Longitudinal Trajectories in Guilford-Zimmerman Temperament Survey Data: Results from the Baltimore Longitudinal Study of Aging

Antonio Terracciano, Ph.D., **Robert R. McCrae, Ph.D.**, and **Paul T. Costa Jr., Ph.D.**
Laboratory of Personality & Cognition, National Institute on Aging, NIH, DHHS, Baltimore, Maryland

Abstract

Developmental trends in personality traits over 42 years were examined using data from the Baltimore Longitudinal Study on Aging (N = 2,359, aged 17 to 98), collected from 1958 to 2002. Hierarchical Linear Modeling analyses revealed cumulative mean-level changes averaging about 0.5 *SD* across adulthood. Scales related to Extraversion showed distinct developmental patterns: General Activity declined from age 60 to 90; Restraint increased; Ascendance peaked around age 60; and Sociability declined slightly. Scales related to Neuroticism showed curvilinear declines up to age 70 and then increased. Scales related to Agreeableness and Openness changed little; Masculinity declined linearly. Significant individual variability in change was found. Although intercepts differed, trajectories were similar for men and women. Attrition and death had no effect on slopes. This study highlights the use of lower-order traits in providing a more nuanced picture of developmental change.

Keywords

HLM; aging; longitudinal study; Five-Factor Model; stability

Introduction

After decades of research, the broad outlines of adult personality development are well understood (Caspi, Roberts, & Shiner, 2005; Helson, Jones, & Kwan, 2002; Jones & Meredith, 1996; McCrae & Costa, 2003; Mroczek & Spiro, 2003; Roberts, Walton, & Viechtbauer, in press; Small, Hertzog, Hultsch, & Dixon, 2003; Steunenberg, Twisk, Beekman, Deeg, & Kerkhof, 2005). But because maturational changes are generally small in magnitude, a more detailed account requires data from large samples followed over long periods. The present study offers new information about personality development by focusing on specific lower-order traits and by adopting advanced statistical methods in the analysis of data from a large, long-term longitudinal study.

Factors and Facets of Personality

Personality psychology has made extraordinary advances in the past 20 years (McCrae, 2002) in large part because of the widespread recognition that most personality traits can be interpreted as aspects of five broad factors that constitute the Five-Factor Model (FFM; Digman, 1990). Literature reviews (e.g., Judge, Heller, & Mount, 2002) now routinely organize findings in terms of the five factors, and, in brief, personality development can be summarized

Address correspondence to: Antonio Terracciano, 5600 Nathan Shock Dr., Baltimore, Maryland, 21224. Email: terraccianoa@grc.nia.nih.gov.

by noting that Neuroticism (N), Extraversion (E), and Openness (O) decline, whereas Agreeableness (A) and Conscientiousness (C) increase throughout adulthood (McCrae & Costa, 2003).

Although categorizing traits in terms of the FFM can bring order to a large body of studies, it does so at a cost: Differences in the developmental trends of specific traits within the same factor are overlooked. Helson and Kwan (2000) pointed to a distinction within the E domain, showing that measures of Social Assurance increased toward middle age, whereas measures of Social Vitality declined. There may be other useful distinctions to be made, and one way to search for them is by analyzing data at the level of discrete facets. Many personality models are hierarchical, with broad factors defined by more specific facets (e.g., Caspi et al., 2005). The Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992b) measures six facets for each factor, and previous facet-level analyses have shown distinct developmental patterns. For example, cross-sectional analyses of self-reports suggest that E5: Excitement-Seeking declines precipitously from early adolescence to middle adulthood, whereas E1: Warmth shows a much slower decline (Costa & McCrae, 1992b), and this distinction is replicated in analyses of observer ratings from around the world (McCrae et al., 2005).

NEO-PI-R scales provide useful assessments of a wide range of specific personality traits, but scientific advances depend upon replications using a variety of methods, including different instruments. In this article we describe normative age trajectories (i.e., those that appear to characterize all people, on average) of personality traits assessed by the Guilford-Zimmerman Temperament Survey (GZTS; Guilford, Zimmerman, & Guilford, 1976), interpreted in terms of higher- and lower-order traits of the FFM.

Multilevel Modeling Studies of Mean-Level Changes in Personality Traits

Recent progress in statistical methods (see Hertzog & Nesselroade, 2003) has provided researchers with new tools to examine longitudinal trajectories. To assess intraindividual growth and describe normative development in personality traits, recent studies have increasingly relied on Structural Equation Modeling (Small et al., 2003), Latent Curve Analysis (Jones & Meredith, 1996), or Multilevel Modeling approaches (Mroczek & Spiro, 2003; Steunenberg et al., 2005), including longitudinal Hierarchical Linear Modeling (HLM; Helson et al., 2002; Terracciano et al., 2005).

Mroczek and Spiro (2003) followed a sample of 1,600 men from the Normative Aging Study for 12 years and found curvilinear slopes for N, which declined up to age 80, and an overall linear trajectory for E indicating no average change. However, E increased in the young cohort and declined in the older cohort. Furthermore, they found significant individual differences in intraindividual change. Steunenberg et al. (2005) found a similar concave curve for N on a random Dutch sample of 2,177 respondents aged 55–85 from the Longitudinal Aging Study Amsterdam.

Helson et al. (2002) conducted HLM analyses on data from two small samples followed for about 40 years. They found that Dominance and Independence (measures of Social Assurance) peaked in middle age, while measures of Social Vitality declined with age. Some measures of Norm-Adherence such as Self-Control increased with age. Somewhat similar results are reported by Jones, Livson, and Peskin (2003).

Terracciano et al. (2005) used HLM to analyze NEO-PI-R data collected between 1989 and 2004 in the Baltimore Longitudinal Study of Aging (BLSA). They found a decline up to age 80 in N, stability and then decline in E, a decline in O, increase in A, and increase up to age 70 in C. Terracciano et al. (2005) also described trends for the facets of each factor. Although

most facets followed the pattern of the factor they define, some facets did not. This was particularly evident for some facets of E, which showed clearly distinct patterns.

Using HLM, the present study attempts to replicate and extend the trends reported in the 15-year longitudinal study by Terracciano et al. (2005) using a different personality questionnaire, the GZTS, and assessment points that span a much longer time interval, up to 42 years. Based on Terracciano et al.'s (2005) results we can formulate specific hypotheses for each GZTS scale, provided that we can interpret the GZTS scales in terms of NEO-PI-R factors and facets. To do so, we take advantage of the fact that the BLSA participants have completed several other personality questionnaires, including the NEO-PI-R. In Table 1 we report correlations between the ten GZTS scales and the NEO-PI-R factor scores in a sample of 900 individuals who were administered both tests during the same visit. Table 1 reports the specific NEO-PI-R facet with which each GZTS scale is most highly correlated and the correlation, and the last column describes the age trajectory observed for that facet in Terracciano et al. (2005).

