

# Age Trends and Age Norms for the NEO Personality Inventory–3 in Adolescents and Adults

**Robert R. McCrae**

*National Institute on Aging*

**Thomas A. Martin**

*Susquehanna University*

**Paul T. Costa, Jr.**

*National Institute on Aging*

*The NEO Personality Inventory–3 (NEO-PI-3) is a modification of the Revised NEO Personality Inventory (NEO-PI-R) designed to be more understandable to adolescents. Data from adults aged 21 to 91 showed that the NEO-PI-3 also functions as well or better than the NEO-PI-R in adults. Age trends from combined adolescent (n = 500) and adult (n = 635) samples confirmed previous cross-sectional findings and demonstrated the importance of studying age changes especially at the facet level and during the decade of the 20s. Normative data for self-report and observer rating forms for adolescents, younger and older adults, and all adults are discussed, as well as for a combined-age group. It is argued that combined-age norms may be most appropriate for depicting the personality scores of individuals, but the utility for some purposes of within-age group scores is also acknowledged.*

**Keywords:** personality; age norms; development; adolescents; test revision

Although there is a wealth of data on age differences and changes in the mean levels of personality traits in adulthood (McCrae & Costa, 2003; Roberts, Walton, & Viechtbauer, in press), there is surprisingly little on personality development in adolescence (McCrae et al., 2002) and across the decade of the 20s. One reason for this gap in our knowledge of personality development is the uncertain applicability of personality measures, typically developed on college students or adults, to younger populations. Recent research using the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) has

suggested that adult measures retain validity when used in children as young as 10. For example, Baker and Victor (2003) replicated the adult factor structure in a predominantly African American sample of 10- to 16-year-olds, and De Fruyt, Mervielde, Hoekstra, and Rolland (2000) showed convergent and discriminant validity for NEO-PI-R scales in correlations with a children's personality inventory in a sample of Flemish children aged 12 to 17. However, both studies also reported that some children have difficulty reading and understanding some items.

---

We thank each of the 56 undergraduate research assistants who recruited participants for this project. Furthermore, we extend special thanks to the research assistants who mailed questionnaires, distributed periodic reminders, scanned data, and managed the databases necessary for this study: Erin Weller, Rachael Gebely, Rebecca Fritschie, Kathy Wulderk, and Amanda Jones. Their dedication to the project, fastidious work, and good humor are deeply appreciated. Robert R. McCrae and Paul T. Costa, Jr., receive royalties from the Revised NEO Personality Inventory. Address correspondence to Robert R. McCrae, Box #03, Gerontology Research Center, 5600 Nathan Shock Drive, Baltimore, MD 21224-6825; e-mail: mcccraej@grc.nia.nih.gov.

*Assessment*, Volume 12, No. 4, December 2005 363-373

DOI: 10.1177/1073191105279724

© 2005 Sage Publications

In consequence, McCrae, Costa, and Martin (2005) developed a more readable version of the NEO-PI-R—the NEO Personality Inventory-3 (NEO-PI-3)—by replacing difficult or poorly performing items. A total of 48 items were candidates for replacement, and two trial items were written for each. Of these 96 items, 37 were selected for the NEO-PI-3. When administered to a sample of adolescents aged 14 to 20, the NEO-PI-3 showed modest improvements in psychometric properties over the NEO-PI-R as well as greater readability. The median item/total correlation for the 37 items that were changed increased from .28 in the NEO-PI-R to .37 in the NEO-PI-3, and the median Flesch-Kinkaid reading grade level decreased from 8.3 to 4.4. McCrae et al. suggested that the NEO-PI-3 might also be useful in adults, especially those with limited education.

The scarcity of personality data on individuals in their 20s has other causes. Much research has, of course, been conducted on college students, including a few longitudinal studies (e.g., Robins, Fraley, Roberts, & Trzesniewski, 2001). Similarly, adult developmentalists have studied middle-aged (Costa, Herbst, McCrae, & Siegler, 2000) and older (Mroczek & Spiro, 2003) adults but, until recently (Neyer & Asendorpf, 2001), have shown less interest in development in early adulthood. This is unfortunate, because a review of lifespan development by Costa and McCrae (2002) suggested that the largest changes (about 0.5 standard deviations) were to be found between age 18 and age 30. During that period, Neuroticism (N) and Extraversion (E) Decline and Agreeableness (A) and Conscientiousness (C) increase at a more rapid rate than in the remainder of life (where a cumulative change of less than 0.5 standard deviations is typically seen). Openness to experience (O) appears to increase after age 12 but to have declined by age 30. It is not clear when O peaks or whether the relatively rapid changes in the other factors are confined to the first part of early adulthood or continue throughout the decade of the 20s. Costa and McCrae (2002) predicted that there would be no noticeable changes in A or C after age 30, but subsequent data support the view that there is gradual but persistent change in these factors as well (Roberts et al., in press).

Although longitudinal data from early adolescence on would be desirable, the present study estimates age changes by examining cross-sectional age differences in adolescence and adulthood. NEO-PI-R and NEO-PI-3 data from McCrae et al. (2005) are combined with new data from adults aged 21 to 91; as a necessary step toward that cross-sectional comparison, we first evaluate the psychometric properties of the NEO-PI-3 in the adult sample.

Aside from its importance in understanding trait development, research on personality traits in different age groups is a practical necessity for preparing appropriate age norms. Researchers and clinicians are routinely taught

that test scores should be interpreted by comparison to normative data, but there is some confusion about which norms are appropriate for different uses. The Civil Rights Act of 1991 banned the use of race and gender norms in job hiring or promotion (Cohen & Swerdlik, 2002), but many clinicians still prefer to interpret scores using within-gender norms. In this article, we describe normative data for the NEO-PI-3 for various age and gender groups to accommodate different needs for normative information. We also discuss the use and interpretation of age norms and suggest that for many purposes in research and individual assessment, combined-age norms are most meaningful.

## STUDY 1: SCALE VALIDATION

### Method

*Participants and procedure.* Participants were 652 individuals ranging in age from 21 to 91 who constituted the second phase of a larger study spanning the adolescent and adult lifespan. Age and sex were stratified, with overrepresentation of the early-adult groups, because evidence suggests that age changes are most marked then (Costa & McCrae, 2002). Participants responded to two questionnaires, one (Form S) to describe themselves and one (Form R) to describe a target individual with whom they were well acquainted. Protocols were screened for validity using the criteria specified in the NEO manual (Costa & McCrae, 1992) for validity checks (which ask respondents if they have answered honestly and in the right spaces), random responding (assessed by repetitive strings of identical responses; e.g., more than six consecutive *strongly disagrees*), and missing items,<sup>1</sup> and all further analyses were conducted on the 635 respondents (356 women, 279 men) who had valid protocols for both Form S and Form R of the NEO-PI-R. Participants resided in 29 states, although most (63.0%) were from Pennsylvania. The sample was predominantly White (92.6%) with 1.6% Asian/Pacific Islander, 1.1% Black, 1.3% Hispanic, and the remainder other or missing ethnicity. High academic achievers were overrepresented in this sample, where 75.6% had more than a high school education and 26.8% more than a baccalaureate degree. Participants were also relatively affluent with only 8.8% reporting annual household incomes less than \$20,000 and 19.3% reporting annual household incomes greater than \$100,000. English was the first language of all but 4 participants.

