportion of the information in the sample covariances accounted for by the model. Generally, fit indices above .90 are considered to be good. The second is the root mean square error of approximation (RMSEA). A RMSEA value less than 0.08 indicates an acceptable fit (Browne & Cudeck, 1993).

Second, the model parameters were developed and replicated in a variety of ways. Several observers (e.g., Breckler, 1990; Cudeck & Browne, 1983) have criticized the iterative procedure of producing a well-fitting model. Specifically, when the model modification procedure is done by using a single data set, the generalizability and validity of the resultant model may not be clear. The risk is that the final model may be specific to only that particular data set, and its generalizability may be severely limited. Cudeck and Browne (1983) recommend model cross-validation with different samples in order to minimize on the capitalization of chance. Accordingly, in the present context, we used Sample 1 to derive a good-fitting cross-sectional model of personality and Sample 2 to cross-validate that model.

**RESULTS**

The results consisted of three main sets of analyses. In the first section, we developed the CFA in Sample 1 and replicate the parameters in Sample 2. Next, we examined the 6-year longitudinal stability of the NEO-PI in terms of the cross-time stability coefficients, as well as the invariance of the factor loading pattern, factor covariances, and variances. Finally, we applied a latent change analysis to examine whether individual differences in personality change, if present, could be predicted by age, gender, years of education, or self-reported health.

**Cross-Sectional Factor Analysis**

CFAs were used to further examine the factor structure of the NEO-PI. We began by specifying a five-factor simple structure model in Sample 1, with each facet loading only on the factor that it was assumed to index. For example, the six N facets loaded only on the N factor. Because the factors A and C had only one scale, they were specified by fixing the error variance to a positive value based on the estimated reliability (Hayduk, 1987). The reliability of the A factor was estimated as .78 in the present sample. Therefore, the unique variance in this variable was fixed at 8.16 $[(1 - .78) \times (37.09)]$, which is the product of the 1 − reliability (1 − .78) and the sample variance of A (37.09). The fixed error variance for C was specified in a similar fashion [unique variance $= 8.41; (1 - .84) \times (52.56)]$. This approach corrects the estimated factor covariances for errors of measurement in A and C, despite the fact that they are measured by only one scale.

Analyses indicated that a CFA model proposing simple structure of the NEO-PI facets fit the data poorly [$\chi^2(162) = 1304.89, p < .001$; CFI = .63; RMSEA = .122]. This suggests that the model that specified that facets had loadings on only one factor did not reproduce the data well. Specifically, modification indices indicated that one of the sources of poor fit was that many of the facets were also associated with factors other than those that they were originally a marker of (e.g., Warmth to A, Fantasy to C).

To produce a good-fitting model, we reduced the number of facets chosen to index each of the factors. The facets selected to define the factors were based on the average intercorrelation among the facets used to define the factor, as well as from the highest loadings from an initial principal components analysis. Using this method, we selected the facets of Anxiety, Depression, and Self-Consciousness to define N; Gregariousness, Excitement Seeking, and Positive Emotions to define E; and Fantasy, Aesthetics, and Feelings to define O. The individual scales continued to define A and C.

Analyses positing a simple structure for the reduced set of indicators produced a better fit to the data [$\chi^2(36) = 211.30, p < .001$; CFI = .848; RMSEA = .101]. However, a number of significant modification indices revealed that the fit of the model could be improved with the addition of several additional cross-loadings. In particular, a positive loading of the Anxiety facet on O factor, a negative loading of Excitement Seeking on A, a positive loading of Positive Emotions on O, and a negative loading of Fantasy on C were observed in the revised model, which provided an excellent fit to the data [$\chi^2(32) = 83.13, p < .001$; CFI = .956; RMSEA = .058]. Table 2 shows the final factor solution for the CFA with the reduced set of NEO facets. In all cases, the factor loadings were statistically significant and the factors themselves were well defined. Further, a number of statistically significant factor correlations were observed (see Table 3).