Although most of these correlations are moderate in magnitude and do not suggest full equivalence of the GZTS scales with any single NEO-PI-R facet, the trends observed can serve as a set of hypotheses for GZTS trajectories. Note that facet correlations for Emotional Stability, Objectivity, and Masculinity are negative, so our hypothesis is that these three scales will show a decelerated increase over the lifespan. Four GZTS scales are related to E, but they tap different facets, and are hypothesized to show different age trajectories: We predicted that General Activity will be stable in young adulthood and then show rapid declines in old age, Ascendance will peak in middle age, Sociability will show no change in mean level, and Restraint, which is related to C6: Deliberation, and inversely to E5: Excitement-Seeking ($r = -.42$) will increases with age. Friendliness was expected to increase; Thoughtfulness was hypothesized to show little change with age; and Personal Relations was expected to peak in midlife.

BLSA Studies of the GZTS

The GZTS was administered from the inception of the BLSA in 1958 until 2002, and several previous studies have reported longitudinal analyses of mean level changes in men (Costa, McCrae, & Arenberg, 1980; Costa, Metter, & McCrae, 1994; Douglas & Arenberg, 1978) and women (Costa & McCrae, 1992a, 1998). In several respects, the present study represents an advance. First, more data have been collected, increasing the number of assessment points and expanding the time interval, at least for a few respondents, to as long as 42 years. Since the last analyses of BLSA men's data were presented in 1994, 390 additional administrations of the GZTS have been gathered from men; since the last analyses of women's data were published in 1998, 237 additional administrations have been gathered from women. Second, HLM analyses offer more refined estimates of maturational trends, based on a more complete use of data. In fact, HLM analyses use all available data, despite variations in retest interval and number of administrations per individual. Previously published GZTS studies used relatively smaller subsamples due to missing data and variations in time interval, for both men (Costa et al., 1994, $N = 205$, retest interval = 20 to 30 years, $M = 24.4$ years) and women (Costa & McCrae, 1998, Sample A: $N = 114$, retest interval = 12 to 16 years, $M = 13.4$ years, Sample B: $N = 211$, retest interval = 6 to 10 years, $M = 7.8$ years). Finally, unlike previous reports on the GZTS from the BLSA, this study includes both men and women, making possible a direct test of gender differences in personality change, as suggested by other researchers (Viken, Rose, Kaprio, & Koskenvuo, 1994; Wink & Helson, 1993).

Terracciano et al. (2005) reported a longitudinal HLM study of changes on NEO-PI-R scales in the BLSA. The GZTS is a different instrument, but the samples used in that and the present study overlap. The two studies differ in the time interval covered: The GZTS was administered from 1958 to 2002, the NEO-PI-R from 1989 to 2004, and 72% of GZTS administrations were gathered before the introduction of the NEO-PI-R. The two studies also examined different

subsets of BLSA participants: 1,443 participants were included in both studies, but 501 had only NEO-PI-R data and 916 had only GZTS data.

Method

Participants

The sample consisted of 2,359 community-dwelling volunteers from the BLSA, an ongoing multidisciplinary study of aging. Across the span of the study, age at assessment ranged from 17 to 98 years ($M = 54.9$, $SD = 16.4$). BLSA participants are generally healthy and highly educated ($M = 16.2$ years of education, $SD = 2.8$); the present sample is 85% Caucasian, 11% African-American, and 4% other. There were 1,472 men and 887 women. Data were collected during regularly scheduled visits, for men starting in October, 1958, and for women in January, 1978, and continuing until May, 2002. The GZTS was administered to all participants at their first or second visit, and subsequently approximately every 6 and then 12 years. Participants had from 1 to 6 administrations (Table 2), and the average time interval between administrations was 7.9 years ($SD = 3.9$). The 2,359 participants provided a total of 4,739 assessments.

About 43.3% of the participants assessed in this study were still active in 2002, 37.5% were deceased, and the remaining 19.2% were dropouts. Dropouts consisted of 4.5% who have formally withdrawn (although they are willing to participate by phone, mail, and/or home visits), 0.7% lost to follow-up, 13.1% at least one year past their due date, and 0.9% who refused to be contacted again. The last four columns of Table 2 provide the proportion of deceased and dropouts by gender and the number of administrations. There is a higher rate of deceased men (women entered the study two decades after men), and the rate of dropouts was slightly higher for participants with only one administration. Indeed, in the active cohort, 60.4% of participants had two or more assessments; the deceased participants had a similar distribution of the number of measurement points (61.7% with two or more assessments); among dropouts only 46.3% had two or more assessments.

At their first GZTS administration, individuals who subsequently dropped out of the study were younger (43 vs. 52 years; $t(1,2357) = 11.0$; $p < .01$, $d = .59$), more likely to be women (45% vs. 36%, $\chi^2(1) = 12.1$, $p < .01$), and less educated (15.9 vs. 16.3 years; $t(1,2236) = 3.1$; $p < .01$, $d = .16$) than participants who did not drop out of the study. After controlling for age, year tested, sex, and education, individuals who dropped out of the study did not differ from those who did not on the ten GZTS scales at their first administration.

At their first GZTS administration, individuals who died in the course of the study were older (64 vs. 42; $t(1,2357) = -38.1$; $p < .01$, $d = 1.64$), more likely to be men (82% vs. 51%, $\chi^2(1) = 229.2$, $p < .01$), and tested earlier in time (1970 vs. 1984; $t(1,2357) = 29.4$; $p < .01$, $d = 1.26$) compared to participants who were still alive at end of this study (2002). After controlling for age, year tested, sex, and education, individuals deceased in the course of the study differed from those who had not died on only one of the ten GZTS scales at their first administration. Estimated marginal means indicated that participants who died were lower on Emotional Stability compared to those who were still alive in 2002 (47.9 vs. 49.9; $p < .05$, partial $\eta^2 = .004$).

Measure

The GZTS (Guilford et al., 1976) is a factor-based personality questionnaire consisting of 300 items, 30 for each of the 10 GZTS scales. For each item, participants choose between '*yes*,' '*no*,' and '?.' Any scale with more than three '?' responses was considered missing, a procedure suggested by Guilford and Zimmerman (1949). Therefore, small variations in the number of

participants will be seen in the analyses for different scales. Raw scores were standardized as *T*-scores (*M* = 50, *SD* = 10) using the mean and standard deviation of all 4,739 administrations to obtain the same metric for each scale and a clear measure of effect size.[1].

