Personality Plasticity After Age 30

Antonio Terracciano, **Paul T. Costa Jr.**, and **Robert R. McCrae**
Laboratory of Personality and Cognition, National Institute on Aging, Baltimore, MD.

Abstract

Rank-order consistency of personality traits increases from childhood to age 30. After that, different summaries of the literature predict a plateau at age 30, or at age 50, or a curvilinear peak in consistency at age 50. These predictions were evaluated at group and individual levels using longitudinal data from the Guilford-Zimmerman Temperament Survey and the Revised NEO Personality Inventory over periods of up to 42 years. Consistency declined toward a non-zero asymptote with increasing time-interval. Although some scales showed increasing stability after age 30, the rank-order consistencies of the major dimensions and most facets of the Five-Factor Model were unrelated to age. Ipsative stability, assessed with the California Adult Q-Set, was also unrelated to age. These data strengthen claims of predominant personality stability after age 30.

Keywords

Five-Factor Model; personality development; long-term stability; individual differences; life-span; older adults

Since data from longitudinal studies appeared in the 1970s (e.g., Block, 1977), it has been clear that individual differences in personality traits are stable over long periods of time. Helson and Wink (1992) reported typical results in a sample of 101 women initially aged 43 and retested after 9 years: The median retest correlation for scales from the California Psychological Inventory (CPI; Gough, 1987) was .73. There is also considerable evidence that personality traits are more stable in adults than in adolescents. For example, Finn (1986) showed that the median 30-year retest correlation for factors from the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943) was .35 for respondents initially aged 17 to 25, and .56 for respondents initially aged 43-53.

However, there is disagreement about the degree of rank-order consistency in different portions of adulthood. Based on their own research and review of the literature, McCrae and Costa (1990) argued that "personality change is the exception rather than the rule after age 30; somewhere in the decade between 20 and 30, individuals attain a configuration of traits that will characterize them for years to come" (p. 10). Ten years later, two meta-analyses were published on the rank-order consistency (or differential stability) of personality traits. Roberts and DelVecchio (2000) reported that rank-order consistency increased with age, even in adulthood. They estimated that 30- and 40-year-olds would show 7-year retest correlations near .60, whereas individuals over age 50 would show retest correlations over .70, and they concluded that recent generations have "stretched the time it takes to fully develop one's traits" (p. 18) past age 30. Curiously, another meta-analysis published the

same year (Ardelt, 2000) concluded that rank-order consistency increased up to age 50, but decreased thereafter.

Meta-analyses combine data from different instruments, samples, and historical times, and may be subject to confounds (e.g., studies of 30-year-olds may have used less reliable instruments than studies of 60-year-olds). An alternative approach would examine rank-order stability in different age groups using the same instrument administered to comparable samples over the same time interval and in the same historical period. Such a design was used by Costa, McCrae, and Arenberg (1980), who examined 6- and 12-year retest coefficients for Guilford-Zimmerman Temperament Survey (GZTS; Guilford, Zimmerman, & Guilford, 1976) scales in male participants in the Baltimore Longitudinal Study of Aging (BLSA; Shock et al., 1984). The mean uncorrected stability coefficients for initially young (17-44), middle-aged (45-59) and old (60-85) men were .76, .77, and .75, respectively, over six years, and .72 .75, and .73, respectively, over 12 years. Similar results were reported by Costa and McCrae (1988) in a six-year study of men and women assessed with the NEO Personality Inventory (Costa & McCrae, 1985), and by McCrae (2001), who analyzed 6-year retest data in spouse ratings and 7-year retest data in peer ratings of personality. All of these studies suggest that personality is quite stable among young and old adults as well as those in midlife.

The present article updates these studies of BLSA participants, using longer retest intervals on the GZTS and new data from the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992b). In addition, longitudinal self-sorts on the California Adult Q-sort (CAQ; Bem & Funder, 1978; Block, 1961) are examined to determine the effect of initial age on the ipsative stability of personality.

The BLSA has both strengths and limitations as a sample in which to test these hypotheses. It is a large sample that has followed some participants for more than 40 years, and the well-educated and cooperative respondents provide data of high quality. However, they are clearly not representative of the population as a whole. In particular, the sample is biased toward older age, making it less than optimal for examining stability in the earliest decade of adulthood. There is, however, general agreement that personality is less stable before 30 than after; the present study therefore examines several forms of stability after age 30, where competing predictions have been made. Group-level analyses (test-retest correlations) facilitate comparisons with previous studies and can address the hypothesis that the degree of stability is related to age. However, we also examine stability at the individual level, computing a measure of stability for each individual that can be used in multilevel analysis to define in greater detail the effect of age on stability, while controlling for other variables.

Fraley and Roberts (2005) have recently emphasized the importance of a second issue in studying stability or change in rank-order consistency across the lifespan: the shape of the function relating consistency to retest interval. Retest correlations are typically lower over longer retest intervals—for example, the median retest correlation for GZTS scales after 6 years was .77; after 24 years it was .65 (Costa & McCrae, 1992c)—and researchers have typically assumed an exponential decay of consistency (e.g., Conley, 1984). Using that model, Costa and McCrae (1992c) projected that the median stability of GZTS true scores after 50 years would be .60, and concluded that "about three-fifths of the variance in personality traits is stable across the full adult age range" (p. 182).

Fraley and Roberts (2005) pointed out that the exponential decay model predicts that eventually rank-order consistency will decline to zero unless there are *developmental constancy factors* (such as one's unchanging genetic endowment) that force consistency to a non-zero asymptote. Such a model is testable only with data from multiple administrations

with widely varying retest intervals. GZTS data in the BLSA include retest intervals as long as 42 years and offer a rare opportunity to test the long-term exponential decay of consistency.

