Faculty Publications, Department of Psychology   Psychology, Department of

1991

Trait Psychology Comes of Age

Paul T. Costa Jr.
National Institute on Aging, Baltimore

Robert McCrae
National Institute on Aging, Baltimore

Trait Psychology
Comes of Age

Paul T. Costa, Jr., and
Robert R. McCrae
National Institute on Aging,
Baltimore

It is sometimes thought that the relation between research on aging
and the basic sciences is one-sided: gerontologists take the methods
and theories of their specialty and apply them to aging populations,
but they rarely initiate theories or report findings that could reshape
the basic disciplines themselves. Whatever the truth of this percep-
tion in general, it is completely false with regard to the psychology
of personality. When Eichorn, Clausen, Haan, Honzik, and Mussen
(1981) published their summary of the Berkeley longitudinal studies,
Sears and Sears (1982) heralded it as "probably the most important
unified research contribution to adult social and personality psy-
chology of the last three decades" (p. 927). According to White
(1964), personality psychology *is* the study of lives, and aging is the
universal dimension along which lives are led. Renewed attention
to the life narrative (McAdams, 1990) and to the adult outcomes of
childhood temperament (Caspi, Elder, & Bem, 1987) show the cen-
tral role that studies of aging must have in many different ap-
proaches to the psychology of personality.

Our focus will be on one branch of personality psychology—dif-
ferential or trait psychology—and what we have learned about it in
the past 15 years. Discoveries about personality structure and stabil-
ity have reshaped both theories of personality and theories of adult

development in ways that few would have predicted a few years ago. Thanks in considerable measure to the findings of longitudinal studies of aging men and women, trait psychology itself has come of age.

Trait psychology has a long history, but it has rarely inspired the partisan loyalties that psychoanalysis, behaviorism, and humanistic psychology did. When, after a decade of research on response sets, Mischel (1968) suggested that traits were of little importance in explaining human behavior and might well be considered cognitive fictions, many personality psychologists seemed to shrug their shoulders and move on to different approaches.

Longitudinal personality researchers were understandably reluctant to do this: they had already invested years in administering and readministering personality questionnaires to find out what happens to personality traits over time. Beginning about 15 years ago, they began to look at their results, and what they found was astonishing: personality traits showed a degree of long-term stability that no one had expected. If traits were mere fictions, or were epiphenomena of response sets and styles, or reflected appraisals of those around us, we would expect little evidence of stability over 10 or 20 years. The retest correlations reported by Block (1977), Leon, Gillum, Gillum, and Gouze (1979), and Siegler, George, and Okun (1979) ranged from .5 to .8—far higher that most of us had expected.

Our own experience with two longitudinal studies—the Normative Aging Study (NAS), where we analyzed the Sixteen Personality Factor Questionnaire (16PF; Cattell, Eber, & Tatsuoka, 1970), and the Baltimore Longitudinal Study of Aging (BLSA; Shock et al., 1984), where we analyzed the Guilford-Zimmerman Temperament Survey (GZTS; Guilford, Zimmerman, & Guilford, 1976) and other instruments—quickly convinced us of two facts: Personality traits are real, and they are remarkably stable in adulthood (Costa & McCrae, 1980; McCrae & Costa, 1990). Stability is seen in two ways: the mean levels of most traits change little with age, and retest correlations show stability of individual differences.

Evidence of stability was more welcomed by trait psychologists than by gerontologists. Our findings contradicted popular theories that called for stages of adult development, a midlife crisis, or disengagement or that predicted increasing rigidity or wisdom, mellowing or crankiness, depression or denial with approaching old age.

What we and other longitudinal researchers found was that most people remained pretty much the same as they aged. This conclusion was radical ten years ago, and it is by no means universally conceded now. But most experts in the field of personality and aging would agree that it is a valid generalization and take issue chiefly with how far and in what ways it can be extended. One of the purposes of this chapter will be to provide new data that extend the finding both in time and across methods of assessing personality.

The Structure and Assessment of Personality

THE FIVE-FACTOR MODEL

Before turning to those data, however, we wish to address briefly the other major development in the field of personality, the five-factor model of personality structure. Trait models of personality all suggest that individuals have pervasive and enduring characteristics that influence their thoughts, feelings, and behaviors and that distinguish them from other people. What distinguished the systems of Allport, Cattell, Eysenck, and Buss and Plomin from each other and from the systems of dozens of other trait psychologists was the particular traits thought to be important. Personality psychologists created hundreds of scales to measure traits, and everyday language includes thousands of trait adjectives like *nervous, enthusiastic, original, accommodating,* and *careful* that have evolved over the centuries to denote characteristics of personality.

The central problem of trait psychology was how to handle this extraordinary diversity of trait constructs, and the solution was to organize them in terms of broad and pervasive factors or dimensions. The particular set of factors we now recognize as basic dimensions of personality were first articulated by Tupes and Christal in 1961, but their five-factor model was not widely accepted until quite recently (Digman, 1990; John, 1990; Wiggins & Trapnell, in press).

The story of how these many trait terms and scales were boiled down to five fundamental factors has been told elsewhere (Digman, 1990; John, Angleitner, & Ostendorf, 1988). Here we need only note that there was tremendous redundancy in both natural and technical languages and that the five dimensions of Neuroticism (N), Ex-

traversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C) summarize the recurring themes very well. These five factors have been found in children, college students, and older adults; in men and women; in self-reports and observer ratings; and in American, German, and Chinese samples.

Our own research on this topic began in attempts to organize the longitudinal personality data at the NAS. We grouped scales at first into three clusters or *domains*, now recognizable as N, E, and O (Costa & McCrae, 1976, 1977). In the process, we became dissatisfied with existing instruments, and we thought a better inventory might be constructed by subdividing these three domains into more discrete *facets* and creating scales for each. The result was the NEO Inventory, a questionnaire measure that we introduced into the BLSA. When, in 1983, we became convinced that A and C were also important domains of personality, we included scales to measure them in an instrument we published as the NEO Personality Inventory (NEO-PI; Costa & McCrae, 1985b, 1989a). In 1990 we developed facet scales for A and C for the forthcoming Revised NEO-PI or NEO-PI-R. We will present some data in this chapter on characteristics of the NEO-PI-R.