The GZTS scales are valid and reliable (Guilford et al., 1976). Internal consistency reliability coefficients range from .75 to .87 (*Med* = .80; Guilford et al., 1976). In the BLSA (McCrae, Costa, & Arenberg, 1980), the structural stability of the GZTS has been shown across age, cohort, and time-of-measurement. Test-retest reliability estimates for the ten scales ranged from .75 to .91 (*Mnd* = .83; Costa et al. 1980). Retest stability coefficients over a 24-year interval ranged from .61 to .71 (*Mdn* = .65; Costa & McCrae 1992a).

Data analysis: HLM

HLM (Raudenbush & Bryk, 2002) is a flexible approach that can be applied to evaluate within-individual change or growth trajectories. In HLM analyses the number and spacing of measurement observations may vary across persons, given that the time-series observations in each individual are used to estimate each individual's trajectory (level-1), and those individual parameters are the basis of group estimates (level-2). Even data from individuals who were tested only on a single occasion can be used to stabilize estimates of mean and variance. In this way, all available data can be included in the analyses. This is a major advantage of conducting analysis within the HLM framework; by contrast, missing data and varying timing pose major problems in conventional repeated measures ANOVA. Furthermore, longitudinal HLM can estimate age-trajectories over a broad age span (from 20 to 90) using data collected in a shorter time interval (up to 42 years).

The analyses were conducted using the program HLM version 6 (Raudenbush, Bryk, & Congdon, 2004). For each trait a stepwise procedure was adopted to evaluate longitudinal trajectories. First level-1 linear and then quadratic models were tested. The quadratic model was chosen when it provided a better fit, according to the chi-square test of deviance at $p < .01$. After the level-1 model was determined, sex (Male = 0, Female = 1) and cohort (year of birth) were entered in the model as level-2 variables. Level-2 variables were retained in the final model if they improved the fit of the model, or in other words, explained a significant amount of the variance in mean intercept or slope. Finally we examined possible effects of attrition and death on the intercept and slopes of the ten scales.

Age in decades was centered on the grand mean (*M* = 54.9 years) to minimize the correlation between the linear and quadratic terms. At level-2, year of birth (in decades) was centered on the sample mean year of birth (1929).

Results

Before describing changes in personality traits we evaluate the proportions of between- and within-subjects variance, and examine to what extent the within-subject variance is explained by age. Between- and within-individual variance estimates[2] from the basic one-way ANOVA with random effects model (Raudenbush & Bryk, 2002, p. 24) were used to estimate the proportion of the stable variance in personality traits, as the ratio of between-subjects variance (u_0: intercept variance) to the total variance (u_0: intercept variance + σ^2: within-subject variance). This ratio, which is an intraclass correlation, indicates that the proportion of variance

---

[1]Means and standard deviations for the first assessment for all participants and by age groups are available from the first author, as are tables showing means and standard deviations contrasting first assessments with second (*M* interval = 7.5 years), third (*M* interval = 14.9 years), and fourth (*M* interval = 22.6 years) assessments.
[2]Estimates of variance components from the basic one-way ANOVA with random effects model are not shown in the Table 3. Using the variance components from the final models reported in Table 3, the proportion of stable variance would have a median of .74 for the ten scales.

that was stable over the course of this study ranged from 67% for Thoughtfulness to 86% for Masculinity (*Mdn* = 71%).

Given that the between-subjects variance accounts for about 70% of the total variance, the remaining 30% is within-subject variance. By comparing the within-subject variance from the above baseline model with the residual within-subject variance from a model that includes age and age squared, it is possible to estimate the proportion of within-subject variance explained by age (Raudenbush & Bryk, 2002, p. 24). The within-subject variance accounted for by age ranged from 4% for Restraint to 26% for General Activity (*Mdn* = 14%). Thus, age accounts on average for about 14% × 30% = 4% of the total variance, a figure consistent with cross-sectional estimates (Terracciano et al., 2005).

HLM results for the final models of the ten GZTS scales are reported in Table 3. We first describe the average trajectories with the fixed effects of the intercept, linear, and quadratic terms respectively (γ_{00}, γ_{01}, γ_{02}), then the associated random effects (u_0, u_1, u_2), and finally the effects of predictors (i.e., gender, cohort, attrition, and death) of intercept and slope variability.

Fixed Effects

Of most interest are the linear and quadratic fixed effects, which determine the shape of the developmental trajectory. Estimated age trajectories for the ten scales are depicted in Figure 1, setting cohort as equal to the sample mean date of birth (1929), and separately for men and women for the scales that showed significant gender differences.

As hypothesized, the four GZTS scales related to E have distinct patterns: General Activity is stable or slightly increases in young adulthood and then declines at an accelerating rate in the very old; Restraint (which is negatively related to E, and positively related to C6: Deliberation) increases linearly; Ascendance has a curvilinear trajectory which peaks around age 60; and Sociability decreases linearly at a very modest rate. Emotional Stability and Objectivity, the two scales most strongly and inversely related to N, have a similar pattern: Both increase at a decelerating rate up to age 70, and then decrease slightly in old age. Personal Relations, which is related to A1: Trust, increases up to age 50 and then declines. Contrary to our hypotheses, Friendliness, the GZTS scale related to A, is stable, and Thoughtfulness, the GZTS scale related to O5: Ideas, follows a concave curve, with some increase in old age. Masculinity declines with age, as hypothesized, but the effect is linear. The changes in adulthood on the ten scales amount, at most, to about one *T*-score point (.1 *SD*) per decade.

Random Effects

The random effects terms (u_0: variance) associated with the intercepts, which reflect between-individual differences in personality traits, were all statistically significant and substantially higher than the within-individual variance (σ^2). General Activity, Ascendance, and Sociability showed less within-individual variability than other scales. In other words, on these scales, individuals are relatively more consistent across assessments. Masculinity shows markedly smaller values of both within- and between-individuals variance values (σ^2 and u_0), setting it apart from the other GZTS scales, perhaps because Masculinity is a marker of sex role attributes more than a personality trait.

None of the variances of the quadratic slopes in Table 3 was significant, but for seven scales, variances associated with the linear slopes were significant, indicating that for those scales, there are individual slopes departing from the overall trends. Next, we attempt to explain such variability with individual difference variables.