Analysis 1: Rank-Order Consistency in Different Age Groups

Most previous analyses of differential stability have examined retest correlations within different age groups. Roberts and DelVecchio (2000) suggested that these correlations are higher among adults over age 50 than among younger adults, including those in their 30s and 40s. Ardelt (2000) reported a curvilinear relation to age, with stability coefficients peaking around age 50. To test those hypotheses, we divided the BLSA sample by initial age into groups aged 30-50, 50-65, and 65+. McCrae and Costa (1990) would expect no differences in rank order consistency between these groups; Roberts and DelVecchio (2000) would hypothesize higher consistency in the two older groups, and Ardelt (2000) would predict higher consistency in the middle group than in the oldest group. We analyze data from the GZTS and the NEO-PI-R, and, to maximize time interval, we also examine data from the NEO Inventory, a precursor of the NEO-PI-R.

Previous analyses (Costa & McCrae, 1988) showed no systematic differences between men and women of different age groups on retest stability in NEO-PI scales, but no such comparisons have been made for the GZTS. We therefore report GZTS data separately for men and women in this study.

Too few individuals ($n = 19$) retested on the NEO-PI-R were initially under age 30 to study that group, but larger numbers of under-30 adults were retested with the GZTS, and we report supplementary analyses on that instrument to test the common view that stability is lower among adults under 30.

Method

Participants and procedure—The BLSA is a multidisciplinary study of normal aging. Participants have agreed to return for repeated assessments of biomedical and psychosocial variables. Recruitment has been continuous over the course of the study, so long-term participants have been assessed more frequently than more recently recruited participants. GZTS data were collected during regularly scheduled visits, starting for men 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. Retest stability over 6 and 12 years has previously been reported for GZTS scales in the BLSA sample (Costa et al., 1980; Costa & McCrae, 1992a, 1998). The present study examined data for individuals whose first and last administrations were separated by at least 6 and as much as 42 years.

The NEO-PI-R was administered by computer between September, 1989, and July, 2004, during regularly scheduled visits (Terracciano, McCrae, Brant, & Costa, 2005). To obtain long-term stability coefficients, retest correlations were computed between the first and last administration for individuals with a time interval between administrations of at least 6 years. Note that there is no overlap between these data and those used in Costa and McCrae's (1988) study of the NEO-PI; further, most of the individuals who took the NEO-PI-R (about 60%) joined the BLSA after the data collection for the earlier study.

Finally, to maximize the time interval, we also compared domain scores from individuals who had completed the NEO Inventory by mail in 1980 (McCrae, 1982) with domain scores from their latest administration of the NEO-PI-R. A description of the subsamples that completed each instrument is given in Table 1.

Measures—The GZTS (Guilford et al., 1976) is a factor-based personality questionnaire composed 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 colleagues. Therefore, there are small variations in the number of participants for different scales. Data were standardized as *T*-scores using the grand mean and standard deviations across all administrations. In the BLSA (McCrae, Costa, & Arenberg, 1980), the structural stability of the GZTS has been shown across age, cohort, and time-of-measurement.

The Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992b), designed to measure the Five-Factor Model (FFM) of personality, consists of 240 items answered on a five-point Likert format ranging from *strongly disagree* to s*trongly agree.* The NEO-PI-R assesses 30 facets, six for each dimension of the FFM: Neuroticism (N), Extraversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C). *T*-scores for facets are standardized scores ($M = 50$, $SD = 10$) based on combined-sex adult norms in the *Manual*; factor scores combining information from each of the 30 facets are also expressed as *T*-scores. Evidence on reliability and validity is presented in the manual (Costa & McCrae, 1992b). The NEO Inventory was a precursor of the NEO-PI-R that assessed only N, E, and O domains. Revisions introduced with the NEO-PI-R changed 10 of the 144 N, E, and O items, but these minor changes are unlikely to distort the long-term stability of the scales and should have no effect on the comparison of age differences in consistency.

Baseline comparison—We compared individuals included in the rank-order correlation analyses (Table 2 and 3) with those who had taken the GZTS or NEO-PI-R at least once, but were not included, because they had dropped out of the study, had a retest interval of less then 6 years, were younger than 30 at their first visit, or had not been in the study long enough to be retested. For the GZTS, those excluded were about 2 years younger, somewhat less educated, and slightly more likely to be female. After controlling for those differences, the excluded respondents scored significantly lower on Emotional Stability, Friendliness, Thoughtfulness, and Personal Relations, but none of the differences exceeded about 0.4 standard deviations in magnitude. Similarly, those excluded from the NEO-PI-R analyses were younger, less educated, and more likely to be female; after controlling for those differences, respondents who were excluded were about 2 *T*-score points higher on O. In general, it appears that respondents included in the analyses were similar to BLSA participants in general.

Results and Discussion

Retest correlations from first and last administrations in the total sample and in various age groups are reported in Table 2 for the GZTS and in Table 3 for the NEO-PI-R. Retest intervals for each group are given at the bottom of each column. All correlations are significant, with GZTS scales and NEO-PI-R facets near .70 and NEO-PI-R factors near .80, consistent with previous literature. The coefficients in Table 2 and 3 are underestimates of the true score stability, given that personality scales are not perfectly reliable. To correct for attenuation of retest coefficients due to measurement error, estimated stability coefficients can be obtained by dividing the observed coefficients by the short-term retest reliability. Short-term (e.g., two-week) retest reliability coefficients have not been reported for the full NEO-PI-R, but using the two-year retest correlations reported by McCrae, Yik, Trapnell, Bond, and Paulhus (1998), corrected stability coefficients for the five factors of the NEO-PI-R were .94, .91, .96, .92, and .92 for N, E, O, A, and C, respectively. Over a two year interval, some real change may have occurred, so the McCrae et al. (1998) values underestimate reliability and lead to an overcorrection: These corrected coefficients might be interpreted as upper-bound estimates of the true stability of personality traits.