Potential participants were contacted by 1 of 56 undergraduate research assistants from two liberal arts colleges located in the Eastern and Southern United States. Most participants were acquaintances of the research assistants.

Research assistants described the study to potential participants, and if an individual expressed interest in participating, the research assistant submitted a referral with the identity, age, and contact information for the participant. Questionnaires were then mailed to participants.

To implement the cross-observer design, an attempt was made to recruit participants in pairs in which both individuals were aged 21 and older and both agreed to rate and be rated by the other. Complete data were obtained from 266 pairs. Individuals not paired chose a target they knew well to rate and were asked to specify their relationship to the target. The targets (341 women, 293 men) were siblings (6.8%), friends (18.3%), spouses (63.7%), domestic partners (8.5%), or parents/children (2.7%). Targets ranged in age from 18 to 90.

Research assistants were instructed to periodically contact individuals they had recruited and to whom questionnaires had been mailed. They asked these prospective participants whether they had questions about the task, reminded them that they were free to break their work on the questionnaires into sessions of a convenient number and length, and encouraged them to complete the questionnaires. To aid this process, research assistants received, on a weekly basis, a list of the names and phone numbers of persons they had referred in the previous 10 weeks who had not yet returned their questionnaires. When completed questionnaires were received, the participant and the research assistant were each paid \$20. These procedures produced response rates of 71.0% for women and 63.1% for men. Although older cohorts generally yielded higher response rates than younger cohorts, the trend was not statistically significant.

**Instruments.** The NEO-PI-R is a 240-item questionnaire that assesses 30 specific traits (or facets), six for each of the five basic personality dimensions: N, E, O, A, and C. Items are answered on a 5-point Likert-type scale from *strongly disagree* to *strongly agree*. Scales are roughly balanced to control for the effects of acquiescence. Data on the reliability and validity of the instrument are summarized in the NEO manual (Costa & McCrae, 1992). Parallel Form S (for *self-report*) and Form R (for *observer rating*) versions have been validated. The instructions and items are the same, except that the items are phrased in the first person in Form S and in the third person in Form R. In the present study, participants were administered a 336-item questionnaire that included the NEO-PI-R items and the 37 replacement items selected for the NEO-PI-3 along with the 59 trial items not selected by McCrae et al. (2005).

## Results

**Preliminary analyses.** NEO-PI-R items with words unfamiliar to adolescents were replaced in developing the

NEO-PI-3. In the present adult sample, respondents were asked to circle any word or phrase they did not understand. Far fewer items were circled in this sample than in the adolescent sample. For example, *fastidious* was circled by 3.1% of the adult sample but by 11.4% of the adolescent sample. Nevertheless, the same words were problematic for both samples: *fastidious* (times circled in Form S by adults = 20), *panhandlers* (9), *lackadaisical* (9), *methodical* (9), *permissiveness* (8), and *aesthetic* (6). Items with all these words were replaced in the NEO-PI-3.

Respondents are instructed to leave items blank if they do not understand them. In analyses of Form S data in the present sample, 237 of the NEO-PI-R items and all 240 of the NEO-PI-3 items were answered by more than 98% of the sample, and all but two of the NEO-PI-3 items were answered by more than 99% of the sample.

To gauge the representativeness of the sample of 635 valid protocols, we scored the NEO-PI-R and compared mean values with the published adult norms (Costa & McCrae, 1992). Both men and women had means within the average range ( $T = 45-55$ ) for all domains and facets. The most noticeable difference was that men in the present sample scored lower in O ( $T = 45$ ) than the published values, perhaps because the original norms included especially well-educated men from the Baltimore Longitudinal Study of Aging (Costa & McCrae, 1986).

To assess the item selection decisions made in developing the NEO-PI-3, we calculated corrected item/facet correlations for the 37 items discarded from the NEO-PI-R and the 37 new items in the NEO-PI-3. For Form S, the correlation for the new item was as high or higher than for the old item in 30 of 37 comparisons, and the median item/facet correlation for these 37 items increased from .36 to .45.<sup>2</sup> (For the 240 NEO-PI-3 items, item/facet correlations ranged from .16-.70, median = .47.) For Form R, the correlation for the new item was as high or higher than for the old item in 27 of 37 comparisons, and the median correlation increased from .38 to .48. The new items appear to be appropriate for adult respondents.

**Scale analyses.** Table 1 reports analyses of scales for both the NEO-PI-R and the NEO-PI-3; they show that both work well in this adult sample, with adequate internal consistency and cross-observer validities that compare favorably to most studies using adults (cf. McCrae, Costa, Martin, et al., 2004). The NEO-PI-3 scales appear to offer a slight improvement in median internal consistency. Columns 2 and 3 report coefficient alphas for the two versions of Form S. At the domain level, the two versions have equivalent reliability (median  $\alpha = .90$ ); at the facet level, revisions led to alphas that are the same or higher in 17 of the 19 cases. However, an alpha of .54 for O4 (actions) is marginal even in the revision. Columns 5 and 6 report the same statistics for Form R with similar results. Across all

**TABLE 1**  
**Internal Consistency, Cross-Version**  
**Correlations, and Cross-Observer Correlations**  
**for Revised NEO Personality Inventory–**  
**Revised (NEO-PI-R) and NEO-PI-3 Scales**