Level-2 Predictors of Intercept and Slope Variability

Gender was a significant predictor of the intercept for six scales. In the terminology of the NEO-PI-R, women were higher on A (Friendliness and low Masculinity), lower in Assertiveness (Ascendance) and higher on N (i.e., lower on Emotional Stability, Objectivity, Personal Relations, and Masculinity), which is consistent with cross-cultural patterns of sex-differences (Costa, Terracciano, & McCrae, 2001). Gender explains 55% of the variance in the intercept of Masculinity, 3.3% for Ascendance, and about 1% for the other scales.

Cohort (year of birth centered on 1929) was a significant predictor of the intercept for seven of the ten GZTS scales. Later-born cohorts have lower intercepts (about 1 T-score point per decade) on Restraint, Friendliness, and Personal Relations, and higher intercepts on General Activity, Ascendance, Emotional Stability, and Thoughtfulness. These effects are net of the estimated effects of age (the level-1 predictor), and explain 7.6% of the variance in the intercept of Personal Relations and 2.6% for Ascendance, but less than 1% for the other scales.

As indicated by the table note in Table 3, gender was a significant predictor of the slope variability for Thoughtfulness, with women declining in young adulthood more, and increasing in old age less, than men (see Figure 1). This gender by age effect explained less than 1% of the slope variance. Birth cohort was a significant predictor of the linear slope of Sociability and Masculinity, explaining 1% and 3.7% of the slopes' variance, respectively. In both cases, later-born cohorts decline less.

Attrition and Death

Although there is little evidence that attrition influences the results of longitudinal studies of personality traits (Roberts et al., in press), we examined whether the trajectories of individuals who dropped out of the study differed from the trajectories of those who remained in this study. We examined also whether individuals who died during the course of this study had different levels (intercept) or slopes compared to individuals still alive. We created a dummy variable, Attrition, that contrasted those who dropped out with those who remained in the study, and another dummy variable, Death, that contrasted those who died with those still alive. We entered the dummy variables as level 2 predictors of intercept, linear, and quadratic terms in the models presented in Table 3.

Overall, HLM analyses indicate that attrition and death had limited effects on the trajectories of the GZTS scales. The only significant attrition effect was found on Emotional Stability: Individuals who dropped out of the study had a 1.3 T-score points lower intercept ($p < .05$; see also Terracciano et al., 2005). The trajectories of participants who died differed significantly from those who were still alive on two of the ten GZTS scales. Those who died had a 2.4 T-score points lower intercept on Emotional Stability ($p < .001$). For Friendliness, those who later died showed a decline of less than 2.5 T-score points from age 20 to 90 ($p < .05$), whereas those still alive were substantially stable or slightly increased.

Supplementary Analyses

The data structure for this study was unusual because data were collected for the first 20 years of the study only from men. To examine whether this feature affected results, we attempted to replicate the results reported in Table 3 on the assessments collected from 1978 to 2002 only, when both men and women were tested. This analysis removes possible Gender × Time-of-Measurement interaction effects. These analyses produced essentially the same results as those reported in Table 3, with only 4 of the 54 significant coefficients reported in Table 3 becoming non-significant.

Next, we repeated the analyses in separate samples of men and women. These analyses cannot be compared to the full model, where gender was a level-2 variable, but the major concern is with the shape of the developmental curves. Of the 14 significant linear and quadratic fixed effects, all except the linear effects for General Activity in men and Sociability in women remained significant, and in the same direction, in the new analyses. In addition, Personal Relations showed a significant linear effect in men. Overall, however, these analyses suggest similar results for independent analyses of men and women.

Finally, we examined whether we could replicate the findings in Terracciano et al. (2005) using the subset of 916 individuals who had not completed the NEO-PI-R. In this subset we can test hypotheses about developmental curves that were derived from a different instrument administered to a non-overlapping sample (see Table 1). We used Overlaping vs. Non-Overlapping with the NEO-PI-R sample as a level-2 variable in analyses of the full GZTS sample. Those analyses showed that Personal Relations declined more steeply in the Non-Overlapping group, but there were no other significant difference between slopes. Thus, consistency of results for the NEO-PI-R and the GZTS is not due simply to overlapping samples.

Discussion

HLM analyses indicate that nine of the ten GZTS scales showed significant average changes in personality traits. The effect size can be evaluated against several standards. Over a single decade, the average changes on the GZTS scales are less than one tenth of one standard deviation, which is less than a "small" effect size (Cohen, 1988). For all scales except General Activity and Emotional Stability, cumulative effects over the entire adult lifespan are less than one-half standard deviation, which is a medium effect size (Cohen, 1988), but would not reach the threshold for clinically sig nificant change in an individual (Jacobson & Truax, 1991). Note that these effects are found over the age range from 20 to 90, which reaches from adolescence to advanced old age. Effects are still smaller between ages 30 and 70, the usual age range in discussions of adult personality (McCrae & Costa, 2003).

However, the changes observed followed expected patterns for most scales. We found distinct trajectories for the four GZTS scales related to E: Ascendance showed a peak in middle age, replicating Helson and colleagues (2002). Sociability declined linearly. In addition, we found that Activity was essentially stable up to age 50 and then declined at an accelerating rate in old age, a finding that is consistent with a biologically-based decrease in energy and tempo with age, and with previous reports from the BLSA (Douglas & Arenberg, 1978; Terracciano et al., 2005). Restraint, which is inversely related to E5: Excitement Seeking, and positively related to C6: Deliberation, increased linearly throughout adulthood, which is consistent with declines in risk-taking and increased impulse control with aging. Although these findings are consistent with the broad generalization that E declines with age, it is clear that at least for this factor, age trajectories are more appropriately described and researched at the facet level.

Emotional Stability, Objectivity, and Personal Relations increased in young adulthood and decline slightly in later adulthood. These findings mirror previous reports on N (Mroczek & Spiro, 2003; Robins, Fraley, Roberts, & Trzesniewski, 2001; Small et al., 2003; Terracciano et al., 2005; but see, Weiss et al., 2005), and suggest that although in young and middle adulthood changes are in the direction of greater maturity (e.g., Caspi et al., 2005), that gain does not necessarily continue in old age. However, despite the decline in activity, ascendance, and emotional stability, these small changes probably do not reflect meaningful declines of functionality in old age.

As in previous reports (Douglas & Arenberg, 1978), Masculinity declined with age for men and women, with men substantially higher than women on average throughout the entire age range. Thoughtfulness was expected to show little change (like O5: Ideas in Terracciano et al., 2005), but it increased slightly after age 50, a finding that contrasts with the overall decline of the O factor in previous studies.