Research on aging did not directly contribute to the discovery of the five-factor model, but it did contribute substantially to demonstrating its comprehensiveness (e.g., Conley, 1985). One of the seldom-recognized advantages of a longitudinal study is that it allows researchers to accumulate large archives of data on the same individuals (Costa & McCrae, in press). Over the years we have administered many personality inventories to BLSA volunteers (see Table 1), and joint factor analyses of these instruments with the NEO-PI provided persuasive evidence for the generality of the five-factor model (McCrae, 1989).

We cannot review all that evidence here, but one example might be useful. This is nominally a symposium on motivation, and motives are often conceived of as traits—pervasive and characteristic needs to achieve, aggress, affiliate, and so on. The most influential catalog of needs was offered by Murray (1938), and one of the best instruments for assessing these needs is Jackson's (1984) Personality Research Form (PRF). In a joint factor analysis of the NEO-PI and the PRF (Costa & McCrae, 1988a), we showed that five factors could be recovered, each marked by one of the NEO-PI factors. Every PRF

Table 1

Instruments Administered to BLSA-Related Samples

Instrument	Date of Administration
Guilford-Zimmerman Temperament Survey (GZTS; Guilford, Zimmerman, & Guilford, 1976)	1960–1990[a]
Profile of Mood States (POMS; McNair, Lorr, & Droppleman, 1971)	1979
Self-Directed Search (SDS; Holland, 1985)	1981
MMPI Personality Disorder Scales (Morey, Waugh, & Blashfield, 1985)	1981–1987
State-Trait Personality Inventory (STPI; Spielberger et al., 1979)	1983
Revised Interpersonal Adjective Scales (IASR; Wiggins, Trapnell, & Phillips, 1988)	1985
Sensation Seeking Scales V (SSS; Zuckerman, 1979)	1985
Myers-Briggs Type Indicator (MBTi; Myers & McCaulley, 1985)	1986
Buss-Durkee Hostility Inventory (BDHI; Buss & Durkee, 1957)	1986
Personality Research Form (PRF; Jackson, 1984)	1986
Revised California Psychological Inventory (CPI; Gough, 1987)	1987
Interpersonal Style Inventory (ISI; Lorr, 1986)	1990

[a]For analyses reported in Table 3, GZTS data are from the period 1960–1978.

scale loaded on one or more of the factors, and all the factors were defined by two or more PRF scales. N was marked by needs for succorance and defendence; E by needs for affiliation and exhibition; O by needs for change and understanding; A by needs for nurturance and abasement; C by needs for order and achievement. The dimensions underlying Murray's needs seem to be the same five found in analyses of trait adjectives and personality scales.

MORE SPECIFIC TRAITS

What are these five factors? One way to understand them is to enumerate the traits that define them. Goldberg (1990), Ostendorf (1990), and others have done this with regard to trait adjectives found in natural languages. We have defined the factors in terms of constructs drawn from the psychological literature that became the basis for NEO-PI-R facet scales. Six facets have been developed for each of the five domains in the NEO-PI-R; these facets are listed in Table 2. It is clear that all five domains are very broad constructs. Chronic anxiety, hostility, depression, and many other variables that

Table 2

Descriptive Statistics and Factor Loadings for Form S NEO-PI-R Scales

NEO-PI-R	Men		Women		Varimax-Rotated Factor				
Scale	Mean	SD	Mean	SD	N	E	O	A	C
Domain									
Neuroticism	75.2	19.9	83.1[b]	21.7					
Extraversion	108.5	18.5	110.3	18.4					
Openness	110.1	17.5	111.0	17.2					
Agreeableness	120.1	16.1	128.5[b]	14.4					
Conscientiousness	123.6	17.4	122.7	17.8					
N facets									
Anxiety	13.3	4.9	15.4[b]	5.4	**.81**	.02	−.01	−.01	−.10
Hostility	12.2	4.5	12.6	4.8	**.63**	−.03	.01	**−.48**	−.08
Depression	11.6	5.2	12.9[b]	5.6	**.80**	−.10	.02	−.03	−.26
Self–Consciousness	13.7	4.3	15.0[b]	4.5	**.73**	−.18	−.09	.04	−.16
Impulsiveness	15.3	4.2	16.3[b]	4.6	**.49**	.35	.02	−.21	−.32
Vulnerability	9.2	3.7	10.9[b]	4.0	**.70**	−.15	−.09	.04	−.38
E facets									
Warmth	22.3	4.0	23.6[b]	3.8	−.12	**.66**	.18	.38	.13
Gregariousness	16.0	4.9	17.0[b]	4.7	−.18	**.66**	.04	.07	−.03
Assertiveness	16.3	4.7	15.4[a]	4.8	−.32	**.44**	.23	−.32	.32
Activity	17.3	4.3	17.8[a]	4.4	.04	**.54**	.16	−.27	**.42**
Excitement Seeking	17.2	4.7	15.7[b]	5.1	.00	**.58**	.11	−.38	−.06
Positive Emotions	19.5	4.3	20.8[b]	4.5	−.04	**.74**	.19	.10	.10
O facets									
Fantasy	17.0	4.7	16.2[a]	5.0	.18	.18	**.58**	−.14	−.31
Aesthetics	16.7	5.4	18.5[b]	5.1	.14	.04	**.73**	.17	.14
Feelings	19.7	3.8	20.8[b]	4.1	.37	**.41**	**.50**	−.01	.12
Actions	16.1	3.8	16.8[a]	3.6	−.19	.22	**.57**	.04	−.04
Ideas	19.8	5.0	18.2[b]	5.0	−.15	−.01	**.75**	−.09	.16
Values	20.8	4.5	20.5	3.8	−.13	.08	**.49**	−.07	−.15
A facets									
Trust	20.9	4.3	21.7[a]	4.0	−.35	.22	.15	**.56**	.03
Straightforwardness	20.3	4.3	22.2[b]	4.3	−.03	−.15	−.11	**.68**	.24
Altruism	22.8	3.6	24.3[b]	3.2	−.06	**.52**	−.05	**.55**	.27
Compliance	18.1	3.7	19.6[b]	4.1	−.16	−.08	.00	**.77**	.01
Modesty	18.1	4.4	19.7[b]	3.8	.19	−.12	−.18	**.59**	−.08
Tender–Mindedness	19.9	3.8	21.0[b]	3.1	.04	.27	.13	**.62**	.00
C facets									
Competence	22.5	3.5	21.8[b]	3.5	**−.41**	.17	.13	.03	**.64**
Order	18.9	4.1	19.1	4.2	−.04	.06	−.19	.01	**.70**
Dutifulness	23.2	3.9	23.2	3.8	−.20	−.04	.01	.29	**.68**
Achievement Striving	19.3	4.1	19.6	3.9	−.09	.23	.15	−.13	**.74**
Self–Discipline	21.8	4.2	21.7	4.4	−.33	.17	−.08	.06	**.75**
Deliberation	17.8	4.0	17.3	4.3	−.23	−.28	−.04	.22	**.57**