Heise (1969) proposed an alternative method to estimate true score stability coefficients that can be used when data from three administrations are available. We applied it in a subsample of 520 individuals with at least three administrations and a minimum of six years between first and last administration. In this subsample the mean interval between the first and the intermediate administration was 5.3 years ($SD = 1.9$); between the first and the last it was 10.1 years ($SD = 2.3$). Heise's formula calculates retest stability as $s_{13} = r_{13}^2 / (r_{12} * r_{23})$, which yielded coefficients of .91, .89, .95, .90, and .95 for N, E, O, A, and C factors, respectively.

Of primary interest are the comparisons of the last three columns of Tables 2 and 3. Consistent with Roberts and DelVecchio's (2000) report, adults aged 30-50 showed significantly lower rank-order consistency (using Fisher's z-test) than adults over age 50 on GZTS Sociability, Emotional Stability, and Thoughtfulness in men, and on General Activity, Ascendance, Emotional Stability, and Masculinity in women, with small effect sizes ranging from q = .10 to .17 (q = | Fisher's $z_1 - z_2$|; Cohen, 1988). The hypothesis from Ardelt (2000) that stability declines after age 50 was supported for Thoughtfulness in men (q = .13), but women initially over age 65 were more rather than less consistent than women 50-65 on Friendliness (q = .20). No significant differences were found for GZTS Restraint, Objectivity, or Personal Relations. These data thus provide mixed support for the hypothesis of continued increase in rank-order consistency between 30 and 50, but no consistent support for the hypothesis that stability declines after age 50. Further, the mean retest interval is substantially longer for the male group aged 30 to 50 compared to the other groups (see Table 2), which has the effect of reducing the retest correlation coefficients and might explain some of the above effects. Analyses at the individual level will address this issue.

Among the five NEO-PI-R factors, significant differences among groups can be seen only for Extraversion, with adults aged 50-65 showing higher consistency than adults over age 65 (q = .06). Among the 30 NEO-PI-R facets, the over-65 group showed lower rank-order stability on E2: Gregariousness (q = .13) and O3: Feelings (q = .11), but higher on O6: Values (q = .16). For the remaining 27 facets, no significant effects were found, providing no support for the hypotheses that rank-order stability of individual differences varies systematically with age after age 30.

In 1980, the NEO Inventory was completed by 257 respondents aged 30 and older who subsequently took the NEO-PI-R an average of 19.0 years later (range = 9-24 years). Rank-order consistency for the N, E, and O domains were .73, .74, and .77, respectively for the full group. Not surprisingly, too few of these individuals were initially over age 65 to allow meaningful analyses. The 112 respondents initially aged 30-50 had retest correlations of .76, .74, and .81 for N, E, and O, respectively, whereas the 145 respondents initially over age 50 had retest correlations of .71, .74, and .74. These data provide no support for the view that long-term stability is higher among those over 50.

We conducted a supplementary analysis in which we examined rank-order consistency for GZTS scales in respondents initially less than 30 years old ($ns = 101-117$ men, 39-46 women) and retested after an average of 17.9 (men) or 10.2 (women) years. As expected, they tended to show lower retest correlations, ranging from .38 to .73 with median values of .58 in men and .64 in women.

Analysis 2: Individual Stability as a Function of Retest Interval

One limitation of the retest correlation analyses is that the retest interval differed for different age groups, and consistency generally declines with longer retest intervals. It would be preferable to control statistically for the retest interval, but the optimal form of that

control depends on the shape of the function relating rank-order consistency to retest interval. Is decline linear, or exponential, or does it approach a non-zero asymptote?

Fraley and Roberts (2005) developed a model based on retest correlations in groups, a strategy that was feasible for them because their meta-analysis provided coefficients from many samples. In our single sample we can address these issues by conducting analyses at the individual level, using Asendorpf's (1992) individual stability coefficients, defined for each individual as

$$1 - \left[(z_1 - z_2)^2 / 2 \right],$$

where z_1 and z_2 are scores for a trait standardized across the full sample at the first and second administrations. Note that the use of z-scores eliminates the effects of mean level differences. The mean of Asendorpf's coefficient across all respondents is equal to the retest correlation, so each coefficient represents the individual's contribution to overall rank-order consistency.[1]

Hierarchical regression analysis with Asendorpf's individual stability coefficients as dependent variable was performed on the combined male and female GZTS sample ($n =$ 1,063), controlling for initial age at step one (which had minor effects), and then introducing time interval and its square at steps two and three. As in Table 2, analyses were limited to respondents initially over age 30 with a minimum retest interval of six years. All ten GZTS scales showed some decline in individual stability with increasing retest interval, but it was significant only for General Activity, Ascendance, Sociability, and Personal Relations, which all declined at a decelerating rate. For these four scales, linear and quadratic terms explain between .6% and 2.3% of the variance in individual stability beyond that accounted for by initial age.

Because all GZTS scales showed a similar form of decay, we created an average individual stability score (Cronbach's alpha = .58) across the ten GZTS scales. As shown in Figure 1, this mean was best predicted by a concave curve; the linear and quadratic terms accounted for 2.7% of the variance. It is notable, however, that the curve does not resemble exponential decay; instead of tending toward zero consistency, it turns up after a retest interval of about 20 years. This should not be interpreted as evidence that stability increases with longer intervals; it is instead simply a quadratic approximation to a plateau. In the language of Fraley and Roberts (2005), developmental constancy factors seem to operate such that retest stability of GZTS scales does not decline further after retest intervals exceed 20 years; there appears to be a non-zero asymptote.