One scale, Friendliness, did not change longitudinally. This finding is somewhat puzzling, because both cross-sectional (McCrae et al., 1999; Weiss et al., 2005) and longitudinal (Terracciano et al., 2005) studies have found increases with age in A, the factor that includes Friendliness. However, one large-scale six-year longitudinal study of men and women at midlife found no longitudinal change in A (Costa, Herbst, McCrae, & Siegler, 2000). Clearly, this is a question that requires further research.

Some large cross-sectional and longitudinal studies (Srivastava, John, Gosling, & Potter, 2003; Viken et al., 1994) have reported different developmental trends for men and women, especially for N. With the exception of Thoughtfulness, in the present study we found no evidence that gender influences personality development. The general absence of gender effects on developmental curves is consistent with other longitudinal (e.g., Helson et al., 2002; Roberts et al., in press) and cross-sectional studies in the US and around the world (McCrae & Costa, in press) using self-report and observer perspectives (McCrae et al., 2004).

HLM analyses in the present study found significant cohort effects on seven scales. The largest effect was found for Personal Relations, with later-born cohorts declining more than 1 *T*-score point per decade, or about .5 *SD* over four decades. Personal Relations is related to Trust, which Robinson and Jackson (2001) found to be declining among Americans born after the 1940s. Personal Relations is also weakly and inversely related to N, which shifted toward substantially higher levels (about 1 *SD*) during recent decades (from 1952 to 1993) according to one meta-analysis (Twenge, 2000). However, in the present study corresponding cohort effects are not found for Emotional Stability or Objectivity, which are both strongly related to N. Later-born cohorts are instead higher in Ascendance, a scale related to E. A study by Twenge (2001) on American college students assessed from 1966 to 1993 found similar increases for E, but with a larger effect size of slightly less than 1 *SD* over 20–25 years. In the present study, over the same time period an estimated effect of one-third of one *SD* was found for Ascendance. Other GZTS scales related to E did not show consistent cohort effects.

The present study has several positive features, including sample size, long duration, and analytical method used, but it also has several limitations. Although the GZTS is a valid and reliable instrument, it is an older personality inventory that does not comprehensively cover the full range of traits in the FFM. The GZTS scales tap the N and E dimensions well, but not O, A, or C. Another limitation is the demographics of the sample. BLSA participants are not representative of the U.S. population, because they are highly educated and have high socioeconomic status. However, similar trajectories for N and E were found in the Normative Aging Study (Mroczek & Spiro, 2003), which differs in education and social-economic status from the BLSA (see also Steunenberg et al., 2005). Furthermore, Small et al. (2003) found years of education to be unrelated to longitudinal change in FFM traits.

GZTS data indicate that over 70% of the variance in personality traits is between individuals (stable individual differences). The within-individual variance is in part accounted for by age (about 14%), but most might be explained by measurement error or biological or environmental factors, which should all be addressed by future research. The present study confirms previous reports showing different trajectories for different components of Extraversion, and suggests that the analysis of lower-order traits can provide a more nuanced picture of developmental

change. HLM analyses indicate that the trends in middle adulthood are in the direction of greater maturity, but changes among the oldest adults are not necessarily positive. Overall, the changes in adulthood seen in this exceptionally long-running longitudinal study are consistent with trends seen in other studies that used cross-sectional (Srivastava et al., 2003) and accelerated longitudinal designs (Steunenberg et al., 2005). Long-term longitudinal studies are crucial for predicting important outcomes and explaining non-normative trajectories, but it is good news for gerontologists that briefer studies give similar results in describing normative developmental trends.

References

Caspi A, Roberts BW, Shiner RL. Personality development: Stability and change. Annual Review of Psychology 2005;56:453–484.

Cohen, J. Statistical power analysis for the behavioral sciences. Vol. 2nd edition. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

Costa PT Jr, Herbst JH, McCrae RR, Siegler IC. Personality at midlife: Stability, intrinsic maturation, and response to life events. Assessment 2000;7:365–378. [PubMed: 11151962]

Costa PT Jr, McCrae RR. Multiple uses for longitudinal personality data. European Journal of Personality 1992a;6:85–102.

Costa, PT., Jr; McCrae, RR. Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment Resources; 1992b.

Costa, PT., Jr; McCrae, RR. Trait theories of personality. In: Barone, DF.; Hersen, M.; Hasselt, VBV., editors. Advanced personality. New York: Plenum; 1998. p. 103-121.

Costa PT Jr, McCrae RR, Arenberg D. Enduring dispositions in adult males. Journal of Personality and Social Psychology 1980;38:793–800.

Costa PT Jr, Metter EJ, McCrae RR. Personality stability and its contribution to successful aging. Journal of Geriatric Psychiatry 1994;27:41–59.

Costa PT Jr, Terracciano A, McCrae RR. Gender differences in personality traits across cultures: Robust and surprising findings. Journal of Personality and Social Psychology 2001;81:322–331. [PubMed: 11519935]

Digman JM. Personality Structure: Emergence of the Five-Factor Model. Annual Review of Psychology 1990;41:417–440.

Douglas K, Arenberg D. Age changes, cohort differences, and cultural change on the Guilford-Zimmerman Temperament Survey. Journal of Gerontology 1978;33:737–747. [PubMed: 299564]

Guilford, JP.; Zimmerman, WS. The Guilford-Zimmerman Survey: Manual of instructions and interpretations. Beverly Hills, CA: Sheridan Supply Co; 1949.

Guilford, JS.; Zimmerman, WS.; Guilford, JP. The Guilford-Zimmerman Temperament Survey Handbook: Twenty-five years of research and application. San Diego, CA: EdITS Publishers; 1976.

Helson R, Jones C, Kwan VSY. Personality change over 40 years of adulthood: Hierarchical linear modeling analyses of two longitudinal samples. Journal of Personality and Social Psychology 2002;83:752–766. [PubMed: 12219867]

Helson, R.; Kwan, VSY. Personality development in adulthood: The broad picture and processes in one longitudinal sample. In: Hampson, S., editor. Advances in personality psychology. Vol. Vol. 1. London: Routledge; 2000. p. 77-106.