Scale	Form S			Form R			Cross-Observer $r$	
	Coefficient $\alpha$			Coefficient $\alpha$				
	NEO-PI-R	NEO-PI-3	$r_{R3}^a$	NEO-PI-R	NEO-PI-3	$r_{R3}$	NEO-PI-R	NEO-PI-3
<b>Domains</b>								
N	.92	.93	.99	.93	.93	.99	.54	.55
E	.89	.89	.99	.90	.90	.99	.64	.65
O	.88	.89	.99	.88	.88	.99	.57	.57
A	.90	.90	.99	.93	.93	.99	.59	.58
C	.91	.92	.98	.93	.94	.99	.54	.52
Median	.90	.90	.99	.93	.93	.99	.57	.57
<b>Facets</b>								
N1	.82	.83	.98	.81	.84	.98	.49	.51
N2 <sup>b</sup>	.75			.85			.53	
N3 <sup>b</sup>	.83			.83			.51	
N4	.70	.77	.93	.74	.77	.94	.38	.43
N5	.68	.66	.92	.73	.71	.93	.44	.42
N6 <sup>b</sup>	.77			.80			.49	
E1 <sup>b</sup>	.79			.83			.58	
E2 <sup>b</sup>	.76			.79			.61	
E3	.76	.77	.98	.77	.77	.97	.55	.56
E4	.68	.69	.95	.71	.73	.95	.53	.53
E5	.68	.69	.96	.69	.72	.96	.65	.68
E6	.79	.80	.96	.79	.81	.97	.43	.45
O1 <sup>b</sup>	.75			.75			.40	
O2	.81	.83	.98	.82	.84	.99	.57	.57
O3 <sup>b</sup>	.71			.74			.44	
O4	.51	.54	.90	.54	.62	.90	.37	.42
O5 <sup>b</sup>	.81			.84			.59	
O6	.70	.70	.92	.69	.68	.93	.52	.51
A1	.83	.82	.97	.86	.86	.98	.46	.44
A2	.75	.76	.98	.81	.81	.98	.38	.39
A3 <sup>b</sup>	.78			.81			.53	
A4	.71	.71	.97	.78	.78	.98	.59	.56
A5	.73	.76	.98	.82	.82	.98	.47	.48
A6	.58	.69	.86	.65	.78	.89	.42	.48
C1	.73	.75	.87	.78	.80	.90	.51	.43
C2	.72	.80	.90	.78	.85	.93	.60	.65
C3	.70	.70	.91	.75	.75	.94	.47	.44
C4	.71	.77	.96	.74	.80	.96	.51	.48
C5 <sup>b</sup>	.78			.85			.46	
C6 <sup>b</sup>	.76			.78			.44	
Median	.75	.76		.78	.80		.50	.49

NOTE:  $N = 635$  except for cross-observer correlations ( $n = 532$ , 266 couples). Median for facets is across all 30 scales. All cross-observer correlations are significant at  $p < .001$ . N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness.

a. Equivalence coefficient, NEO-PI-R with NEO-PI-3 scale.

b. Facet unchanged.

30 facet scales, the median coefficient alpha increases from .75 to .76 for Form S and from .78 to .80 for Form R.

For Form S and Form R, respectively, columns 4 and 7 of Table 1 show equivalence correlations between the old

and new versions of the scales that were revised. The domain scores are virtually unchanged, with all correlations at .98 or higher. In self-reports, the facet scale correlations range from .86 for A6: Tender-Mindedness (where half the items were changed) to .98. Comparable results are seen for observer ratings. All these correlations support the view that the new scales measure equivalent constructs. However, paired  $t$  tests showed significant differences between the old and new versions of 14 of the 19 changed facet scales ranging in magnitude from 0.17 to 1.19 raw score points. Although these differences are rather small, and NEO-PI-R norms might continue to be used with the NEO-PI-3, the present study presents an opportunity to prepare contemporary normative data with different age groups.

Columns 8 and 9 of Table 1 report cross-observer correlations. Revisions made little difference at the domain level. At the facet level, notable improvements were found for N4: Self-Consciousness), O4: Actions, A6: Tender-Mindedness, and C2: Order. However, cross-observer validity declined in this sample for C1: Competence.

**Factor structure.** Principal component analysis was conducted on Form S and Form R facets for both the NEO-PI-R and the NEO-PI-3. In each case, parallel analysis (Cota, Longman, Stewart, Holden, & Fekken, 1993) clearly indicated the presence of five factors. McCrae, Zonderman, Costa, Bond, and Paunonen (1996) recommended that factor replication be evaluated after rotation toward the normative adult Form S structure (Costa & McCrae, 1992); results for Form S are reported in Table 2. The structures of the original and revision are very similar, and both closely approximate the adult NEO-PI-R structure, with factor congruence coefficients ranging from .97 to .99. Results for Form R are similar, with factor congruence coefficients ranging from .94 to .98 after targeted rotation.

**Gender differences.** The present sample shows familiar gender differences, with women scoring higher than men on N, E, O, and A factors. We computed  $d$  scores contrasting men and women on the facet scales; rank-order correlations across the 30 facets showed that almost identical patterns of gender differences were found for the old and new versions of Form S scales ( $r = .96$ ,  $p < .001$ ) and Form R scales ( $r = .98$ ,  $p < .001$ ) and that all four sets of scales showed patterns similar to that seen in Costa, Terracciano, and McCrae (2001) for American data ( $r_s = .74$  to  $.85$ ,  $p_s < .001$ ).

## STUDY 2: AGE DIFFERENCES IN NEO-PI-3 SCALES

Analyses in Study 1 suggest that the NEO-PI-3 works at least as well as the NEO-PI-R when administered to

**TABLE 2**  
**Factor Structure for Form S Revised NEO Personality Inventory–(R) and NEO**  
**Personality Inventory–3 (3) Facet Scales After Targeted Rotation (N = 635)**