Exponential decay to a non-zero asymptote can be modeled with the Nonlinear Estimation program of Statistica (SoftStat, 1995) using the equation

$$\text{Individual Stability} = c + \left(b_0 * e^{b_1 * t} \right),$$

where c is the asymptote, b_0 and b_1 are exponential decay coefficients, and t is the time interval in years. It is clear from Figure 1 that c must be between .6 and .7, and we varied it systematically across this range. The optimal value was .655, leading to the equation

---

[1]Fifteen respondents had mean individual stability coefficients less than 0.0, suggesting major and pervasive changes in personality trait levels. The largest of these, marked by dramatic decreases in scales measuring emotional stability and Extraversion, was coincident with an episode of depression at the time of retest (cf. Costa, Bagby, Herbst, & McCrae, 2005).

$$Individual\ Stability = .655 + \left(.413\ ^* e^{-.187^* t}\right),$$

which is plotted in Figure 1. This model accounted for 3.0% of the variance in mean individual stability scores, somewhat more than the quadratic model. Because the mean of the individual stability coefficients is equal to the retest correlation, the asymptote can also be regarded as an estimate of the long-term stability for GZTS scales: The lower bound for observed coefficients is about .65, and (using reliability estimates from Costa et al, 1980) the lower bound for true scores is close to .80.

Analysis 3: Individual Consistency Scores as Function of Age

The retest correlations in Table 2 and 3 are easily understood, but they are not optimal tests of the effects of age on rank-order consistency for two reasons. First, they utilize only a portion of the available data, namely, the first and last administrations for individuals whose retest interval is at least 6 years. Second, they do not control for the time interval. For example, the retest interval for the GZTS was substantially longer in men 30-50 (M = 20.2 years) than in men 65+ (M = 9.5 years), and stability declines over longer time intervals, at least up to 20 years.

To address these problems, for each scale we calculated a measure of individual consistency, the standard deviation for each individual across all available administrations. Because SDs are a measure of inconsistency, we reflected the values (1 – SD) to obtain an absolute measure of consistency.[2] Note that this is a conservative measure of rank-order stability, because normative changes in mean level, which do not affect rank-order, do lower these consistency scores. Mean level changes, however, are generally modest in these data (Terracciano et al., 2005,2006). All respondents with at least two data points were included in this analysis, and data from more than two administrations were used when available to obtain a best estimate of individual consistency. We also calculated the maximum retest interval as a control variable.

Methods

Analyses of individual consistency were based on 3,281 administrations of the GZTS from 1,194 individuals, and 4,217 administrations of the NEO-PI-R from the 1,178 individuals who had at least two administrations. Note that all the data reported in Table 2 and 3 are included here, but they are supplemented by data from additional administrations and retest intervals outside the limits used in the retest correlation analyses. Examination of scatterplots showed a small number of outliers (from 0 to 6 per scale), which were recoded as three standard deviations above the mean. For each scale, we conducted hierarchical multiple regressions in which we successively entered maximum retest interval and interval-squared, age, and age-squared as predictors.

Because gender does not have systematic effects on either rank-order consistency or the relation of rank-order consistency to age (e.g., Table 2), data from men and women were combined. Only data from respondents initially over age 30 were included.

---

[2]The correlations of Asendorpf's (1992) individual stability coefficient with our measure of consistency in the subsample with first and last administration at least 6 years apart ranged from .67 to .80 for the GZTS scales and NEO-PI-R factors.

Results and Discussion

Maximum retest interval and its square were entered as a block that was a significant predictor of consistency for all GZTS scales and NEO-PI-R factors, accounting for from 0.4% to 4.1% of the variance in individual consistency scores. As expected, longer retest intervals were modestly associated with lower consistency.

In the next block, age was a significant linear predictor of stability for five GZTS scales: Consistency on Restraint, Ascendance, and Personal Relations increased cross-sectionally with age, whereas consistency on General Activity declined.[3] All of these effects were less than 1%, accounting for from 0.4% to 0.7% of the variance in consistency scores. There were no significant linear age effects on Emotional Stability, Objectivity, Friendliness, Thoughtfulness, or Masculinity. Only Sociability showed significant curvilinear effects, accounting for 0.7% of the variance in consistency scores.

The only GZTS scale that showed increasing consistency with older age, at both the group (Table 2, women only) and individual level (Analysis 3), was Ascendance.

Multiple regression showed that individual consistency of the five NEO-PI-R factors was unrelated to age or to age-squared. After age 30, consistency appears to have reached a plateau, a result at the individual level that confirms the group level findings. Of the 30 NEO-PI-R facets only E3: Assertiveness, which is strongly correlated with Ascendance (Terracciano et al., 2006), showed a significant linear increase in consistency, explaining 0.4% of variance. Curvilinear trends were found for O2: Openness to Aesthetics, which increased up to age 60, and for O4: Openness to Actions and A4: Compliance that both showed the lowest consistency values in middle adulthood. Age and age-squared combined accounted for from 0.4% to 1.0% of the variance.

Analysis 4: Ipsative Stability as a Function of Age

Ipsative stability (or "personcentered continuity;" Caspi & Roberts, 2001, p. 52) refers to the stability of the configuration of personality traits in each individual. Concern for ipsative stability or change was inaugurated by Block (1971), who believed that it more adequately captured the integrated functioning of traits within the individual. The interpersonal activities of a sociable person may be inhibited by self-consciousness; a small change in the relative balance of these two dispositions might lead to large changes in social behavior. From childhood to adolescence ipsative stability appears to be rather limited, and there are substantial individual differences (Caspi & Roberts, 2001).

Ipsative stability is most frequently assessed using longitudinal Q-sort data. Q-sort instruments require that a set of items be ordered from most to least characteristic of the person, usually with a fixed distribution. Block's (1961) CAQ was originally intended for use by expert raters, but was modified by Bem and Funder (1978) for use as a self-sort. Ipsative stability is usually quantified as a Q- or inverse correlation, that is, the correlation for each individual of first and second sort across the 100 items. Costa, McCrae, and Siegler (1999) reported Q-correlations for 273 BLSA participants retested on average after 6.6 years. These values ranged from .12 to .86, with a median of .71 for men and .72 for women. In the present study we report data from an expanded BLSA sample, stratified by age group.

---

[3]The declining consistency in General Activity may seem puzzling, given the high retest correlations for that scale in Table 2. This effect is probably due to the accelerated decline in the mean level of General Activity in old age (Terracciano et al., 2005) to which our measure of consistency is sensitive.