Note. N = 500 men, 500 women. Factor loadings > ± .40 are given in boldface.
[a]Difference between men and women significant at $p < .05$
[b]Difference between men and women significant at $p < .001$

refer to psychopathology or maladjustment are included in N. E includes not only interpersonal traits like gregariousness and assertiveness, but also temperamental traits such as activity level and positive emotionality. O can be seen in an active fantasy life and in differentiated feelings, but also in liberal values and intellectual curiosity. A includes both interpersonal traits like compliance and attitudinal traits like tender-mindedness, and C is defined not only by the constraining facets of orderliness and deliberation, but also by the proactive facets of self-discipline and need for achievement.

The data in this table come from three subsamples: 405 men and women from the BLSA who completed the NEO-PI in 1986, and supplemental items to measure A and C in 1990, as part of mailed questionnaire batteries; 329 men and women who completed the NEO-PI-R by computer administration between 1989 and 1991; and 1,520 men and women who took the NEO-PI-R as part of a job performance study (Costa, McCrae, & Dye, 1991). Despite the differences in samples, times, and conditions of administration, the three samples were generally comparable in personality scores. When the three samples were contrasted on the five domains within sex, there was only one notable difference: men in the job performance study were just over one-half standard deviation higher in E than BLSA men originally tested in 1986.

Previous adult norms for the NEO-PI have been based on responses from BLSA subjects and their spouses and peers. This sample is almost entirely white and exceptionally well educated, and it overrepresents the older segment of the population. To form a more representative group, 500 men and 500 women were selected from the three samples to match U.S. census projections for 1995 in the distribution of age and race groups. By selecting subjects with fewer years of education from the job performance sample, the average level of education was also reduced. These subjects ranged in age from 21 to 96.

A factor analysis (see Table 2) makes it is clear that the facet scales fit the hypothesized structure rather well: all have their highest loading on the intended factor, and the secondary loadings, such as the negative loading of Hostility on the A factor, make conceptual sense. But do the individual facet scales have *differential* validity? Does Anxiety measure a different kind of N than Depression? Do Trust and Altruism have different, if related, correlates?

Table 3

Correlates of Form S Revised NEO Personality Inventory Facet Scales

Neuroticism Facets		Extraversion Facets		Openness Facets		Agreeableness Facets		Conscientiousness Facets	
Correlate	*r*	Correlate	*r*	Correlate	*r*	Correlate	*r*	Correlate	*r*
Anxiety		*Warmth*		*Fantasy*		*Trust*		*Competence*	
ISI Stable	-.77	IASR Aloof-Introverted	-.68	MBTI Intuition	.43	ISI Trusting	.68	CPI Ach. via Conformance	.56
GZTS Emotional Stability	-.60	PRF Affiliation	.64	CPI Flexibility	.42	CPI Tolerance	.48	CPI Independence	.49
STPI Anxiety	.55	MBTI Introversion	-.61	MBTI Perception	.38	BDHI Suspicion	-.46	ISI Orderly	.43
POMS Tension	.54	ISI Sociable	.57	PRF Sentience	.36	ISI Conscientious	.40	POMS Depression	-.43
MMPI Compulsive	.51	GZTS Sociability	.54	MMPI Borderline	.34	GZTS Personal Relations	.39	IASR Assured-Dominant	.42
Hostility		*Gregariousness*		*Aesthetics*		*Straightforwardness*		*Order*	
BDHI Irritability	.65	ISI Sociable	.75	SDS Artistic	.56	IASR Arrogant-Calculatng	-.53	PRF Order	.71
ISI Tolerant	-.65	MMPI Schizoid	-.66	MBTI Intuition	.56	CPI Self-Control	.46	ISI Orderly	.68
PRF Aggression	.62	PRF Affiliation	.64	PRF Sentience	.54	MMPI Anisocial	-.44	MBTI Perception	-.44
STPI Hostility	.60	MBTI Introversion	-.59	PRF Understanding	.45	PRF Aggression	-.41	ISI Persistent	.43
MMPI Borderline	.47	GZTS Sociability	.57	CPI Empathy	.36	ISI Conscientious	.39	MMPI Passive-Aggressive	-.40
Depression		*Assertiveness*		*Feelings*		*Altruism*		*Dutifulness*	
STPI Anxiety	.68	ISI Directive	.73	PRF Sentience	.44	IASR Warm-Agreeable	.60	CPI Ach. via Conformance	.43
ISI Stable	-.61	IASR Assured-Dominant	.69	CPI Empathy	.42	IASR Aloof-Introverted	-.48	MMPI Antisocial	-.41
POMS Depression	.58	CPI Dominance	.69	CPI Self-Acceptance	.39	ISI Nurturant	.42	ISI Approval Seeking	.40
BDHI Irritability	.55	PRF Dominance	.64	GZTS Ascendance	.37	POMS Friendliness	.34	CPI Good Impression	.39
MMPI Dependent	.52	MBTI Introversion	-.59	MBTI Feeling	.33	CPI Femininity	.34	PRF Endurance	.37