Before modeling the observed factor structure longitudinally, we independently cross-validated the accepted model derived in Sample 1 in Sample 2. The fit was good [$\chi^2(32) = 115.19, p < .001$; CFI = .936; RMSEA = .072]. All specified factor
loadings were significantly different from zero, and the basic pattern of loadings was replicated. We further evaluated the comparability of the results by conducting a simultaneous factor analysis in both samples. The replication procedure began by using the measurement model produced in the first sample as the basic model without any constraints on parameters between the two samples, providing a test of configural invariance across the two samples. The fit indices indicated that the solution did hold \( \chi^2(64) = 191.68, p < .001; \) CFI = .946; RMSEA = .046. We then provided a test of metric equivalence by constraining the estimated factor loadings to be equal across the two samples. Although this did produce a slight loss of fit \( \chi^2(74) = 207.77, p < .001; \) CFI = .944; RMSEA = .044; \( \Delta \chi^2(10) = 16.09, p > .05 \), the difference was not statistically significant. The final step was to constrain the factor covariances equal across the two samples. Doing this altered the fit of the model only slightly \( \chi^2(84) = 215.53, p < .001; \) CFI = .945; RMSEA = .041; \( \Delta \chi^2(10) = 7.76, p > .05 \). In general, the cross-validation results demonstrated that the model derived in Sample 1 generalized to the independent sample of adults and therefore could be used as the basis for the longitudinal models to follow.

**Longitudinal CFA**

The general goal of this set of analyses was to examine whether there was evidence of longitudinal stability of the NEO factors in Sample 1. Longitudinal correlations of NEO-PI scales (a common index of stability of personality) are reported in Table 4 (first column). The data revealed uniformly high 6-year stability coefficients. The correlations for the original scales N, E, and O were all above .80, and the correlations of A and C longitudinally were approximately .70. The stabilities were also calculated with the factors derived from the reduced set of indicators that were used in the confirmatory factor models (see Table 4, middle column). The magnitude of these relationships was almost identical to those seen with the full-scale scores. We then examined longitudinal changes at the mean level (another common index of stability of personality) for the original NEO factors. A repeated measures multivariate analysis of variance was computed on the data from Time 1 and Time 2 and revealed an overall effect of time [Wilks’ \( \lambda = .94, F(5, 218) = 2.86, \) and \( p = .02 \)], indicating that statistically significant changes were observed. At the univariate level, only the O factor exhibited statistically significant changes over time \( F(1, 222) = 6.49, p = .01, \) with declines on this factor \( (M_{T1} = 110.64, M_{T2} = 108.97) \) being observed. The univariate analyses for N \( (M_{T1} = 74.93, M_{T2} = 73.82) \), E \( (M_{T1} = 99.13, M_{T2} = 98.99) \), A \( (M_{T1} = 50.33, M_{T2} = 50.82) \), and C \( (M_{T1} = 50.83, M_{T2} = 50.84) \) were not statistically significant.

The next step in our longitudinal analyses was to examine the stability of personality within a longitudinal factor analysis strategy (see Hertzog, 1990 for a review of this statistical method). We began by applying the structure derived from the cross-sectional sample to the data from both times of measurement. In addition to the basic model, additional covariances between the cross-occasion factors (e.g., N at NT1 and N at NT2) were also added in order to provide an estimate of the stability coefficients disattenuated for measurement unreliability. Cross-occasion covariances for the residual variances were added to the model. In the next set of analyses, we generally followed Meredith’s (1993) recommendations for evaluating factor invariance. First, to establish metric invariance, we constrained the estimated factor loadings at each time of measurement to be equal. Second, we ran a model in which the factor covariances were constrained equal, and this was followed by one that constrained the factor variances equal. These analyses allow us to examine both the stability of personality at the factor level and the stability of the structure of personality longitudinally. In all cases any difference in the fit of these models was assessed with sequential \( \chi^2 \) tests and comparisons of CFIs.

The initial longitudinal confirmatory factor model produced a good fit to the data \( \chi^2(171) = 252.38, p < .001; \) CFI = .973; RMSEA = .047. This model without constraints provides an indirect test of configural invariance over time (Hertzog & Schaie, 1986). In general, the factor structure was well replicated at both points in time. The analyses also revealed that the disattenuated stability coefficients produced by this analysis were substantially higher than the uncorrected stabilities seen earlier. Table 4 (third column) also reports the disattenuated stabilities; in most cases they were above .90.