Hertzog C, Nesselroade JR. Assessing psychological change in adulthood: An overview of methodological issues. Psychology and Aging 2003;18:639–657. [PubMed: 14692854]

Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology 1991;59:12–19. [PubMed: 2002127]

Jones CJ, Livson N, Peskin H. Longitudinal hierarchical linear modeling analyses of California psychological inventory data from age 33 to 75: An examination of stability and change in adult personality. Journal of Personality Assessment 2003;80:294–308. [PubMed: 12763702]

Jones CJ, Meredith W. Patterns of personality change across the life span. Psychology and Aging 1996;11:57–65. [PubMed: 8726370]

Judge TA, Heller D, Mount MK. Five-Factor Model of personality and job satisfaction: A meta-analysis. Journal of Applied Psychology 2002;87:530–541. [PubMed: 12090610]

McCrae RR. The maturation of personality psychology: Adult personality development and psychological well-being. Journal of Research in Personality 2002;36:307–317.

McCrae RR, et al. Universal features of personality traits from the observer's perspective: Data from 50 cultures. Journal of Personality and Social Psychology 2005;88:547–561. [PubMed: 15740445]

McCrae, RR.; Costa, PT, Jr. Personality in adulthood: A Five-Factor Theory perspective. Vol. 2nd ed. New York: Guilford Press; 2003.

McCrae, RR.; Costa, PT, Jr. Cross-cultural perspectives on adult personality trait development. In: Mroczek, D.; Little, T., editors. Handbook of personality development. Hillsdale, NJ: Erlbaum; in press

McCrae RR, Costa PT Jr, Arenberg D. Constancy of adult personality structure in adult males: Longitudinal, cross-sectional and times of measurement analyses. Journal of Gerontology 1980;35:877–883. [PubMed: 6969272]

McCrae RR, Costa PT Jr, de Lima MP, Simões A, Ostendorf F, Angleitner A, et al. Age differences in personality across the adult life span: Parallels in five cultures. Developmental Psychology 1999;35:466–477. [PubMed: 10082017]

McCrae RR, Costa PT Jr, Hrebíčková M, Urbánek T, Martin TA, Oryol VE, et al. Age differences in personality traits across cultures: Self-report and observer perspectives. European Journal of Personality 2004;18:143–157.

Mroczek DK, Spiro A. Modeling intraindividual change in personality traits: Findings from the normative aging study. Journals of Gerontology: Psychological Sciences 2003;58B:P153–P165.

Raudenbush, SW.; Bryk, AS. Hierarchical Linear Models: Applications and data analysis methods. Vol. 2nd ed.. Thousand Oaks, California: Sage Publications; 2002.

Raudenbush, SW.; Bryk, AS.; Congdon, R. HLM (Version 6). Lincolwood, IL: Scientific Software International; 2004.

Roberts RE, Walton KE, Viechtbauer W. Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. Psychological Bulletin. in press

Robins RW, Fraley RC, Roberts BW, Trzesniewski KH. A longitudinal study of personality change in young adulthood. Journal of Personality 2001;69:617–640. [PubMed: 11497032]

Robinson RV, Jackson EF. Is trust in others declining in America? An age-period-cohort analysis. Social Science Research 2001;30:117–145.

Small BJ, Hertzog C, Hultsch DF, Dixon RL. Stability and change in adult personality over 6 years: Findings from the Victoria Longitudinal Study. Journal of Gerontology: Psychological Sciences 2003;58B:P166–P176.

Srivastava S, John OP, Gosling SD, Potter J. Development of personality in early and middle age: Set like plaster or persistent change? Journal of Personality and Social Psychology 2003;84:1041–1053. [PubMed: 12757147]

Steunenberg B, Twisk JWR, Beekman ATF, Deeg DJH, Kerkhof A. Stability and change of neuroticism in aging. Journals of Gerontology Series B-Psychological Sciences and Social Sciences 2005;60:P27–P33.

Terracciano A, McCrae RR, Brant LJ, Costa PT Jr. Hierarchical linear modeling analyses of NEO-PI-R scales in the Baltimore Longitudinal Study of Aging. Psychology and Aging. 2005

Twenge JM. The Age of Anxiety? Birth cohort change in anxiety and neuroticism, 1952–1993. Journal of Personality and Social Psychology 2000;79:1007–1021. [PubMed: 11138751]

Twenge JM. Birth cohort changes in extraversion: A cross-temporal meta-analysis, 1966–1993. Personality and Individual Differences 2001;30:735–748.

Viken RJ, Rose RJ, Kaprio J, Koskenvuo M. A developmental genetic-analysis of adult personality - Extroversion and neuroticism from 18 to 59 years of age. Journal of Personality and Social Psychology 1994;66:722–730. [PubMed: 8189349]

Weiss A, Costa PT Jr, Karuza J, Duberstein PR, Friedman B, McCrae RR. Cross-sectional age differences in personality among medicare recipients aged 65 to 100. Psychology and Aging 2005;20:182–185. [PubMed: 15769223]