Facet Scale	Procrustes-Rotated Principal Component											
	N		E		O		A		C		VC <sup>a</sup>	
	R	3	R	3	R	3	R	3	R	3	R	3
N1: Anxiety	<b>.83</b>	<b>.84</b>	.00	.00	-.06	-.07	-.05	-.03	.00	-.02	.99**	.99**
N2: Angry hostility <sup>b</sup>	<b>.62</b>	<b>.63</b>	-.05	-.04	-.09	-.09	<b>-.53</b>	<b>-.53</b>	-.03	-.04	.99**	.99**
N3: Depression <sup>b</sup>	<b>.80</b>	<b>.80</b>	-.12	-.11	-.03	-.03	-.02	-.01	-.24	-.27	.99**	.99**
N4: Self-Consciousness	<b>.76</b>	<b>.74</b>	-.24	-.31	-.06	-.07	.15	.15	-.15	-.14	.99**	.98**
N5: Impulsiveness	<b>.47</b>	<b>.53</b>	.36	.35	-.06	-.02	-.16	-.20	-.39	-.37	.99**	.99**
N6: Vulnerability <sup>b</sup>	<b>.70</b>	<b>.69</b>	-.09	-.07	-.08	-.11	.01	.03	<b>-.43</b>	<b>-.46</b>	.99**	.99**
E1: Warmth <sup>b</sup>	-.12	-.13	<b>.74</b>	<b>.74</b>	.10	.11	<b>.43</b>	<b>.43</b>	.06	.06	.99**	.99**
E2: Gregariousness <sup>b</sup>	-.10	-.12	<b>.72</b>	<b>.72</b>	.06	.04	.04	.06	-.19	-.20	.97**	.97**
E3: Assertiveness	-.28	-.27	<b>.45</b>	<b>.46</b>	.14	.15	<b>-.46</b>	<b>-.46</b>	.26	.26	.97**	.97**
E4: Activity	-.03	-.04	<b>.48</b>	<b>.53</b>	.11	.10	-.30	-.31	<b>.41</b>	<b>.41</b>	.99**	.99**
E5: Excitement-Seeking	-.05	-.04	<b>.55</b>	<b>.50</b>	.18	.22	<b>-.40</b>	<b>-.42</b>	-.15	-.12	.98**	.97**
E6: Positive Emotions	-.19	-.19	<b>.65</b>	<b>.65</b>	.35	.34	.19	.18	.15	.17	.95**	.95**
O1: Fantasy <sup>b</sup>	.19	.19	.12	.12	<b>.64</b>	<b>.64</b>	-.10	-.10	-.15	-.15	.97**	.97**
O2: Aesthetics	.13	.10	.03	.04	<b>.77</b>	<b>.76</b>	.20	.22	.03	.02	.99**	.99**
O3: Feelings <sup>b</sup>	.28	.28	.43	.43	<b>.52</b>	<b>.52</b>	.13	.11	.29	.28	.95**	.96**
O4: Actions	-.29	-.35	.20	.27	<b>.55</b>	<b>.54</b>	.04	.03	-.15	-.13	.98**	.97**
O5: Ideas <sup>b</sup>	-.13	-.12	.01	.00	<b>.78</b>	<b>.79</b>	-.08	-.08	.20	.18	.99**	.99**
O6: Values	-.05	-.05	.23	.21	<b>.46</b>	<b>.53</b>	.02	.08	.03	.03	.87*	.88*
A1: Trust	-.28	-.28	.31	.29	.06	.06	<b>.65</b>	<b>.65</b>	.00	.00	.98**	.98**
A2: Straightforwardness	-.05	-.03	-.07	-.10	-.08	-.07	<b>.73</b>	<b>.74</b>	.20	.22	.99**	.99**
A3: Altruism <sup>b</sup>	-.04	-.04	<b>.44</b>	<b>.44</b>	.02	.04	<b>.65</b>	<b>.64</b>	.26	.28	.98**	.98**
A4: Compliance	-.22	-.22	-.09	-.14	.01	-.02	<b>.77</b>	<b>.74</b>	-.03	-.08	.99**	.99**
A5: Modesty	.18	.18	-.14	-.18	-.09	-.09	<b>.68</b>	<b>.68</b>	-.02	-.05	.98**	.98**
A6: Tender-Mindedness	.12	.18	.27	.31	.14	.16	<b>.62</b>	<b>.56</b>	-.01	.08	.99**	.97**
C1: Competence	-.36	<b>-.41</b>	.17	.15	.11	.17	.09	.06	<b>.73</b>	<b>.74</b>	.99**	.99**
C2: Order	.08	.06	.05	.05	-.07	-.16	-.06	-.06	<b>.68</b>	<b>.64</b>	.97**	.98**
C3: Dutifulness	-.14	-.12	.00	.01	-.03	.03	.29	.34	<b>.74</b>	<b>.74</b>	.99**	.99**
C4: Achievement-Striving	-.07	-.11	.16	.23	.13	.13	-.19	-.17	<b>.79</b>	<b>.76</b>	.99**	.99**
C5: Self-Discipline <sup>b</sup>	-.30	-.29	.07	.06	-.02	-.01	.06	.05	<b>.78</b>	<b>.79</b>	.99**	.99**
C6: Deliberation <sup>b</sup>	-.23	-.22	-.27	-.29	.02	.02	.23	.24	<b>.60</b>	<b>.61</b>	.99**	.99**
Congruence <sup>c</sup>	.99**	.98**	.99**	.98**	.98**	.98**	.99**	.99**	.98**	.97**	.98**	.98**

NOTE: Components are rotated toward the adult normative structure (Costa & McCrae, 1992). Loadings greater than .40 in absolute magnitude are given in boldface. N = neuroticism; E = extraversion; O = openness; A = agreeableness; C = conscientiousness; VC = variable congruence.

a. Total congruence coefficient in the last row.

b. Scale unchanged in revision.

c. Congruence with adult normative NEO-PI-R structure.

\*Congruence higher than that of 95% of rotations from random data. \*\*Congruence higher than that of 99% of rotations from random data.

adults, and thus, it can appropriately be used for age comparisons with adolescents.

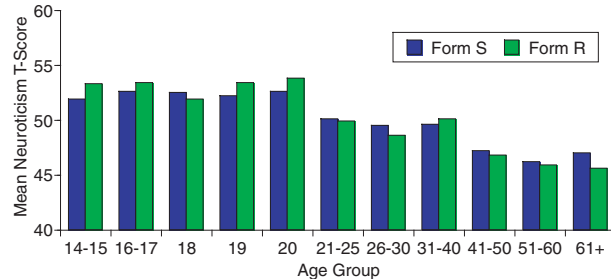
## Method

Age comparisons are based on a sample that combines the adults in Study 1 with adolescents described in McCrae et al. (2005). Briefly, the adolescents consisted of 242 males and 258 females aged 14 to 20, recruited and tested by methods similar to those described in Study 1. They were predominantly White (84.6%), and high academic achievers were overrepresented in the sample. They com-

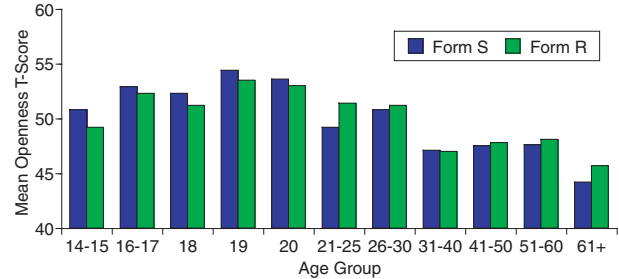
pleted the 336-item questionnaire described in Study 1 from which the NEO-PI-3 was scored. Adolescents also provided observer ratings of another adolescent, either a paired sibling or a target they chose.

In the combined sample, participants and targets were divided into 11 age groups: 14 to 15 ( $n_s = 94$  for Form S, 94 for Form R), 16 to 17 (110, 119), 18 (100, 82), 19 (98, 80), 20 (98, 90), 21 to 25 (119, 134), 26 to 30 (99, 100), 31 to 40 (59, 63), 41 to 50 (155, 151), 51 to 60 (100, 98), and 61+ (103, 103). Note that age was missing or out of range for 21 Form R targets who were excluded from subsequent analyses.