Self-Consciousness
Scale	r
MMPI Avoidant	.58
STPI Anxiety	.55
CPI Independence	-.51
IASR Assured-Dominant	-.50
ISI Stable	-.50

Impulsiveness
Scale	r
CPI Self-Control	-.46
GZTS Restraint	-.43
MMPI Borderline	.43
BDHI Indirect Hostility	.43
ISI Deliberate	-.42

Vulnerability
Scale	r
STPI Anxiety	.62
ISI Stable	-.58
MMPI Dependent	.52
POMS Tension	.51
IASR Assured-Dominant	-.50

Activity
Scale	r
GZTS General Activity	.67
MBTI Introversion	-.42
IASR Unassured-Submissive	-.41
IASR Gregarious-Extraverted	.41
CPI Self-Acceptance	.39

Excitement Seeking
Scale	r
PRF Harm Avoidance	-.46
ISI Deliberate	-.43
CPI Self-Control	-.42
GZTS General Activity	.41
MMPI Antisocial	.40

Positive Emotions
Scale	r
IASR Gregarious-Extraverted	.58
MBTI Introversion	-.51
PRF Play	.50
GZTS General Activity	.49
CPI Sociability	.48

Actions
Scale	r
PRF Change	.56
MBTI Intuition	.53
SSS Experience Seeking	.43
CPI Flexibility	.37
MMPI Histrionic	.37

Ideas
Scale	r
PRF Understanding	.67
MBTI Intuition	.56
GZTS Thoughtfulness	.46
STPI Curiosity	.45
SDS Investigative	.43

Values
Scale	r
SSS Experience Seeking	.46
ISI Rule Free	.46
CPI Ach. via Independence	.43
CPI Flexibility	.41
MBTI Intuition	.39

Compliance
Scale	r
BDHI Verbal Hostility	-.63
ISI Tolerant	.63
PRF Aggression	-.60
GZTS Friendliness	.51
CPI Self-Control	.43

Modesty
Scale	r
IASR Unassuming-Ingenuous	.36
PRF Exhibition	-.35
ISI Directive	-.33
PRF Abasement	.32
CPI Social Presence	-.31

Tender-Mindedness
Scale	r
ISI Nurturant	.45
IASR Cold-hearted	-.44
MBTI Feeling	.39
SDS Social	.36
PRF Nurturance	.35

Achievement Striving
Scale	r
PRF Achievement	.59
IASR Assured-Dominant	.52
CPI Ach. via Conformance	.49
GZTS General Activity	.47
ISI Persistent	.47

Self-Discipline
Scale	r
ISI Orderly	.58
CPI Ach. via Conformance	.57
ISI Persistent	.51
PRF Order	.47
MMPI Passive-Aggressive	-.44

Deliberation
Scale	r
ISI Deliberate	.64
PRF Cognitive Structure	.43
ISI Orderly	.41
PRF Impulsivity	-.41
MBTI Perception	-.39

Note. For the PRF and CPI, correlations are given with NEO-PI facets and preliminary Agreeableness and Conscientiousness scales (Costa & McCrae, 1989b). All other correlations are with NEO-PI-R scales. Correlations with the GZTS are for men only; all other correlations are based on combined sex samples. All correlations are significant at $p < .001$. ISI = Interpersonal Style Inventory, $N = 115$. MBTI = Myers-Briggs Type Indicator continuous scales, $N = 388$. BDHI = Buss-Durkee Hostility Inventory, $N = 268$. GZTS = Guilford-Zimmerman Temperament Survey, $N = 131$. IASR = Revised Interpersonal Adjective Scales, $N = 221$. SDS = Self-Directed Search, $N = 208$. STPI = State-Trait Personality Inventory, $N = 246$. MMPI = Minnesota Multiphasic Personality Inventory Personality Disorder Scales, $N = 170$. POMS = Profile of Mood States "as you felt during the past week," $N = 130$. PRF = Personality Research Form, $N = 203–296$. CPI = Revised California Personality Inventory, $N = 216–348$. SSS = Sensation Seeking Scales, $N = 217$.

There are several ways to approach this question. One is to exploit the longitudinal archives by examining the correlations of the individual facets with scales from the instruments listed in Table 1 in the core sample of 405 BLSA participants and in an additional sample of BLSA peers. Together these instruments have 116 different scales (disregarding the total and validity scales), so there are 3,480 correlations with the 30 NEO-PI-R facets—an intimidating wealth of data. To summarize it, we first identified all correlations above .30 in absolute magnitude. With our sample sizes, all these correlations were significant at at least $p < .001$; by chance, we would expect only 3 or 4 correlations this large. Instead, we found 648. This is dramatic testimony to the pervasiveness of the five factors.

To further summarize, we sorted the correlations for each facet scale in descending order; Table 3 lists five of the highest correlates for each. There is some redundancy in our criteria: Instruments like the Revised California Personality Inventory (CPI) have extensive item overlap, and half the scales of the Revised Interpersonal Adjective Scale (IASR) were designed to be polar opposites of the other half. For that reason we examined the top eight correlates for each scale and selected five that seemed to give the broadest picture of the facet. Where possible, only one scale was selected from any single instrument.