Wink P, Helson R. Personality change in women and their partners. Journal of Personality and Social Psychology 1993;65:597–605. [PubMed: 8410654]
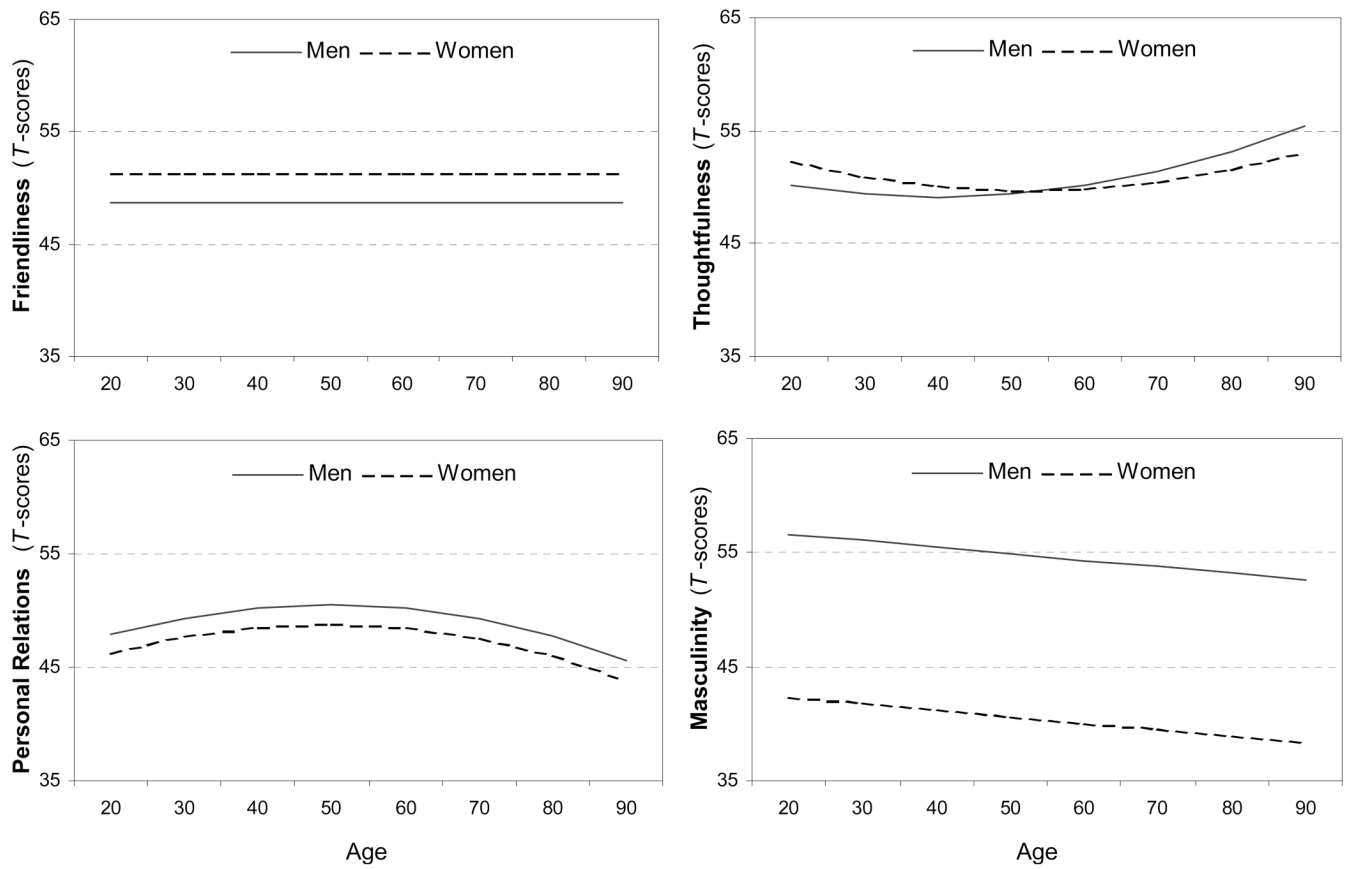
Acknowledgments
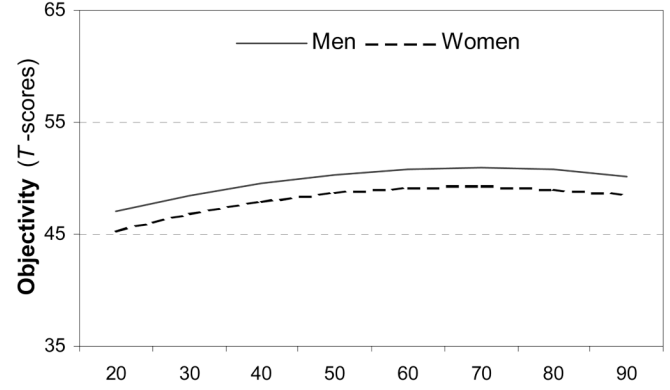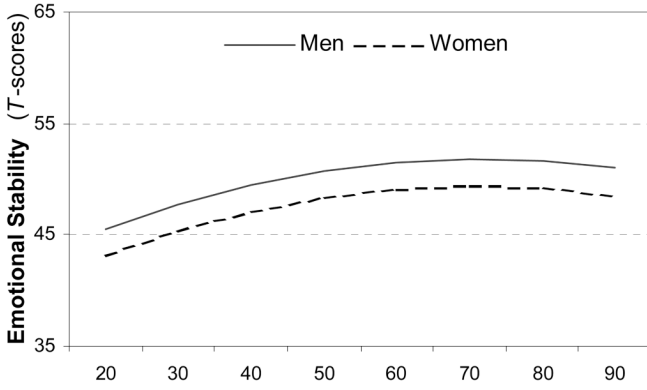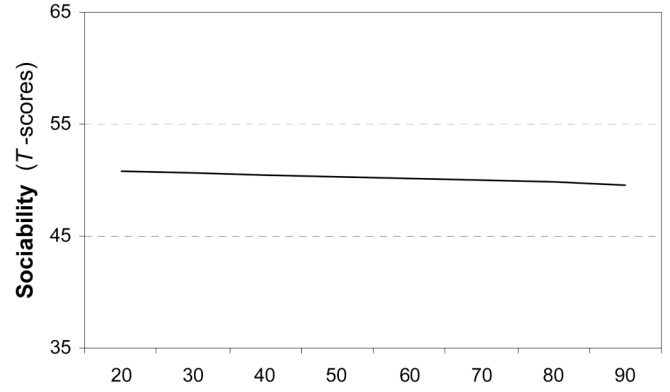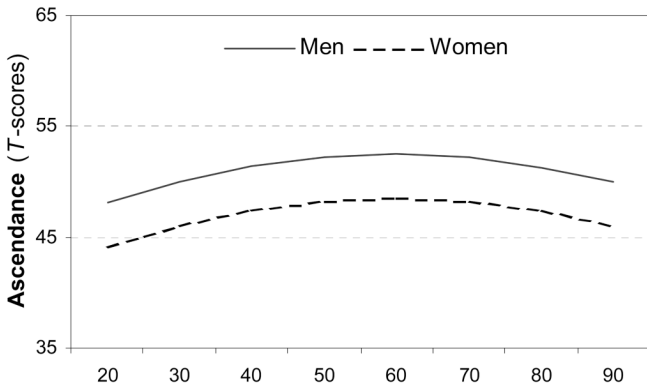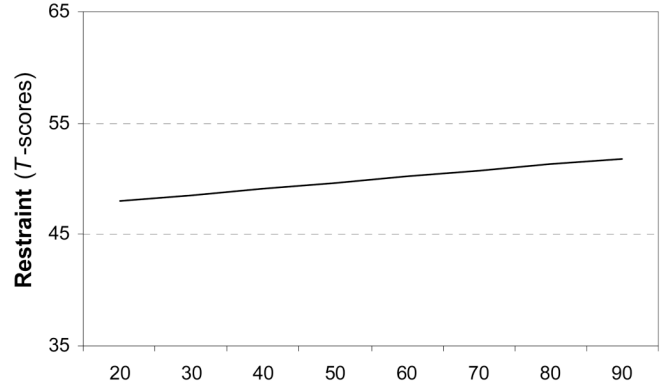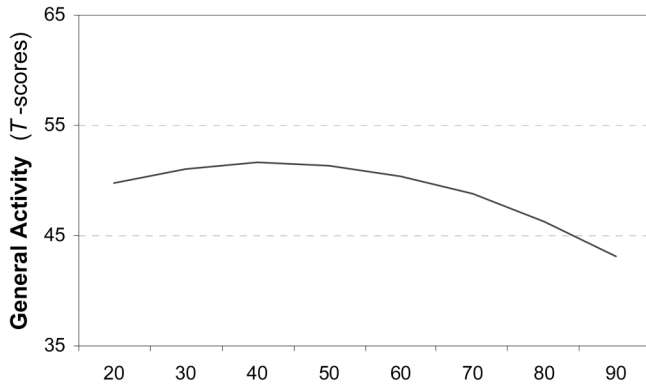
Figure 1.
Trajectories of Guilford-Zimmerman Temperament Survey scales (*T* scores: *M* = 50, *SD* = 10).