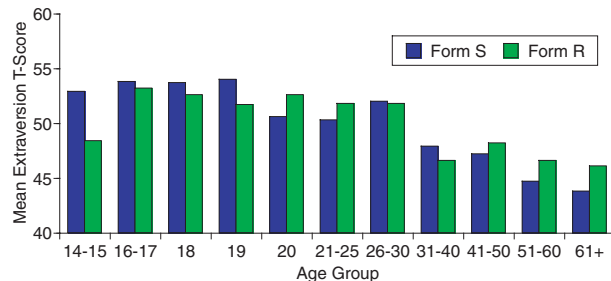
**FIGURE 1**  
**Mean Levels of NEO Personality Inventory–3**  
**Neuroticism *T*-Scores in 11 Age Groups for**  
**Self-Reports (Form S) and Observer Ratings**  
**(Form R)**



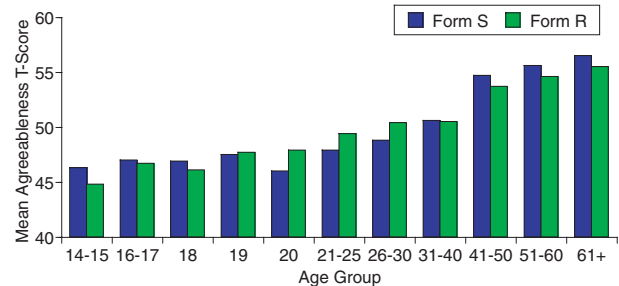
**FIGURE 3**  
**Mean Levels of NEO Personality Inventory–3**  
**Openness *T*-Scores in 11 Age Groups for**  
**Self-Reports (Form S) and Observer Ratings**  
**(Form R)**



**FIGURE 2**  
**Mean Levels of NEO Personality Inventory–3**  
**Extraversion *T*-Scores in 11 Age Groups for**  
**Self-Reports (Form S) and Observer Ratings**  
**(Form R)**



**FIGURE 4**  
**Mean Levels of NEO Personality Inventory–3**  
**Agreeableness *T*-Scores in 11 Age Groups for**  
**Self-Reports (Form S) and Observer Ratings**  
**(Form R)**



## Results

Two-way ANOVAs were first conducted on Form S and Form R data, using gender and the age group as classifying factors. The expected gender differences were found, with women scoring higher than men in N, E, O, and A, but not C. All 35 domains and facets showed significant age main effects in either self-reports or observer ratings, and 29 of them showed significant, replicated effects in both. However, there were only six significant Gender  $\times$  Age Group interactions for Form S, and only four for Form R, and none of these remained significant after Bonferroni correction.<sup>3</sup> All subsequent analyses were therefore conducted on combined gender groups.

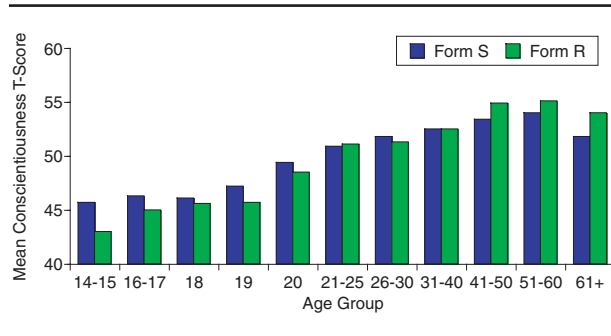
Age effects for the five domains, all significant at  $p < .001$ , are depicted in Figures 1 to 5. Age trends are entirely consistent with previous findings: N and E decline cross-sectionally with age, whereas A and C increase; O shows a

curvilinear trend peaking at age 19. Visual inspection suggests that, except in the case of O, there is no change during adolescence (age 14 to 20), and Scheffé post hoc tests show that these five groups do not differ significantly on any of the domains, including O. It is thus sensible to combine them to define adolescent norms.<sup>4</sup>

Similarly, groups older than age 30 show very similar mean levels. By Scheffé tests, only self-reported A shows evidence of an age difference (adults aged 31 to 40 score lower than adults older than 40). It would be meaningful to combine these four groups into an older adult category.

It is less clear how to handle adults in their 20s. With respect to E and O, they appear to resemble adolescents; for N and A, they seem closer to older adults; and in C, they are intermediate. Scheffé tests show that individuals from the first and latter half of this decade do not differ from each other, so it is possible to make separate norms applicable to men and women in their 20s.

**FIGURE 5**  
**Mean Levels of NEO Personality Inventory–3**  
**Conscientiousness T-Scores in 11 Age Groups**  
**for Self-Reports (Form S) and Observer Ratings**  
**(Form R)**



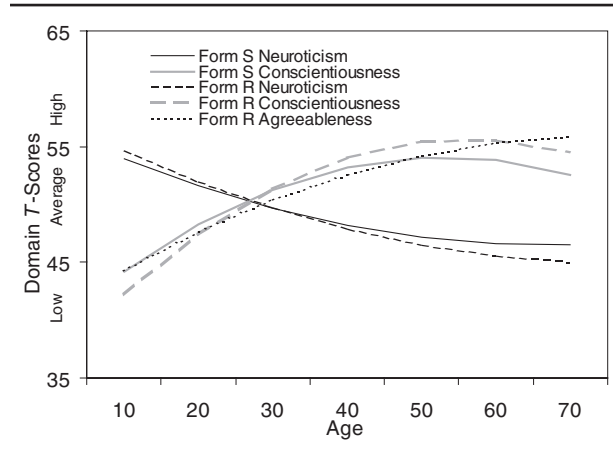
For practical purposes, however, having multiple sets of norms is cumbersome. Individuals older than age 21 are legally adults, and the difference between adults in their 20s and older adults is small in magnitude, so it may be reasonable to continue the practice of the NEO-PI-R and offer a single set of adult norms for everyone 21 and older.

The data in the figures are plotted against one last standard—namely, norms for the total group. Given that the structure of personality is essentially unchanged across the lifespan after age 14 and that the same set of questions validly assesses the constructs in different age groups, it is meaningful to compare scores to norms from all age groups combined.

The hypothesis of Costa and McCrae (2002) that the rate of change would be higher in early than in later adulthood was partially supported. Curvilinear regression on the continuous data showed significant quadratic effects for N and C in both Form S and Form R data and for A in Form R data. As Figure 6 depicts, N declined and C and A increased cross-sectionally more rapidly for younger respondents. However, only linear effects were significant for E and O, and for A in self-reports. Details of the regressions are reported in Table 3, which shows that the linear effects (first line for each entry, first data column) ranged from  $-1.16$  *T*-score points per decade for observer-rated O to  $2.17$  *T*-score points per decade for self-reported A. These rates of change are consistent with previous results in several cultures (McCrae, Costa, Hřebíčková, et al., 2004).

Figures 7 through 11 present developmental curves for the 30 facet scales for Form R data. (Form S data showed very similar curves.) Where significant quadratic effects were found, the quadratic regression is plotted; otherwise, the linear regression is plotted.<sup>5</sup> In general, facets followed the trend of the domain to which they are assigned, but

**FIGURE 6**  
**Estimated Age Curves From Regressions**  
**Predicting Self-Reported and Observer-Rated**  
**Neuroticism and Conscientiousness**  
**Observer-Rated Agreeableness Domain**  
**Scores From Age and Age Squared**



considerable variation was found in trajectories in the O and especially E domains. O1: Fantasy shows a steep curvilinear decline, whereas O5: Ideas does not decline with age. Among E facets, E1: Warmth increases (perhaps because it has a secondary loading on A), E3: Assertiveness has a curvilinear shape peaking in middle age (cf. Helson & Kwan, 2000), and E5: Excitement-Seeking declines dramatically (cf. Zuckerman, 1979). Similar effects have been reported for E facets in longitudinal hierarchical linear modeling analyses (Costa, McCrae, & Terracciano, 2004). In general, age trends for the facets resemble those found in other cultures (McCrae & Costa, in press). Thus, the NEO-PI-3 facet scales appear to show the same pancultural developmental trends as the NEO-PI-R facet scales.