Q-correlations based on raw CAQ data represent the degree to which individuals report the same ordering of characteristics across two time points, and, to the extent that the reports are accurate, gives a straightforward estimate of the stability of the trait configuration. However, there is a sense in which such Q-correlations are inflated, because the items of the CAQ have different normative values. At any given time, almost all individuals are likely to claim that "is genuinely dependable" is more characteristic of them than "is guileful, deceitful" (McCrae, Terracciano, Costa, & Ozer, 2006). Because of these differences in item desirability, Q-correlations between any two people are usually positive and often substantial (Ozer & Gjerde, 1989). It is not clear from an analysis of raw CAQ items how much of the observed stability is due to the relative permanence of trait configurations in the individual, and how much is due to enduring social norms about item endorsement.

McCrae et al. (2006) therefore recommended that CAQ items first be standardized across persons to remove differences due to item endorsement norms. Q-correlations can then be calculated on the standardized items, and age differences in standardized ipsative stability can be examined. These values are likely to be substantially less than the median values of . 70 typically reported in adult samples, because they exclude stability attributable to enduring social norms of item endorsement.

However, lower standardized ipsative stability coefficients would not necessarily imply that there are major changes in the configuration of traits across adulthood, because single items are apt to be unreliable. More reliable assessments of personality are given by factors from the CAQ items, which can be interpreted in terms of the dimensions of the FFM. Q-correlations can be computed across these five factor scores on two occasions.

Method

Participants and procedures—BLSA participants completed the Bem and Funder (1978) modification of Block's (1971) CAQ during their regular visits to the Gerontology Research Center. Sorting 100 items into 9 fixed categories is a challenging task, especially for older participants, and a first longitudinal analysis (Costa et al., 1999) suggested that a few participants had been confused about the direction of the sort, putting *most characteristic* items in the *least characteristic* bin. In subsequent analyses (McCrae et al., 2006) we correlated each Q-sort with the normative values of the 100 items, defined by their means across all administrations. Of 2,289 administrations, 31 showed a negative correlation with the normative values and were discarded. The CAQ was readministered approximately every 6 years; for the present study we analysed data from the first two administrations, with a minimum retest interval of 4 years. Characteristics of this subsample are given in the last line of Table 1. A comparison of individuals with and without a second CAQ sort showed that those not retested were more likely to be women (50% vs. 38%), were initially tested about 6 years later, and, curiously, were about one-eighth standard deviation higher in Conscientiousness. There were no differences on the other personality factors.

Instrument—The CAQ was developed to provide a comprehensive description of personality traits, with a particularly emphasis on clinical description (Block, 1961). In the self-sort version, respondents arrange the 100 items into 9 categories, with a fixed number of items in each, approximating a normal distribution. Analyses of some of the first administration data in the BLSA (McCrae, Costa, & Busch, 1986) showed that the CAQ included items from all five factors of the FFM, and that a five-factor structure captured most of the common variance in CAQ items (cf. Lanning, 1994). Five varimax-rotated factors were extracted from the intercorrelation of the CAQ items on both occasions, yielding the expected N, E, O, A, and C factors. Retest correlations for these factors ranged

from .70 for C to .80 for E. Q-correlations across the two occasions were computed across these five factor scores.

Results and Discussion

In this subsample of BLSA participants, Q-correlations based on raw CAQ items ranged from −.06 to .89; the median was .70 for men and .73 for women. Table 4 reports mean ipsative stability coefficients for the total and by age groups. The rows report results for raw CAQ items, standardized CAQ items, and CAQ factors. As expected, the standardized items showed lower levels of ipsative stability than the raw items, because stability coefficients are not inflated by item response norms. The clearest evidence of ipsative stability is given by the analysis of the more reliable CAQ factors, which are not inflated by item response norms, and provide a measure of the degree to which the relative ordering (i.e., profile) of the five factor scores within an individual persists over time.

Of primary interest here are the comparisons of ipsative stability across age groups. Preliminary analyses including gender as a classifying factor showed that there were no Gender × Age Group interactions, so further analyses were conducted on the combined sample. Ipsative stability based on standardized CAQ items was significantly higher in middle-aged men and women than in older adults, which is consistent with Ardelt's (2000) claim that stability declines after age 50. Here age accounted for 1.8% of the variance in Q-correlations. However, there were no significant age differences between younger and middle-aged adults, and there were no significant differences between any of the groups when ipsative stability was based on raw CAQ items or CAQ factors.[4] Results were unchanged when retest interval was used as a covariate.

General Discussion

The retest correlations in Tables 2 and 3 confirm the well-established fact that the rank-order consistency of personality traits in adulthood is quite high. For the total group, the median retest correlations across all scales is .70. This value is almost as high as that estimated by Roberts and DelVecchio (2000) for their most stable age group (.75 at age 50-59), despite the fact that the retest interval in the present study ranged from 10 to 16 years, whereas Roberts and DelVecchio's was only 6.7 years. High rank-order consistency in adulthood is equally characteristic of men and women.

Consistency as a Function of Age

The focus of interest in the present study was the relation of age to rank-order consistency after age 30. McCrae and Costa (1990) and meta-analyses by Roberts and DelVecchio (2000) and Ardelt (2000) all suggested that stability should be lower in the decade of the 20s than in later age periods, and that view was supported by supplementary analyses of the GZTS scales. The main issue the present study hoped to resolve was whether rank-order consistency reached a plateau by age 30 or continued to increase after that age. The evidence was mixed: For some GZTS scales there were significant differences between retest correlations for respondents under 50 and over age 50, and in every case, the lower correlations were found in the younger group. However, most scales, including all NEO-PI-R factors and most facets, failed to show significant differences between these age groups. Analyses of individual consistency, which controlled for retest interval, provided partial support for Roberts and Delvecchio's (2000) hypothesis of increasing consistency for four GZTS scales, but not for the other GZTS scales or for any of the NEO-PI-R factors.