There are several ways to interpret the data in Table 3. Examining each facet's correlates provides evidence of convergent validity that is generally substantial: 66 of the 150 correlations in the table are greater than .50 in absolute magnitude. Comparing the correlates of different facets within a domain speaks to the differential validity of facets. For example, Openness to Aesthetics is strongly related to Artistic vocational interests, whereas Openness to Ideas is related to Investigative interests. We can see from this table that some traits, like Assertiveness or dominance, are often measured in very similar ways. Others, like Openness to Fantasy, are less often or less clearly found in personality inventories.

However, there is also another important way to look at these data. Normally construct validity studies are conducted by administering all the relevant measures at the same time. As Table 1 shows, the intervals between the administration of the NEO-PI-R and criteria in Table 3 vary from years to decades. Consider the relation between GZTS General Activity and NEO-PI-R Activity. The interval

between administration of these two inventories ranged from 7 to 26 years, with a mean of 20 years. Yet the correlation is .67, as high as most cross-instrument validity coefficients based on data collected on the same occasion. This is impressive, albeit indirect, testimony to the stability of personality in adulthood (cf. Costa & McCrae, 1985a).

The Long-Term Stability of Personality

By now most psychologists interested in the course of adult personality have been persuaded that there is considerable continuity in most traits in most people; few still imagine that a midlife crisis or menopause or retirement brings dramatic swings in basic dispositions. The relative degree of stability is still debated, however, chiefly in the question of cumulative changes over time. True, personality may be predominantly stable from age 30 to 40, or 40 to 50, or 50 to 60—but how predictable is personality at age 60 from personality at age 30? Isn't there a steady decay that leads to increasingly small associations as the prediction interval increases?

  This is a plausible argument, and it seems to be supported by literature reviews by Schuerger, Zarrella, and Hotz (1989) and by Conley (1984), both of whom summarized their data in graphs that show progressive declines in retest correlations with increasing retest interval. However, such metanalyses are often difficult to interpret. For example, both Conley and Schuerger et al. combined adolescents with adults. There is strong evidence that personality continues to change during the interval from college age to mid-30s, so some portion of the apparent decay may be limited to changes in these younger subjects. The literature does not contain enough long-term studies of older subjects to justify a metanalysis. Further, the instruments examined vary in their reliability, again making comparisons difficult.

  Perhaps the most satisfying design would retest the same individuals several times on the same instrument. We know of only one study that does this—an analysis of Minnesota Multiphasic Personality Inventory (MMPI) clinical scales conducted by Leon, Gillum, Gillum, and Gouze (1979), who tested a sample of 71 middle-aged men in 1947 and retested them in 1953, 1960, and 1977. The 30-year

stability coefficients ranged from .28 for 1 (Hypochondriasis) and 8 (Schizophrenia) to .74 for 0 (Social Introversion), with a median of .39. By comparison, the median stability coefficient for the first 6-year interval was .59, suggesting some decay of individual ranking, even though there is still considerable continuity present after 30 years.

The major shortcoming of this study is that the MMPI is not a very good measure of personality traits. Its content is saturated with psychopathology, and many of the items refer to physical health or somatic complaints. It is reasonable to expect that changes in health—which are an undeniable accompaniment of aging—contribute to the changes in MMPI scores. This would explain the particularly low stability in such scales as Hypochondriasis and Schizophrenia, which include a disproportionate number of somatic items.

This interpretation is supported by a reanalysis conducted by Finn (1986). Instead of the clinical scales, he used item factor scales developed by Johnson, Butcher, Null, and Johnson (1984). These scales are more internally consistent and show predictable relations to other measures of normal personality (Costa, Busch, Zonderman, & McCrae, 1986). The median stability for these scales was .56. However, three of the scales did not show significant retest correlations: Denial of Somatic Problems, Depression, and Neurasthenic Somatization. In each case it is probably changes in physical health, rather than personality, that led to changes on these scales. Scales measuring Neuroticism, Social Extraversion, Intellectual Interests, and Cynicism—scales we would interpret as representing N, E, O, and low A—all showed 30-year stability coefficients of .56 or better.

THIRTY-YEAR STABILITY OF GZTS SCORES

We would like to be able to present 30-year retest data from the NEO-PI-R, but those data will not be available for another few decades. In the meantime, data on the Guilford-Zimmerman Temperament Survey (GZTS; Guilford, Zimmerman, & Guilford, 1976) for a small group of BLSA men are worth examining. The GZTS was first administered to BLSA participants in 1960, and participants were supposed to be retested every 6 years. After analyses of the 12-year data (Costa, McCrae, & Arenberg, 1980; Douglas & Arenberg, 1978),

Table 4

Stability Coefficients for Guilford-Zimmerman Temperament Survey Scales in Men Assessed on Five Occasions

GZTS Scale	Correlation Across Occasions T1 to T5 (retest interval in years)									
	T1–T2 (6.7)	T2–T3 (5.8)	T3–T4 (6.0)	T4–T5 (10.4)	T1–T3 (12.5)	T2–T4 (18.5)	T3–T5 (16.4)	T1–T4 (18.5)	T2–T5 (22.2)	T1–T5 (28.9)
General Activity	.80	.77	.76	.82	.79	.81	.65	.89	.74	.71
Restraint	.62	.79	.79	.76	.57	.86	.74	.41[a]	.88	.67
Ascendance	.77	.84	.80	.96	.78	.87	.88	.88	.91	.85
Sociability	.89	.71	.57	.64	.83	.61	.86	.67	.69	.80
Emotional Stability	.62	.86	.74	.67	.67	.59	.68	.38[a]	.77	.51
Objectivity	.59	.66	.71	.72	.49	.64	.32[a]	.55	.64	.63
Friendliness	.83	.72	.84	.64	.65	.60	.65	.62	.59	.63
Thoughtfulness	.75	.86	.82	.60	.79	.72	.55	.58	.71	.51
Personal Relations	.80	.89	.79	.83	.70	.82	.73	.71	.80	.61
Masculinity	.64	.77	.75	.61	.83	.82	.63	.71	.72	.52

Note. N = 16 to 23 men. Except as marked, all correlations are significant at $p < .05$.
[a]Not significant.

changes appeared to be so small that the retest interval was changed to 12 years. Because of scheduling problems the actual interval varied, but 23 men were eventually tested five times, with successive retest intervals ranging from 4 to 12 years. The full interval was 28 to 30 years. At the first administration, these men ranged in age from 25 to 57 (M = 42.4).