Although the previous analysis suggested that the factor configuration remained stable across time, it did not tell us whether the specific parameters (factor loadings, factor covariances, or factor variances) remained invariant across time. The following models provided a statistical test of each of these hypotheses. The first model tested constrained the factor loadings to be invariant across time. This model produced
a good fit to the data \( \chi^2(181) = 263.52, p < .001; \text{CFI} = .972; \text{RMSEA} = .046 \), and the small loss of statistical fit was not statistically significant \( [\Delta \chi^2(10) = 11.14 \text{ and } p > .05] \). This indicated that the factor loadings exhibited metric invariance across the 6-year follow-up interval. The additional constraint that the within-occasion factor covariances be constrained equal across time also provided a good fit to the data \( \chi^2(191) = 281.84, p < .001; \text{CFI} = .970; \text{RMSEA} = .047; \Delta \chi^2(10) = 18.32, p = .05 \). Finally, the factor variances were constrained equal across time and the model was rerun. This set of constraints also failed to degrade the fit of the model significantly \( \chi^2(196) = 288.96, p < .001; \text{CFI} = .969; \text{RMSEA} = .046; \Delta \chi^2(5) = 7.12, p > .05 \). Taken together, these results indicate that there is evidence for stability in both the factor structure of the reduced NEO-PI and relative stability of individual differences in personality over the 6-year interval. This suggests that not only is the between-person level of stability relatively stable over time, but the nature of the interrelationships among the subscales and factors also remains invariant longitudinally.

### Longitudinal Latent Change Analysis

Although the previous results provided evidence of stability of personality at several levels of analysis, no direct test of factor-based individual differences in longitudinal changes in personality has yet been conducted. That is, although the cross-time stability coefficients were uniformly high, the 95% confidence intervals shown in Table 4 (right-hand column) for the disattenuated stability coefficients failed to include 1.0 for any of the factors. This indicates that individual differences in change do exist. In other words, the hypothesis of perfect 6-year stability, as manifested in a standardized stability coefficient of 1.0, could be rejected. Next, we proceeded to compute a more direct test of the hypothesis of true individual differences in personality change.

The two occasion-specific factors for the five personality dimensions were each reconfigured into two latent variables (a latent initial level or status factor and a latent change factor, over the 6-year longitudinal interval). All latent level and latent change factors were allowed to covary. The overall fit of the model was good \( \chi^2(201) = 339.35, p < .001; \text{CFI} = .968; \text{RMSEA} = .047 \). Table 5 reports the latent Time 2—Time 1 change variances, along with associated standard errors and \( z \) tests. For the latent change variables, the \( z \) statistics test the null hypothesis that the population variance of personality change is zero (i. e., a complete absence of individual differences in change). Consistent with the 95% confidence intervals, all of

The five factors exhibited significant individual differences in change across the follow-up interval. The correlations between initial level and change for the five personality factors are also shown in Table 5. In all cases, the negative correlations indicated that individuals who were initially lower on the personality dimensions exhibited greater increases across the follow-up period. All of the level–change correlations were statistically reliable, with the exception of E. Taken together, although the stability coefficients indicated substantial longitudinal stability among the NEO factors, this analysis revealed that in all cases there were significant individual differences as well.

Next, an evaluation of the significant individual differences in change involved relating these changes to demographic factors and self-reported health. To do this, we included age, gender, and years of education in the personality latent change model, together with initial level and change factors for self-reported health. The self-reported health factor was indexed by self-reports of the prevalence of chronic health conditions, ratings of instrumental health, and a subjective self-rated health measure (see Hultsch, Hertzog, Small, & Dixon, 1999). The overall fit of this new model was good \( \chi^2(337) = 452.46, p < .001; \text{CFI} = .965; \text{RMSEA} = .041 \). Further, the self-reported health latent change factor exhibited significant individual differences in change across the follow-up period (variance = .16; standard error = .04; \( z = .410; p < .001 \)). The correlations between initial level and changes in personality and the demographic and self-reported health measures are shown in Table 6. The correlations between initial level of the personality factors are shown in the upper half of Table 6. Several of the estimated correlations were statistically significant. Higher N scores at baseline were associated with female gender and more health conditions. Lower scores on E were significant correlated with older age. For the O factor, female gender was associated with higher scores. Female gender was also significantly correlated with higher A scores at baseline. Further, higher scores on A were associated with more years of education, better health at baseline, and greater negative change in health status over the follow-up period. Finally, initial level on C was not significantly correlated with any of the individual difference predictors.

For changes in personality (bottom half of Table 6), only three of the estimated correlations were statistically significant. For N, advanced age was associated with increases in N, whereas female gender was related to decreases in N across the retest interval. Gender was also related to changes in A, with women exhibiting increases in A over time. None of the other correlations with years of education or the two self-reported health factors with personality variables were statistically

### Table 4. Longitudinal Stabilities of the NEO-PI

<table>
<thead>
<tr>
<th>Personality</th>
<th>Scale</th>
<th>Reduced Facet</th>
<th>LFA (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Full Facet</td>
<td>.84</td>
<td>.84</td>
</tr>
<tr>
<td>E</td>
<td>.82</td>
<td>.82</td>
<td>.93 (89.97)</td>
</tr>
<tr>
<td>O</td>
<td>.85</td>
<td>.87</td>
<td>.94 (90.98)</td>
</tr>
<tr>
<td>C</td>
<td>.73</td>
<td>—</td>
<td>.88 (83.94)</td>
</tr>
<tr>
<td>A</td>
<td>.69</td>
<td>—</td>
<td>.91 (85.97)</td>
</tr>
</tbody>
</table>

Notes: NEO-PI = NEO Personality Inventory; LFA = longitudinal factor analysis; CI = Confidence interval; N = Neuroticism; E = Extroversion; O = Openness to Experience; C = Conscientiousness; A = Agreeableness.