---

[4]A supplementary analysis including a group of 54 men and women aged 17 to 30 showed that they did not differ from any of the older groups on the three measures of ipsative stability.

Ardelt (2000) reported a curvilinear relation of age to stability, with declines after age 50. We found no support for that view, despite the large number of old and very old participants in the BLSA. Provided that they remain cognitively intact (Siegler et al., 1991), older individuals show remarkable rank-order consistency in personality traits.

Adults of all ages also show considerable ipsative stability of personality. There were no consistent age differences in ipsative stability, and the most accurate and reliably assessed measure, based on the profile of five factor scores derived from the CAQ, showed no significant age differences. The stable "configuration of traits" predicted by McCrae and Costa (1990, p. 10) was clearly found here for adults of all ages.

All of the indicators of rank-order stability are lowered by unreliability, and several previous studies (e.g., Costa et al., 1980; Costa & McCrae, 1992c) have demonstrated that disattenuation for retest unreliability substantially increases coefficients, with estimates of the true score stability as high as .87 over a 24-year interval (Costa & McCrae, 1992c). Personality traits are indeed enduring dispositions. Nevertheless, it is also true that these estimated consistency coefficients rarely attain 1.0. The present analyses of retest interval confirmed earlier findings that stability decays slowly with the passage of time (though never approaching zero). These declines in consistency may have two sources: They may represent small accumulated effects due to random processes that affect all individuals, such as the gradual atrophy of the brain (Resnick, Pham, Kraut, Zonderman, & Davatzikos, 2003), or they may be due to the presence in the sample of a small number of individuals who show substantial personality change, perhaps as a result of traumatic events or an episode of depression. (Costa & McCrae, 1994, described these two kinds of change as the "crumbling" and "cracking" of the set plaster of personality.) The existence of a small subset of individuals with significant change can be seen in Figure 1 and had been suggested by HLM analyses of NEO-PI-R and GZTS data, where significant individual variations in longitudinal slopes were found (Terracciano et al., 2005, 2006). In this study we used a novel, simple, absolute measure of consistency at the individual level that could be related to biological and environmental variables in future studies.

The present data provide at least some support for the view that there are subtle increases in rank-order consistency between age 30 and age 50, especially for measures of ascendance or assertiveness. Roberts and DelVecchio (2000) interpreted such findings as evidence of cultural change in the century since William James opined that personality was "set like plaster" after age 30 (James, 1890/1981, p. 126). Data from that era are not available, and so subtle a change might simply have escaped James's notice. Roberts and DelVecchio's hypothesis could be tested, however, in cross-cultural longitudinal studies. If personality development is paced by the demands and opportunities of the culture, then stability should be reached more quickly in underdeveloped nations, where full adult responsibilities are assumed at earlier ages. At present, there are no longitudinal studies of personality traits in underdeveloped nations; testing the Roberts and DelVecchio hypothesis would be one of many reasons to inaugurate them.

Consistency as a Function of Time Elapsed

Although a decline in stability with increasing time interval is generally found over relatively short time intervals (ten years or less), few longitudinal studies have examined the effect of elapsed time on the consistency of personality traits over several decades. The exceptionally long time interval between GZTS assessments was used to examine the effect of time on stability at the individual level. Surprisingly, the expected pattern of continuing exponential decay was not observed: For six of the GZTS scales, individual stability coefficients were not significantly related to the length of the retest interval; for four scales,

and for mean stability, the data were best modeled by a concave quadratic curve, suggesting a decelerating pattern of decline up to about a 20 years interval and then a plateau.

Fraley and Roberts (2005) attributed non-zero asymptotic declines in rank-order consistency to developmental constancy factors such as enduring DNA. They argued that other factors, including person-environment transactions and stochastic-contextual processes, were also needed to account for long-term patterns of continuity, and claimed that "a model that excludes the role of environmental experiences … simply cannot account for the empirical patterns of stability and change that we have presented" (p. 71). But their interpretation goes beyond the data examined. It is true that there must be sources of medium-term stability that affect trait levels for periods of a few years and then fade away, but these sources need not be environmental. Genes are activated and silenced during the course of development (Fraga et al., 2005; Plomin, 1986); perhaps a subset of personality-related genes are switched on for several years and then switched off. The data Fraley and Roberts examined, like ours, do not speak to this issue.

These GZTS data are certainly not definitive. They are based on a single instrument, and relatively few of the respondents (none of the women) had retest intervals over 30 years (see Figure 1). A further complication is introduced by the fact that many participants had taken the GZTS more than twice, and repeated "practice" on the same instrument might conceivably inflate stability. However, if these GZTS data accurately reflect the developmental course of personality traits, then previous estimates of lifetime stability must be revised upwards. Costa and McCrae (1992c), using a model of exponential decay, estimated that 60% of true score variance was constant over the full 50-year adult lifespan. The present data suggest that perhaps as much as 80% is constant.

Acknowledgments

References

Ardelt M. Still stable after all these years? Personality stability theory revisited. Social Psychology Quarterly 2000;63:392–405.

Asendorpf JB. Beyond stability: Predicting inter-individual differences in intra-individual change. European Journal of Personality 1992;6:103–117.

Bem DJ, Funder DC. Predicting more of the people more of the time: Assessing the personality of situations. Psychological Review 1978;85:485–501.

Block, J. The Q-sort method in personality assessment and psychiatric research. Charles C Thomas; Springfield, IL: 1961.

Block, J. Lives through time. Bancroft Books; Berkeley, CA: 1971.

Block, J. Advancing the psychology of personality: Paradigmatic shift or improving the quality of research?. In: Magnusson, D.; Endler, NS., editors. Personality at the cross-roads: Current issues in interactional psychology. Erlbaum; Hillsdale, NJ: 1977. p. 37-64.

Caspi A, Roberts BW. Personality development across the life course: The argument for continuity and change. Psychological Inquiry 2001;12:49–66.