Table 4 presents stability coefficients across each occasion, grouped according to increased retest interval. Because of missing data, the Ns for these correlations range from 16 to 23; despite these small numbers, 97 of the 100 correlations are significant. The last column shows 29-year retest correlations that range from .51 to .85. Is there a pattern of declining correlation with increasing interval? The median value for the first four columns (adjacent administrations) is .77; the median for the next three columns (across every second administration) is .71; the median for the next two columns (across every third administration) is also .71; and the median of the final column is .63. These data suggest that there is indeed a decline in the strength of the relation as the interval increases, but it is relatively small.

It may seem unwise to interpret data based on so few cases, but comparisons with other data suggests that these correlations, partic-

ularly the medians, are reasonably accurate. For example, the median 6-year stability of GZTS scales in the full BLSA sample was .77; the median 12-year stability was .73 (Costa et al., 1980)—values close to the .77 and .71 we see for comparable intervals in the present sample. For the longer interval, we can turn to a second, independent subsample of BLSA participants. Many individuals missed one or more administrations; some dropped out of the study and were persuaded to return many years later. From among the whole group it was possible to find 150 men, initially aged 30 or over, whose last administration was at least 20 years (and at most 29 years) after their first. These subjects ranged in age from 30 to 67 at their first administration, and the mean retest interval was 23.7 years.

Table 5 shows the retest correlations, which range from .61 to .71. The median value of .65 is about the same as we find in our small five-administrations sample. But with somewhat more confidence we can manipulate these data to answer some more interesting questions. First, we know that all personality scales are more or less unreliable, and that in consequence observed retest correlations underestimate the true stability. The reliability estimates from an earlier study of the BLSA (Costa et al., 1980) are given in the second column of Table 5; by dividing the first column by the second we obtain the third, which estimates the stability we would see over a 24-year period if we had perfectly reliable measures. The median value of this column is .79, and this might be interpreted to mean that about four-fifths of the variance in personality traits is stable over a quarter century of adult life.

Assuming exponential decay, as Conley (1984) did, we can make a more extended prediction. What is the projected stability of personality traits over the full adult age span—a half century? The last column of Table 5 gives these values, which estimate the 50-year stability coefficients we would see if we had perfectly reliable measures. The coefficients range from .47 to .75 with a median value of .60. Thus it appears that about three-fifths of the variance in personality traits is stable across the full adult age range. Is there change as well as stability in individual differences? Yes, of course—but it appears that stability has an edge!

What is perhaps most impressive about these data is the generality of the finding of stability across traits. Even though some traits may change in an individual, others are likely to remain the same,

Table 5

*Stability, Reliability, and Estimated and Projected Stability
Coefficients for Guilford-Zimmerman Temperament Survey Scales
in Men Initially 30 and Over*

GZTS Scale	Observed 24-Year Retest	Reliability[a]	Estimated 24-Year Stability	Projected 50-Year Stability
General Activty	.64	.88	.72	.50
Restraint	.64	.80	.79	.61
Ascendance	.71	.82	.86	.74
Sociability	.68	.91	.74	.53
Emotional Stability	.62	.89	.70	.47
Objectivity	.62	.86	.72	.50
Friendliness	.65	.83	.78	.59
Thoughtfulness	.66	.78	.84	.70
Personal Relations	.61	.75	.81	.64
Masculinity	.69	.79	.87	.75

Note. $N = 133$–142 men. All observed correlations are significant at $p < .001$. Estimated stabilities are taken from the formulas of Heise (1969) and Converse and Markus (1979).
[a]Reliability estimates are from Costa, McCrae, and Arenberg (1980).

and the total picture, the personality as a whole, is one of remarkable consistency across time. This can be shown graphically in personality profiles. Figure 1 gives two examples of GZTS profiles taken at random from the sample of subjects with all five administrations. Of course, scores are not identical at each time; they would not be identical if the test were administered twice in the same day. Within the bounds of the reliability of the scales, there is remarkable similarity here: these two men have retained their distinctive pattern of traits over an interval of 30 years.

When we first analyzed GZTS scales in 1980 we found evidence for stability in all ten scales. At that time, however, we had no clear idea how well the GZTS covered the full range of personality traits. Were there other important traits not measured by this instrument that would show change with age? The five-factor model provides a basis for answering this question. The scales of the GZTS include several measures of N and E but few clear markers of O, A, or C (McCrae, 1989). To answer fully the question of change or stability in personality traits, we needed to measure the full range of traits. An-
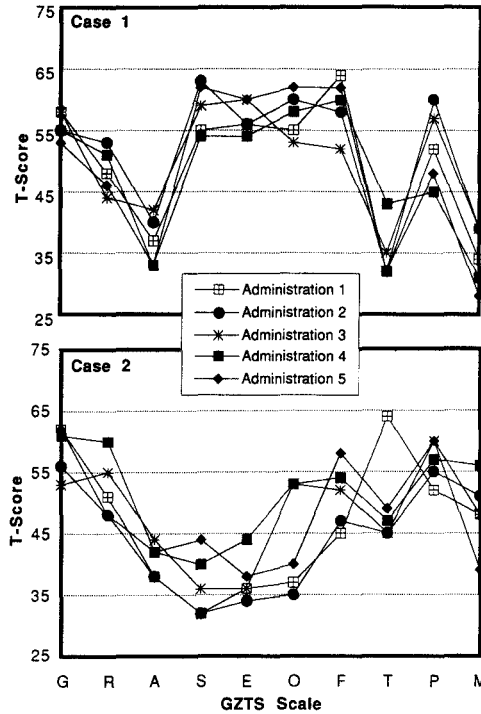
FIGURE 1. Guilford-Zimmerman Temperament Scale profiles for two men assessed at five administrations. G = General Activity, R = Restraint, A = Ascendance, S = Sociability, E = Emotional Stability, O = Objectivity, F = Friendliness, T = Thoughtfulness, P = Personal Relations, M = Masculinity.

alyses of the NEO-PI (Costa & McCrae, 1988b) suggested that traits in the domains of O, A, and C were also stable, but in the case of A and C these findings were limited to a 3-year interval and to self-report personality assessments. The next step in this program of research was a longer-term longitudinal study of all five domains using observer ratings in place of self-reports.