### Table 5. Factor Latent Change Variances and Level–Change Correlations for the Five Personality Scales

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N_change</td>
<td>3.29</td>
<td>.74</td>
<td>4.42***</td>
<td>-.45***</td>
</tr>
<tr>
<td>E_change</td>
<td>.93</td>
<td>.46</td>
<td>2.00*</td>
<td>-.21</td>
</tr>
<tr>
<td>O_change</td>
<td>1.16</td>
<td>.38</td>
<td>3.04***</td>
<td>-.23*</td>
</tr>
<tr>
<td>C_change</td>
<td>10.60</td>
<td>2.49</td>
<td>4.26***</td>
<td>-.28*</td>
</tr>
<tr>
<td>A_change</td>
<td>5.06</td>
<td>1.78</td>
<td>2.84***</td>
<td>-.43**</td>
</tr>
</tbody>
</table>

Note: N = Neuroticism; E = Extroversion; O = Openness to Experience; C = Conscientiousness; A = agreeableness; SE = standard error.

*\(p < .05\); **\(p < .01\); ***\(p < .001\).
significant. Further, because the standard errors used to evaluate whether the factor covariances are statistically significant are unique to each comparison, there is not a perfect relationship between the magnitude of the standardized covariances (correlations) and statistical significance. This is the reason why the correlation between years of education and change in E is not statistically reliable, although the correlation is larger than the significant relationship between change in N and years of education or initial level nor changes in self-reported health were related to changes in personality across time.

Thus, our results suggest that personality change can be observed in old age, reinforcing recent reports from studies using the CPI (e.g., Helson et al., 2002; Roberts et al., 2002). The present data demonstrate that these changes can also be observed with the NEO-PI, discounting the possibility that reported personality change is an artifact of the method used to measure personality or the selection of particular personality facets in selected in a particular inventory.

From a life-span perspective, the finding of substantial personality change in old age is not surprising. Baltes and colleagues (e.g., Baltes et al., 1999) emphasized the potential importance of nonnormative life events in old age as triggers for psychological change, and they discussed the critical importance of individuals’ compensating for biological, psychological, and social losses in old age by adjusting goals, beliefs, and behaviors. Just as personality change in midlife may reflect adjustments to critical life events (Costa et al., 2000), personality change in old age could be caused by critical events such as retirement or loss of a spouse. It could also reflect more subtle and continuous adjustments in attitudes and behaviors such as retirement or loss of a spouse. It could also reflect more subtle and continuous adjustments in attitudes and behaviors associated with reduced personal control and adaptation to major chronic illness, such as osteoarthritis. Of course, our results could also be caused by ontogenetic processes that cannot be attributed to life events, social change, or psychological processes of adjustment. That is, as in accounts of early personality development (McCrae et al., 2000), nature could be seen as an important and ongoing process that leads to personality change in midlife.

The evidence for personality change we have just reported is consistent with previous research on personality change, which has been observed in various populations and contexts, including general populations, clinical samples, and older adults. The results of the present study reveal that personality change is not limited to young adulthood and can be observed in old age, providing further support for the idea that personality change is a lifelong process. This finding has important implications for understanding the nature and course of personality change, as well as for designing interventions aimed at promoting positive outcomes for older adults.
personality factor structure derived from CFA methods has not been previously published. Thus, our results indicate measurement equivalence for the NEO-PI scales over time in older samples. This outcome justifies treatment of quantitative measures of stability (i.e., mean changes in factor scores and individual differences in such changes) as reflecting change in equivalently defined constructs over time (Baltes & Nesselroade, 1970; Laboviu, 1980).