Conley JJ. The hierarchy of consistency: A review and model of longitudinal findings on adult individual differences in intelligence, personality, and self-opinion. Personality and Individual Differences 1984;5:11–26.

Costa PT Jr. Bagby RM, Herbst JH, McCrae RR. Personality self-reports are concurrently reliable and valid during acute depressive episodes. Journal of Affective Disorders 2005;89:45–55. [PubMed: 16203041]

Costa, PT., Jr.; McCrae, RR. The NEO Personality Inventory manual. Psychological Assessment Resources; Odessa, FL: 1985.

Costa PT Jr. McCrae RR. Personality in adulthood: A six-year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. Journal of Personality and Social Psychology 1988;54:853–863. [PubMed: 3379583]

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. Psychological Assessment Resources; Odessa, FL: 1992b.

Costa, PT., Jr.; McCrae, RR. Trait psychology comes of age. In: Sonderegger, TB., editor. Nebraska Symposium on Motivation: Psychology and aging. University of Nebraska Press; Lincoln, NE: 1992c. p. 169-204.

Costa, PT., Jr.; McCrae, RR. "Set like plaster"? Evidence for the stability of adult personality. In: Heatherton, T.; Weinberger, J., editors. Can personality change?. American Psychological Association; Washington, DC: 1994. p. 21-40.

Costa, PT., Jr.; McCrae, RR. Trait theories of personality. In: Barone, DF.; Hersen, M.; Hasselt, VBV., editors. Advanced personality. Plenum; New York: 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.; McCrae, RR.; Siegler, IC. Continuity and change over the adult life cycle: Personality and personality disorders. In: Cloninger, CR., editor. Personality and psychopathology. American Psychiatric Press; Washington, DC: 1999. p. 129-154.

Finn SE. Stability of personality self-ratings over 30 years: Evidence for an age/cohort interaction. Journal of Personality and Social Psychology 1986;50:813–818. [PubMed: 3712223]

Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, et al. Epigenetic differences arise during the lifetime of monozygotic twins. Proceedings of the National Academy of Sciences 2005;102:10604–10609.

Fraley RC, Roberts RW. Patterns of continuity: A dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course. Psychological Review 2005;112:60–74. [PubMed: 15631588]

Gough, HG. California Psychological Inventory administrator's guide. Consulting Psychologists Press; Palo Alto, CA: 1987.

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

Hathaway, SR.; McKinley, JC. The Minnesota Multiphasic Personality Inventory. rev. ed.. University of Minnesota Press; Minneapolis: 1943.

Helson R, Wink P. Personality change in women from the early 40s to the early 50s. Psychology and Aging 1992;7:46–55. [PubMed: 1558705]

James, W. The principles of psychology. Vol. Vol. 1. Harvard University Press; Cambridge, MA: 1981. Original work published 1890

Lanning K. Dimensionality of observer ratings on the California Adult Q-Set. Journal of Personality and Social Psychology 1994;67:151–160.

McCrae RR. Consensual validation of personality traits: Evidence from self-reports and ratings. Journal of Personality and Social Psychology 1982;43:293–303.

McCrae RR. Traits through time. Psychological Inquiry 2001;12:85–87.

McCrae, RR.; Costa, PT, Jr.. Personality in adulthood. Guilford; New York: 1990.

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. Busch CM. Evaluating comprehensiveness in personality systems: The California Q-Set and the Five-Factor Model. Journal of Personality 1986;54:430–446.

McCrae RR, Terracciano A, Costa PT Jr. Ozer DJ. Person-factors in the California Adult Q-Set: Closing the door on personality trait types? European Journal of Personality 2006;20:29–44.

McCrae RR, Yik MSM, Trapnell PD, Bond MH, Paulhus DL. Interpreting personality profiles across cultures: Bilingual, acculturation, and peer rating studies of Chinese undergraduates. Journal of Personality and Social Psychology 1998;74:1041–1055. [PubMed: 9569658]

Ozer DJ, Gjerde PF. Patterns of personality consistency and change from childhood through adolescence. Journal of Personality 1989;57:483–507. [PubMed: 2769564]

Plomin, R. Development, genetics, and psychology. Erlbaum; Hillsdale, NJ: 1986.

Resnick SM, Pham DL, Kraut MA, Zonderman AB, Davatzikos C. Longitudinal Magnetic Resonance Imaging studies of older adults: A shrinking brain. Journal of Neuroscience 2003;23:3295–3301. [PubMed: 12716936]

Roberts BW, DelVecchio WF. The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. Psychological Bulletin 2000;126:3–25. [PubMed: 10668348]

Shock, NW.; Greulich, RC.; Andres, R.; Arenberg, D.; Costa, PT., Jr.; Lakatta, EG., et al. Normal human aging: The Baltimore Longitudinal Study of Aging. National Institutes of Health; Bethesda, MD: 1984. (NIH Publication No. 84-2450)

Siegler IC, Welsh KA, Dawson DV, Fillenbaum GG, Earl NL, Kaplan EB, et al. Ratings of personality change in patients being evaluated for memory disorders. Alzheimer Disease and Associated Disorders 1991;5:240–250. [PubMed: 1781966]

SoftStat, Inc.. Statistics II [Computer software and manual]. Vol. Vol. 3. Author; Tulsa, OK: 1995. Statistica.

Terracciano A, McCrae RR, Costa PT Jr. Longitudinal trajectories in Guilford-Zimmerman Temperament Survey data: Results from the Baltimore Longitudinal Study of Aging. Journal of Gerontology: Psychological Sciences 2006;61B:P108–P116.

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;20:493–506. [PubMed: 16248708]
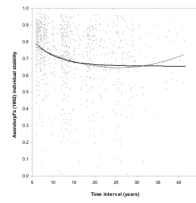
Figure 1.
Mean individual stability coefficients ($N = 1,063$) across ten GZTS scales as a function of retest interval in years. Significant quadratic (dotted line; $R^2 = .027$) and exponential decay (solid line; $R^2 = .030$) curves are superimposed. Data from 15 individuals whose mean coefficients are less than 0.0 do not appear in the Figure, but were used in the analyses.