## DISCUSSION

This study is the first to compare adolescents and adults on a revised personality inventory designed to be readable across this full age span. Study 1 showed that the NEO-PI-3 functioned as well in adults as it did in adolescents (McCrae et al., 2005). Study 2 showed familiar age trends suggesting that O peaks about age 19 and demonstrated that individuals in the decade of the 20s are, as expected, intermediate between adolescents and adults in the mean levels of their personality traits.

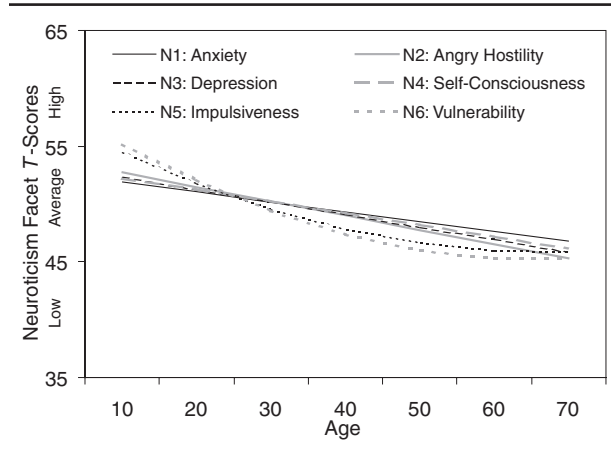
The chief purpose of creating the NEO-PI-3 was to make it useable for respondents with a wider range of

**TABLE 3**  
**Regressions Predicting Form S and Form R NEO Personality Inventory—3 Domain T-Scores**  
**From Age in Decades and Age Squared**

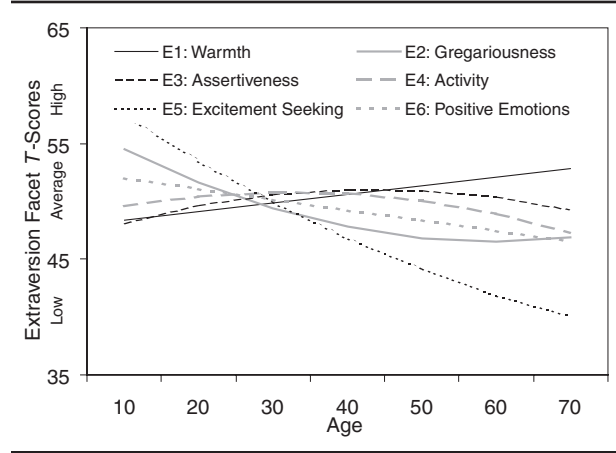
Domain	Age in Decades			Adjusted R <sup>2</sup>	Age Squared		
	B	β	t		B	β	t
<b>Form S (self-reports)</b>							
Neuroticism	-1.25 (0.16)	-.22	-7.63***	.048			
	-3.08 (0.84)	-.54	-3.67***	.051	0.23 (0.10)	.33	2.22*
Extraversion	-1.92 (0.16)	-.34	-12.12***	.114			
	-3.05 (0.81)	-.55	-3.76***	.115	0.14 (0.10)	.20	1.42
Openness	-1.54 (0.16)	-.27	-9.51***	.073			
	-1.90 (0.83)	-.33	-2.28*	.072	0.04 (0.10)	.06	0.44
Agreeableness	2.17 (0.16)	.38	13.92***	.145			
	3.57 (0.80)	.63	4.48***	.147	-0.17 (0.10)	-.25	-1.79
Conscientiousness	1.31 (0.16)	.23	7.99***	.053			
	5.71 (0.83)	1.01	6.88***	.075	-0.54 (0.10)	-.79	-5.40***
<b>Form R (observer ratings)</b>							
Neuroticism	-1.59 (0.16)	-.28	-9.71***	.077			
	-3.33 (0.82)	-.59	-4.06***	.080	0.21 (0.10)	.31	2.17*
Extraversion	-1.19 (0.17)	-.21	-7.16***	.043			
	-1.53 (0.84)	-.27	-1.82	.043	0.04 (0.10)	.06	0.41
Openness	-1.16 (0.17)	-.21	-6.98***	.041			
	-1.43 (0.84)	-.25	-1.70	.040	0.03 (0.10)	.05	0.32
Agreeableness	1.89 (0.16)	.33	11.79***	.110			
	4.13 (0.81)	.73	5.13***	.116	-0.28 (0.10)	-.40	-2.84**
Conscientiousness	1.95 (0.16)	.34	12.24***	.118			
	7.08 (0.79)	1.25	8.97***	.151	-0.63 (0.10)	-.92	-6.63***

NOTE: For each entry, the first line reports the linear regression and the second the quadratic. Standard errors of Bs are given in parentheses. For Form S, *N* = 1,135; for Form R, *N* = 1,114.  
 \**p* < .05. \*\**p* < .01. \*\*\**p* < .001.

**FIGURE 7**  
**Estimated Age Curves From Regressions**  
**Predicting Observer-Rated Neuroticism Facet**  
**Scores From Age and Age Squared**



**FIGURE 8**  
**Estimated Age Curves From Regressions**  
**Predicting Observer-Rated Extraversion Facet**  
**Scores From Age and Age Squared**



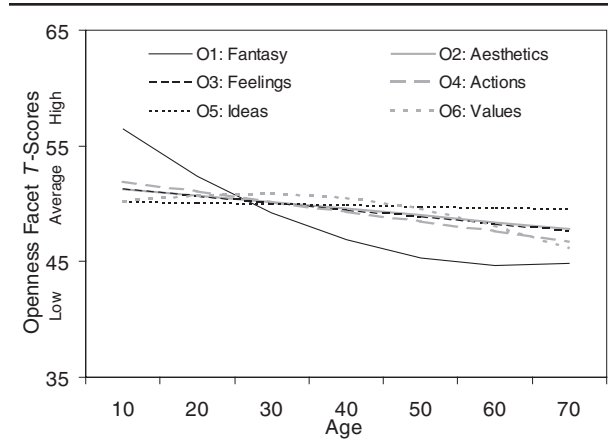
reading skills. In the process, the psychometric properties of the instrument were improved modestly. The most notable improvements were for N4: Self-Consciousness, A6: Tender-Mindedness, and C2: Order, which showed size-

able increases in both internal consistency and cross-observer agreement.