Our results are also relevant to issues regarding the definition of the five personality factors from the NEO-PI, which has been studied extensively in cross-sectional data. Although our evidence indicated invariance in factor structure of the NEO-PI, it is clear that the NEO-PI scales do not approximate simple structure in their mappings on the Big Five personality dimensions. Specifically, our initial simple factor structure did not fit the data, and it was only after reducing the number of facets and adding additional cross-loadings that we could achieve a good-fitting solution. These results are consistent with an increasing number of studies that have examined the factor structure of the NEO-PI with CFA techniques (Borkenau & Ostendorf, 1990; Church & Burke, 1994; McCrae et al., 2002; McCrae, Zonderman, Costa, Bond & Paunonen, 1996). For example, Church and Burke (1994) reported that only after the addition of 22 cross-loadings was a good fitting solution to the model derived. Similarly, McCrae and colleagues (1996) also failed to observe simple structure among the NEO factors. However, they concluded that this may reflect flaws in the statistical procedure of CFA, especially as it is applied to reducing the number of dimensions in personality tests. Although CFA may have some limitations in this regard, it is also important to note that the EFA procedures they typically use, principal axes factor extraction followed by a targeted (Procrustes) rotation, also have limitations and detractors (e.g., Horn & Knapp, 1973). In addition, Little and colleagues (1999) recently reported simulation results indicating that CFA consistently yielded more valid and unbiased estimates of the relationships among constructs, as compared with EFA. Finally, the lack of a simple confirmatory structure may also reflect conceptual, rather than solely statistical, grounds. Hofstee, de Raad, and Goldberg (1992) argued that the five factors represented in the NEO-PI are not truly orthogonal but reflect a blend of multiple factors (also see Goldberg, 2001). According to this view, the failure of an NEO-PI simple structure is not terribly surprising.

Our CFA results also suggested an oblique factor structure for the NEO-PI, in contrast to the assumption of orthogonal factors often applied to the NEO-PI (including McCrae & Costa, 1989; McCrae et al., 1996). Although the magnitude of factor correlations was relatively low, the concern is not trivial. If personality factors are truly correlated, a misspecified orthogonal rotation would absorb the factor loadings into the factor pattern weights (factor loadings), possibly biasing attempts to estimate the relationships of variables to factors (Little et al., 1999). Undoubtedly, the costs and benefits of both kinds of factor solutions for personality will continue to be evaluated. One clear advantage of the CFA specification is that it allows explicit tests of longitudinal invariance and optimal estimates of quantitative changes in factor scores over time.

Although the present results are informative, there are a number of limitations to the study that must be acknowledged. Foremost was the use of the NEO-PI, rather than the NEO-PI-R (Costa & McCrae, 1992). As a result, we were unable to model the A and C factors with individual facets as they were not available in this version of the NEO. However, we do not believe that this fact had a negative impact on the results seen here. Rather than modeling these factors with multiple facets, we constrained the error variance to the reliability estimate derived from the current data. A second limitation emanates from the research design. The 6-year follow-up interval available in the VLS may have limited our opportunity to observe even more sizable individual differences in change. Furthermore, the relatively short interval may have restricted the opportunity to observe even greater interrelationships for the demographic characteristics and self-reported health. In contrast, this study may be viewed as providing a relatively conservative test of the hypothesis regarding individual differences in personality change in late adulthood. The third limitation is the breadth of individual difference predictors of personality change that were examined here. Had other variables been examined, especially those potentially more relevant to changes in personality (e.g., significant life events and alterations in social structure), we may have witnessed additional predictive relationships with the changes in personality that we did observe. For example, Mroczek and Spiro (in press) recently examined life event predictors of individual differences in rate of change for neuroticism and extraversion in a sample of adults measured over 12 years. They reported that death of spouse and remarriage were associated with a decline in neuroticism. Such variables are not currently available in the VLS.

Overall, our study provides evidence for significant longitudinal stability and instability of the NEO-PI across a 6-year follow-up period. In addition, we found that a portion of this instability was related to participants’ age and gender. It is hoped that these results will help to stimulate future attempts in exploring the instability of personality with age, as well as to predict these individual differences with changes in health status, social structure, and the occurrence of major life events.

Acknowledgments
This research was supported by Grant AG08235 from the National Institute on Aging. We thank all staff members and volunteer participants of the VLS for collaboration and data collection.

Address correspondence to Brent J. Small, Department of Gerontology—SOC 204, University of South Florida, 4202 E. Fowler Avenue, Tampa, FL 33620. E-mail: bsmall@luna.cas.usf.edu

Correspondence concerning the VLS should be addressed to Roger Dixon, Department of Psychology, University of Alberta, Edmonton, Alberta T6G 2E9, Canada. E-mail: rdixon@ualberta.ca

References
LONGITUDINAL STABILITY OF PERSONALITY


Received February 14, 2001
Accepted December 9, 2002