Table 1

Description of the Samples.

| Instrument | N | | Initial Age | | 1st. Administration | |
	Males	Females	M	Range	M	Range
GZTS, Men	737	—	50.9	30-87	1969	1959-1996
GZTS, Women	—	326	52.8	30-82	1983	1978-1996
NEO-PI-R	367	309	60.6	30-89	1992	1989-1998
NEO Inventory	186	71	51.7	30-80	1980	—
CAQ	322	195	48.2	30-83	1986	1981–1994

Note. Initial age is given in years.

Table 2

Rank-Order Consistency Coefficients for GZTS Scales for the Full Sample and Different Age Groups

		Age Group		
Scale/Factor	Total	30-50	50-65	>65
Men				
General Activity	.75	.71	.78	.75
Restraint	.68	.62	.70	.72
Ascendance	.78	.77	.77	.79
Sociability	.73	.68[a]	.79	.74
Emotional Stability	.64	.59[a]	.70	.66
Objectivity	.67	.63	.72	.68
Friendliness	.66	.63	.71	.68
Thoughtfulness	.65	.60[a]	.73[b]	.61
Personal Relations	.67	.66	.72	.68
Masculinity	.70	.67	.74	.69
Mdn	*.68*	*.65*	*.73*	*.69*
n	600-682	315-350	189-215	96-117
M retest interval (range)	16.6 (6-42)	20.2 (6-41)	14.7 (6-37)	9.5 (6-21)
Women				
General Activity	.76	.69[a]	.82	.78
Restraint	.70	.72	.63	.75
Ascendance	.78	.72[a]	.85	.81
Sociability	.78	.75	.82	.80
Emotional Stability	.69	.59[a,c]	.76	.76
Objectivity	.69	.67	.67	.77
Friendliness	.66	.64	.57[b]	.77
Thoughtfulness	.71	.69	.70	.75
Personal Relations	.68	.66	.69	.72
Masculinity	.76	.68[a,c]	.82	.83
Mdn	*.71*	*.69*	*.73*	*.77*
n	252-305	116-139	82-101	54-68
M retest interval (range)	10.5 (6-24)	10.8 (6-24)	11.2 (6-21)	9.1 (6-15)

Note. All coefficients are significant at *p* < .001.

[a]Correlation for Age 30-50 differs from correlation for Age 50-65, *p* < .05.

[b]Correlation for Age 50-65 differs from correlation for Age > 65, *p* < .05.

[c]Correlation for Age 30-50 differs from correlation for Age > 65, *p* < .05. GZTS = Guilford Zimmermann Temperament Survey.

Table 3

Rank-Order Consistency Coefficients for NEO-PI-R Scales for the Full Sample and Three Age Groups

		Age Group		
Factor/Facet	Total (*n* = 676)	30-50 (*n* = 151)	50-65 (*n* = 259)	>65 (*n* = 266)
N: Neuroticism	.78	.79	.79	.78
E: Extraversion	.83	.84	.85[a]	.79
O: Openness	.85	.82	.87	.86
A: Agreeableness	.80	.79	.80	.80
C: Conscientiousness	.81	.76	.82	.83
Mdn	*.81*	*.79*	*.82*	*.80*
N1:Anxiety	.71	.76	.69	.71
N2:Angry Hostility	.66	.68	.63	.69
N3:Depression	.62	.62	.63	.64
N4:Self-Consciousness	.67	.64	.71	.65
N5:Impulsiveness	.62	.59	.62	.64
N6:Vulnerability	.65	.61	.65	.67
E1:Warmth	.71	.73	.71	.70
E2:Gregariousness	.77	.81[b]	.81[a]	.68
E3:Assertiveness	.79	.76	.80	.80
E4:Activity	.74	.68	.74	.73
E5:Excitement-Seeking	.74	.71	.77	.71
E6:Positive Emotions	.71	.64	.73	.71
O1:Fantasy	.73	.70	.73	.71
O2:Aesthetics	.82	.81	.83	.81
O3:Feelings	.69	.75[b]	.65	.64
O4:Actions	.73	.71	.72	.73
O5:Ideas	.79	.78	.78	.80
O6:Values	.68	.56[b]	.67	.73
A1:Trust	.67	.65	.71	.64
A2:Straightforwardness	.66	.61	.71	.63
A3:Altruism	.65	.60	.64	.69
A4:Compliance	.71	.70	.72	.70
A5:Modesty	.70	.72	.69	.71
A6:Tender-Mindedness	.64	.68	.66	.61
C1:Competence	.66	.61	.66	.64
C2:Order	.75	.75	.76	.74
C3:Dutifulness	.57	.59	.55	.58
C4:Achievement Striving	.72	.68	.75	.70
C5:Self-Discipline	.70	.65	.70	.72
C6:Deliberation	.67	.63	.71	.65
Mdn	*.70*	*.68*	*.71*	*.70*
M retest interval (range)	*10 (6-15)*	*10.1 (6-15)*	*10.3 (6-14)*	*9.6 (6-14)*

Note. Note. All coefficients are significant at *p* < .001.

[a]Correlation for Age 50-65 differs from correlation for Age > 65, *p* < .05.

[b]Correlation for Age 30-50 differs from correlation for Age > 65, *p* < .05.

Table 4

Mean Q-Correlations for Different Age Groups

Q-Correlation for	Age Group			
	Total	30-50	50-65	65+
Raw CAQ Items	.69	.69	.70	.68
Standardized CAQ Items	.45	.45	.46[a]	.41[a]
CAQ Factors	.72	.71	.73	.75
n	463	245	138	80
M retest interval (range)	6.7 (4-11)	6.9 (4-11)	6.6 (5-10)	6.4 (4-10)

[a]Groups are significantly different by Scheffé post-hoc test ($p < .05$).