The case of O4: Actions requires comment. Modifications to the NEO-PI-R led to small improvements in inter-

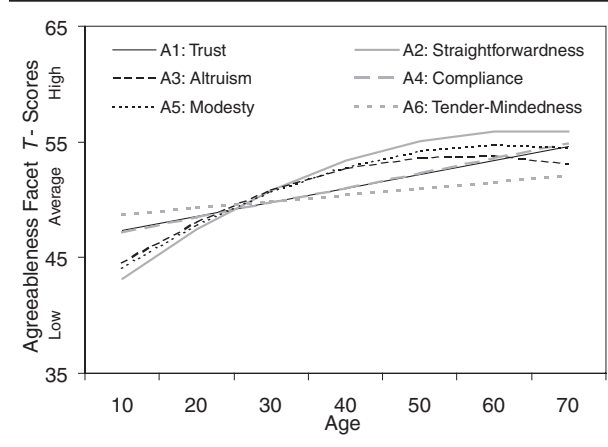


**FIGURE 9**  
**Estimated Age Curves From Regressions Predicting Observer-Rated Openness Facet Scores From Age and Age Squared**



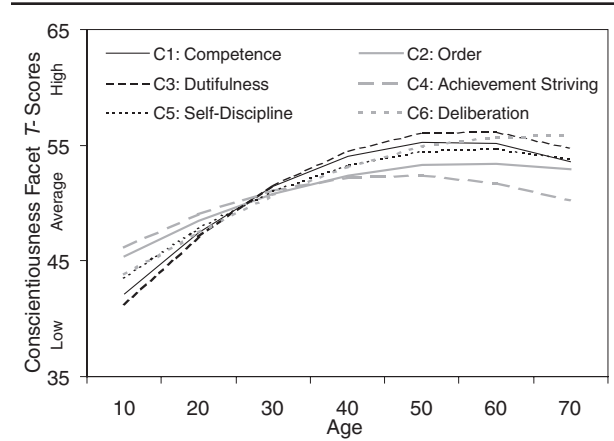
NOTE: The regression for O5 (ideas) is nonsignificant.

**FIGURE 10**  
**Estimated Age Curves From Regressions Predicting Observer-Rated Agreeableness Facet Scores From Age and Age Squared**



nal consistency, especially in Form R, but the NEO-PI-3 Form S value of .54 is still below the conventional standards for an internally consistent scale. However, there is a good deal of evidence that O4 validly assesses its intended construct. In the present study, it shows cross-observer agreement ( $r = .42$ ) and loads strongly (.54) on the O factor. Other studies have shown that the NEO-PI-R O facet is longitudinally stable (Costa et al., 2000) and heritable (Jang, McCrae, Angleitner, Riemann, & Livesley, 1998) at levels comparable to other NEO-PI-R facets, and O4: Actions has also demonstrated discriminant validity (Costa & McCrae, 1995; McCrae & Costa, 1992). In this instance,

**FIGURE 11**  
**Estimated Age Curves From Regressions Predicting Observer-Rated Conscientiousness Facet Scores From Age and Age Squared**



relatively poor internal consistency does not appear to diminish the validity or utility of the scale.

In Study 2, the NEO-PI-3 was used to refine our understanding of developmental trends. Individuals aged 18, 19, and 20 were oversampled with the expectation that change would be most pronounced in these years (cf. Robins et al., 2001). However, Figures 1 to 5 show that in the present study, little change occurs in these years; the period of transition appears instead to be in the decade of the 20s. Given that the rate of change of most personality traits is less than 2 *T*-score points per decade, much larger sample sizes are probably needed to estimate precisely when change occurs.

**The Interpretation and Use of Age Norms**

We have proposed NEO-PI-3 norms for different age and gender groups. The effect of using gender-specific norms is to eliminate gender differences, which might be appropriate, for example, if a researcher wished to understand the influence of personality net of gender effects. However, gender-specific norms cannot be used in hiring or promotion decisions, so combined gender norms are also needed.

Although the rate of change in personality trait mean levels is slow, the cumulative effect during a lifetime approaches one full standard deviation, so the issue of age norms must also be considered. We have suggested one to three age groupings, although it seems likely that published versions of the NEO-PI-3 will report norms only for adolescents and adults 21 and older. That choice would make NEO-PI-3 data more comparable to NEO-PI-R data;

the major innovation would be the demonstrated applicability of adolescent norms to respondents as young as 14.

The consequence of using a single set of adult norms is that respondents aged 21 to 30 will score a few *T*-score points higher on N, E, and O and a few points lower on A and C than they would if early adult norms were used (and the reverse for adults older than age 30). It may seem that this introduces a small distortion into the scores of all adults, but an argument can be made that it does not.

Norming test scores serves two fundamentally different purposes. The first is to make different tests comparable. Raw scores on the NEO-PI-3 N scale could not be compared to raw scores from the Eysenck Personality Questionnaire (Eysenck & Eysenck, 1975) N scale, but if both were converted to *T*-scores using comparable normative samples, meaningful comparisons could be made. Similarly, if the NEO-PI-3 functioned differently in young adults and in older adults—if raw scales failed to show scalar equivalence (Van de Vijver & Leung, 1997) across age groups—then age norming would be required.

But there is no evidence that the NEO-PI-3 functions differently in these two age groups. Adults in their 20s score higher on N than older adults because they are, in fact, higher in N, as is seen in the higher incidence of N-related psychopathology such as borderline personality disorder (American Psychiatric Association, 1994). Adults in their 20s score lower on A and C because they, in fact, are less agreeable and conscientious, as is seen in the higher rates of criminal behavior in this age group (e.g., murder rates; see Federal Bureau of Investigation, 2003). From this perspective, a single set of adult age norms provides a more accurate depiction of personality than the use of more differentiated age norms would. By extension, the same could be said for using a single set of norms for both adolescents and adults. That practice would preserve the age differences shown in Figures 1 to 5 and would be a reminder to clinicians and others dealing with adolescents that they tend to be high in N, E, and O and low in A and C.

There is, however, a second legitimate use for norms, which is to describe an individual relative to a known group. For example, counselors who work exclusively with adolescents may already have a good sense of what adolescent personality is like, and it may be more useful to them to see scores that reflect a client's standing relative to that reference group: Is the client even less conscientious than the average teenager? Such a view of the data may suggest the need for an intervention that might be ignored if low C were regarded simply as a normal attribute of adolescents. Particularly for those accustomed to their use, age norms may be helpful in understanding personality scores.

## Notes

1. The NEO manual (Costa & McCrae, 1992) specifies that a protocol is invalid if more than 40 items are missing. Because we included an additional 96 items in the test booklet, we increased that value proportionately to 56.

2. The item/facet correlations for one N5 (impulsiveness) item declined from .42 to .17 after revision, but reading level changed from 9.3 to 2.1. The six other items that did not increase in item/facet correlations after revision showed modest decreases ranging from .02 to .10.