The Stability of Personality as Assessed by Peer Ratings

Although most of longitudinal studies of personality have employed self-reports, there is also an important tradition of studying

continuity and change in personality through observer ratings. Kelly (1955) collected both spouse and peer ratings in the 1930s as supplements to self-reports in his landmark study of personality, and Conley (1985) used these data to support his views on the stability of personality. A number of studies from Berkeley (Block, 1971; Mussen, Eichorn, Honzik, Bieber, & Meredith, 1980; Field & Millsap, 1991) have employed expert raters working from interview materials. Our own longitudinal studies have included a 6-year analysis of spouse ratings for three of the five major dimensions—N, E, and O (Costa & McCrae, 1988b). Analyses of peer ratings on all five domains over a 7-year interval provide new support for the view that personality is stable in adulthood.

It is customary to speak of observer ratings as a unitary method, set in opposition to self-reports. In fact, ratings differ both in methodology and in the relationship of the rater to the target person, and these differences can have consequences for the conclusions drawn from the data. Expert raters are often asked to make global judgments (e.g., Mussen et al., 1980), presumably on the assumption that their expertise is sufficient to overcome the psychometric limitations of single-item scales. If it is not, the result is likely to be an underestimate of true stability. At the other extreme, observers are sometimes given the rather arduous task of performing a Q-sort of a large number of items for each of the individuals they are asked to judge (e.g., Block, 1971). Here the sheer quantity of work may overtax the discriminating abilities of the judge. Our approach has been to use an instrument that parallels our self-report inventory—Form R of the NEO-PI-R—and to ask each rater to complete it for one other person. This strategy has the advantage of using multi-item scales of known reliability without overburdening research subjects.

Who is the ideal rater for personality research? If you believe that personality assessment requires lengthy training and an understanding of personality structure and dynamics, you are likely to choose experts who base their evaluations on interviews or direct observation, usually over a fairly limited period. The assessment weekend utilized at the Institute for Personality Assessment and Research adopted this strategy (John, 1990).

If you believe that specialized training is not necessary (provided questions are asked in an appropriate way, as any good questionnaire would do), then you may opt for ratings by significant

others, those who have intimate knowledge based on a shared life and shared confidences. Edwards and Klockars (1981) recommended such raters, and our research on spouse ratings confirms the utility of this source of data. But Kammann, Smith, Martin, and McQueen (1984) countered by arguing that spouse ratings may simply be self-reports at one remove: couples disclose their feelings to one another, and loving couples may uncritically adopt the self-presentations of their spouses.

An attractive alternative that combines the familiarity of long acquaintance with perhaps greater objectivity is the peer rating. As Funder (1991) states, "Peers' judgments have the advantage of being based on large numbers of behaviors viewed in realistic, daily contexts, and on the filtering of these behavioral observations through an intuitive system capable of adjusting for both immediate situational and long-term individual context" (p. 35). Among peers, those who are more familiar with the target person are better raters (Norman & Goldberg, 1966); our study involves friends and neighbors who have known the targets for many years in a variety of settings (McCrae & Costa, 1987).

A LONGITUDINAL STUDY OF RATINGS

In 1983, as part of our initial research on the five-factor model, we administered what was to become the NEO Personality Inventory to a large group of peer raters (McCrae & Costa, 1987). These raters had been nominated by participants in the BLSA, who were asked to provide the names of "three or four individuals who know you very well *as you are now*. They can be friends, neighbors, or co-workers, but they should *not* be relatives. These should be people who have known you for a least one year and have seen you in a variety of situations." Complete data were obtained from 743 raters.

In 1990 we approached those BLSA subjects who were still active in the study and who had been rated in 1983 and asked them for permission to make contact with the original raters and for current addresses. Because retirement and relocation are common in this sample, we also asked subjects if they had seen the raters within the past year: we wanted to make sure the second ratings were based on the rater's current view of the subject, so that any intervening

changes in personality traits could be detected. In the final sample, all but ten of the raters had been seen within the past year, and these ten had maintained contact by mail or telephone. A total of 285 longitudinal raters were sent questionnaires; of these, 97 raters of 54 different men and 60 raters of 37 different women provided complete data. The target men ranged in age from 33 to 81 (M = 58.8) when they were first rated in 1983; the target women were aged 31 to 79 (M = 52.6).

In 1986 we had invited a new group of BLSA participants to join our study (Costa & McCrae, 1988b), and in 1990 we also asked these individuals to nominate peer raters, using the same instructions we had used in 1983. This provided a comparison group that could be used in cross-sequential analyses and broadened the base for cross-sectional analyses. Questionnaires were sent to 203 new raters; of these, 50 raters of 21 men and 82 raters of 35 women returned complete data. The target men in this new sample ranged in age from 31 to 87 (M = 72.4) in 1990; the target women ranged in age from 29 to 93 (M = 68.9). The new sample is thus considerably older than those remaining from the original sample, as well as having a higher proportion of women.

All these raters were given several questionnaires, including Form R of the NEO-PI and a set of 120 supplemental items that allowed us to score the revised NEO-PI, or NEO-PI-R. Longitudinal comparisons must of course be based on the NEO-PI itself, but cross-sectional comparisons and other analyses can employ the extended NEO-PI-R.