Table 1

Correlations between GZTS Scales and NEO-PI-R Factors, Most Strongly Related Facet, and the Trend of the NEO-PI-R Facet

GZTS scale	NEO-PI-R Factor					NEO-PI-R Facet	r	Age trajectories[a]
	N	E	O	A	C			
General Activity	−.11	**.51**	.08	−.28	.37	E4: Activity	**.75**	Accelerated decline
Restraint	.11	−**.48**	−.01	.15	.39	C6: Deliberation	**.46**	Increase
Ascendance	−.39	**.47**	.26	−.33	.22	E3: Assertiveness	**.69**	Convex curve, peak at 60
Sociability	−.35	**.69**	.19	.01	.03	E2: Gregariousness	**.70**	Near flat
Emotional Stability	−**.74**	.11	−.08	.00	.21	N3: Depression	−**.75**	Decelerated decline
Objectivity	−**.63**	.07	.05	.07	.00	N4: Self-Consciousness	−**.53**	Decelerated decline
Friendliness	−.21	−.15	−.01	**.47**	−.22	A4: Compliance	**.51**	Increase
Thoughtfulness	.16	−.08	**.46**	.05	.15	O5: Ideas	**.47**	Near flat
Personal Relations	−.31	.07	.04	.14	−.03	A1: Trust	**.46**	Convex curve, peak at 60
Masculinity	−**.41**	−.11	−.07	−.33	−.06	N1: Anxiety	−.33	Decelerated decline

Note. $N = 757$ to 854. Correlations over ± .40 are given in bold. GZTS = Guilford-Zimmerman Temperament Survey; NEO-PI-R = Revised NEO Personality Inventory; N = Neuroticism; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness.

[a] HLM-estimated trajectory of the NEO-PI-R facet (Terracciano et al., 2005). Because Emotional Stability, Objectivity, and Masculinity are inversely related to facets of N, our hypothesis is that these three scales will show a decelerated increase over the lifespan.

Table 2

Frequency of Guilford-Zimmerman Temperament Survey Administrations and Rate of Deceased and Dropouts for Men and Women.

Administrations	N		Mean Age		Age Range		Time interval		Deceased		Dropout	
	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
1	536	451	53	49	18–98	19–95	—	—	49%	18%	21%	30%
2	391	280	58	55	17–93	18–90	9 (1–42)	7 (1–22)	52%	20%	17%	16%
3	317	147	55	56	21–93	22–87	17 (6–41)	13 (7–24)	47%	16%	13%	16%
4	163	9	56	62	17–90	39–84	24 (8–39)	17 (14–23)	54%	—	16%	33%
5	60	—	56	—	20–89	—	30 (14–41)	—	42%	—	10%	—
6	5	—	53	—	25–81	—	41 (40–41)	—	—	—	—	—
Total	1472	887	55	52	17–98	18–95	16 (1–42)	10 (1–24)	49%	17%	18%	23%

Note. *N* refers to the number of individuals having exactly the specified number of administrations. Mean age is averaged across all administrations. Age range is from youngest age at first administration to oldest age at last administration. Time interval is the mean and range of years between first and last administrations.

Table 3

HLM Coefficients and Variance Estimates of Intercept, Linear, and Quadratic terms for GZTS scales

GZTS scales	σ^2: Residual Within-Subject Variance	Intercept				Linear		Quadratic	
		γ_{00}: M	γ_{01}: Gender	γ_{02}: Cohort	u_0: Variance	γ_{10}: M	u_1: Variance	γ_{20}: M	u_2: Variance
General Activity	18.13	51.0 (.24)**		0.05 (.01)**	79.27**	−0.95 (.12)**	4.32*	−0.38 (.04)**	0.23
Restraint	29.06	49.9 (.19)**		−0.07 (.01)**	60.74**	0.55 (.12)**	1.25**		
Ascendance	16.17	52.4 (.64)**	−3.90 (.43)**	0.13 (.01)**	78.98**	0.27 (.11)*	2.12	−0.28 (.04)**	0.29
Sociability	19.32	50.2 (.24)**			73.94**	−0.18 (.09)* c	2.48**		
Emotional Stability	24.24	51.2 (.65)**	−2.40 (.44)**	0.04 (.01)**	74.85**	0.77 (.13)**	4.69**	−0.24 (.05)**	0.34
Objectivity	26.18	50.6 (.61)**	−1.62 (.41)**		73.76**	0.46 (.10)**	2.43	−0.16 (.05)**	0.44
Friendliness	28.24	48.7 (.62)**	2.57 (.43)**	−0.11 (.01)**	60.06**	−0.01 (.13)	2.24**		
Thoughtfulness	28.24	49.7 (.24)**		0.05 (.01)**	75.55**	0.75 (.13)** g	3.15	0.25 (.05)**	0.20
Personal Relations	24.55	50.4 (.65)**	−1.67 (.44)**	−0.13 (.01)**	73.06**	−0.33 (.13)	3.31*	−0.30 (.05)**	0.29
Masculinity	14.25	54.7 (.45)**	−14.25 (.31)**		39.26**	−0.57 (.07)** c	0.39**		

Note. N = 2,166 to 2,308. Standard errors are shown in parenthesis. Significance of coefficients is tested by t ratio (coefficients/Standard error). Significance of variance terms is tested with chi-square. Quadratic effects are presented only when the quadratic model fit the data better than the linear model. Negative quadratic terms indicate convex trajectories; positive quadratic terms indicate concave trajectories. Coefficients are applied to (Age − 54.9) in decades. Gender: 0 = Male, 1 = Female. Cohort is year of birth centered on the mean (1929).

c Cohort effect on linear change γ_{12} = .013 (.00)**.

g Gender effect on linear change γ_{11} = −.65 (.20)**.

*
  $p < .05$;

**
  $p < .01$.