3. Only one interaction was replicated across forms: E5 (excitement seeking) declined faster among women after age 30. This effect accounted for less than 2% of the variance.

4. Means and standard deviations for all the normative groups discussed in this article are available from the first author.

5. These figures show many crossovers about age 30; this phenomenon occurs simply because the data for all scales were standardized in this sample where the mean age is 31.8 years. To attain a grand mean of *T* = 50, if people younger than age 30 score higher than 50 on a trait, people older than age 30 must score lower, and vice versa.

## References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Baker, S. R., & Victor, J. B. (2003, August). *Adolescent self-reports of personality and temperament: NEO-PI-R and TPQ*. Paper presented at the XIth European Conference on Developmental Psychology, Milan, Italy.
- Cohen, R. J., & Swerdlik, M. E. (2002). *Psychological testing and assessment: An introduction to tests and measurement* (5th ed.). Boston: McGraw Hill.
- Costa, P. T., Jr., Herbst, J. H., McCrae, R. R., & Siegler, I. C. (2000). Personality at midlife: Stability, intrinsic maturation, and response to life events. *Assessment, 7*, 365-378.
- Costa, P. T., Jr., & McCrae, R. R. (1986). Cross-sectional studies of personality in a national sample: 1. Development and validation of survey measures. *Psychology and Aging, 1*, 140-143.
- Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T., Jr., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment, 64*, 21-50.
- Costa, P. T., Jr., & McCrae, R. R. (2002). Looking backward: Changes in the mean levels of personality traits from 80 to 12. In D. Cervone & W. Mischel (Eds.), *Advances in personality science* (pp. 219-237). New York: Guilford.
- Costa, P. T., Jr., McCrae, R. R., & Terracciano, A. (2004, August). *Longitudinal changes in extraversion and its facets: HLM analyses*. Paper presented at the Annual Convention of the American Psychological Association, Honolulu, HI.
- Costa, P. T., Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81*, 322-331.
- Cota, A. A., Longman, R., Stewart, R., Holden, R. R., & Fekken, G. C. (1993). Comparing different methods for implementing parallel analysis: A practical index of accuracy. *Educational and Psychological Measurement, 53*, 865-876.
- De Fruyt, F., Mervielde, I., Hoekstra, H. A., & Rolland, J.-P. (2000). Assessing adolescents' personality with the NEO-PI-R. *Assessment, 7*, 329-345.

- Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. San Diego, CA: EdITS.
- Federal Bureau of Investigation. (2003). *Murder offenders by age, sex, and race, 2003*. Retrieved December 4, 2004, from www.fbi.gov/ucr/cius\_03/xl/03tbl2-5.xls
- Helson, R., & Kwan, V. S. Y. (2000). Personality development in adulthood: The broad picture and processes in one longitudinal sample. In S. E. Hampson (Ed.), *Advances in personality psychology* (pp. 77-106). Philadelphia: Taylor & Francis.
- Jang, K. L., McCrae, R. R., Angleitner, A., Riemann, R., & Livesley, W. J. (1998). Heritability of facet-level traits in a cross-cultural twin sample: Support for a hierarchical model of personality. *Journal of Personality and Social Psychology, 74*, 1556-1565.
- McCrae, R. R., & Costa, P. T., Jr. (1992). Discriminant validity of NEO-PI-R facets. *Educational and Psychological Measurement, 52*, 229-237.
- McCrae, R. R., & Costa, P. T., Jr. (2003). *Personality in adulthood: A five-factor theory perspective* (2nd ed.). New York: Guilford.
- McCrae, R. R., & Costa, P. T., Jr. (in press). Cross-cultural perspectives on adult personality trait development. In D. Mroczek & T. Little (Eds.), *Handbook of personality development*. Hillsdale, NJ: Lawrence Erlbaum.
- McCrae, R. R., Costa, P. T., Jr., Hřebíčková, M., Urbánek, T., Martin, T. A., Oryol, V. E., et al. (2004). Age differences in personality traits across cultures: Self-report and observer perspectives. *European Journal of Personality, 18*, 143-157.
- McCrae, R. R., Costa P. T., Jr., & Martin, T. A. (2005). The NEO-PI-3: A more readable Revised NEO Personality Inventory. *Journal of Personality Assessment, 84*, 261-270.
- McCrae, R. R., Costa, P. T., Jr., Martin, T. A., Oryol, V. E., Rukavishnikov, A. A., Senin, I. G., et al. (2004). Consensual validation of personality traits across cultures. *Journal of Research in Personality, 38*, 179-201.
- McCrae, R. R., Costa, P. T., Jr., Terracciano, A., Parker, W. D., Mills, C. J., De Fruyt, F., et al. (2002). Personality trait development from 12 to 18: Longitudinal, cross-sectional, and cross-cultural analyses. *Journal of Personality and Social Psychology, 83*, 1456-1468.
- McCrae, R. R., Zonderman, A. B., Costa, P. T., Jr., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology, 70*, 552-566.
- Mroczek, D. K., & Spiro, A., III. (2003). Modeling intraindividual change in personality traits: Findings from the Normative Aging Study. *Journal of Gerontology: Psychological Sciences, 58B*, P153-P165.
- Neyer, F. J., & Asendorpf, J. B. (2001). Personality-relationship transaction in young adulthood. *Journal of Personality and Social Psychology, 81*, 1190-1204.
- Roberts, B. W., Walton, K. E., & Viechtbauer, W. (in press). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*.
- Robins, R. W., Fraley, R. C., Roberts, B. W., & Trzesniewski, K. H. (2001). A longitudinal study of personality change in young adulthood. *Journal of Personality, 69*, 617-640.
- Van de Vijver, F. J. R., & Leung, K. (1997). Methods and data analysis of comparative research. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology: Vol 1: Theory and method* (pp. 257-300). Boston: Allyn & Bacon.
- Zuckerman, M. (1979). *Sensation seeking: Beyond the optimal level of arousal*. Hillsdale, NJ: Lawrence Erlbaum.

**Robert R. McCrae, Ph.D.**, is a research psychologist in the Intramural Research Program of the National Institute on Aging. Since 1976, he has conducted research on personality structure, assessment, and development. With Paul T. Costa, Jr., he is the author of *Personality in Adulthood: A Five-Factor Theory Perspective*.

**Thomas A. Martin, Ph.D.**, is a professor of psychology at Susquehanna University. He is a clinical psychologist whose interests include test development and personality assessment, and he has conducted extensive research in Russia with colleagues at Yaroslavl State University.

**Paul T. Costa, Jr., Ph.D.**, is chief of the Laboratory of Personality and Cognition, National Institute on Aging, and a professor of psychiatry at the Johns Hopkins University School of Medicine. His research interests include health psychology, personality disorders, and the molecular genetics of personality.