Table 6 provides some descriptive data on this instrument in the combined sample. The first four columns give means and standard deviations for ratings of men and women. (Note that all the data in this table are based on single peer ratings rather than mean peer ratings.) There are remarkably few sex differences: None of the five domains and only 8 of the 30 facet scales show significant effects for gender. Women tend to be rated as being somewhat higher in Anxiety, Activity, Positive Emotions, and Openness to Aesthetics, Feelings, and Actions. Men are rated higher in Openness to Ideas and Deliberation.

The next three columns give information on the psychometric properties of Form R of the NEO-PI-R. After all, if the intrument were not reliable and valid, there would be no reason to use it to assess

Table 6

Descriptive Statistics, Reliability, and Correlates of Form R NEO-PI-R Scales

NEO-PI-R	Men		Women		Coeffi-cient	Correlations		
Scale	Mean	SD	Mean	SD	Alpha	Intraclass	Self-Peer	Age
Domains								
Neuroticism	64.6	17.8	68.9	22.3	.93	.43***	.36***	−.15*
Extraversion	115.0	19.6	119.0	18.9	.90	.42***	.44***	−.16**
Openness	106.6	16.0	110.4	17.4	.89	.45***	.53***	−.16**
Agreeableness	127.6	23.6	130.0	22.5	.95	.49***	.41***	.18**
Conscientiousness	137.0	18.1	137.7	17.2	.92	.22**	.40***	.05
N facets								
Anxiety	12.1	4.2	13.5ª	5.2	.82	.30***	.35***	−.11
Hostility	11.1	5.4	11.5	6.2	.86	.42***	.34***	−.14*
Depression	10.0	3.8	11.0	5.2	.81	.38***	.34***	−.07
Self-Consciousness	11.4	3.6	11.6	4.1	.73	.27***	.26***	−.08
Impulsiveness	12.1	4.2	12.8	4.1	.69	.33***	.26***	−.29***
Vulnerability	7.9	3.3	8.4	3.9	.81	.19**	.25***	.05
E facets								
Warmth	24.8	4.8	25.7	3.9	.81	.37***	.33***	.02
Gregariousness	19.0	5.1	20.0	5.2	.79	.31***	.33***	−.07
Assertiveness	19.1	4.8	19.5	4.8	.76	.40***	.50***	−.09
Activity	17.5	4.6	18.6ª	4.9	.77	.41***	.47***	−.22***
Excitement Seeking	15.7	4.5	14.9	4.6	.74	.48***	.48***	−.21***
Positive Emotions	18.9	4.9	20.3ª	4.7	.82	.39***	.28***	−.08
O facets								
Fantasy	14.6	3.8	14.7	3.8	.72	.35***	.43***	−.24***
Aesthetics	16.9	4.6	18.9ᵇ	4.7	.81	.45***	.52***	.03
Feelings	19.1	3.4	20.3ª	4.2	.69	.15*	.37***	−.17**
Actions	14.9	3.4	16.8ᵇ	3.7	.60	.36***	.36***	−.16**
Ideas	21.8	5.1	19.6ᵇ	5.3	.87	.37***	.38***	.05
Values	19.3	3.7	20.2	4.0	.69	.48***	.35***	−.25***
A facets								
Trust	22.6	4.6	23.3	4.6	.90	.43***	.16*	.15*
Straightforwardness	22.3	4.8	22.7	5.3	.84	.43***	.33***	.20***
Altruism	23.9	4.3	24.7	4.2	.80	.36***	.33***	.06
Compliance	19.6	4.5	19.7	5.5	.78	.54***	.47***	.15*
Modesty	19.3	5.9	19.1	5.1	.83	.21**	.28***	.20***
Tender-Mindedness	20.0	3.9	20.6	3.4	.69	.27***	.25***	.08
C facets								
Competence	25.1	3.4	25.6	3.2	.73	.25***	.26***	−.01
Order	20.2	4.2	20.5	4.3	.71	.19**	.36***	−.01
Dutifulness	24.8	3.7	25.2	3.3	.70	.20**	.28***	.11
Achievement Striving	21.2	4.0	21.3	3.9	.70	.33***	.42***	.02
Self–Discipline	23.7	4.2	24.3	4.1	.82	.12	.33***	.01
Deliberation	21.9	3.6	20.8ª	3.9	.73	.20**	.35***	.13*

Note. Ns = 143 ratings of 73 men and 134 ratings of 69 women. Intraclass correlations are based on 193 pairs of ratings; self-peer correlations are based on 250 pairs of self-reports with single peer ratings. * $p < .05$. ** $p < .01$. *** $p < .001$.
ªDifference between men and women significant at $p < .05$
ᵇDifference between men and women significant at $p < .001$

stability or change in personality. The data are in fact quite strong. Alpha coefficients ranging from .60 to .95 show internal consistency for all domains and facets. Intraclass correlations (computed as the correlation between all possible pairs of raters, using the double-entry method) are significant for all domains and for 29 of the 30 facet scales; the median value is .36.

Intraclass correlations are sometimes viewed as indicators of interrater reliability, and as reliability coefficients these are modest figures. But interrater reliability properly refers to the agreement among judges who assess the same sample of behavior—for example, the interview records reviewed by Vaillant and Vaillant (1990). Peer raters form their impressions based on their own history of interaction with the target person, which may vary considerably from peer to peer. The rater who knows the target from work may see a different side than the rater who has seen the person only in social settings. These intraclass correlations should be regarded as cross-observer validity coefficients, and in this regard they are substantial for most of the scales.

Correlations between Form S and Form R of the NEO-PI-R show self-peer agreement that is significant for all 35 scales and comparable in magnitude to the agreement seen between peers. Of particular note is that the Conscientiousness scales, which showed relatively low intrapeer agreement, show substantial agreement between self-reports and ratings.

The Form R facet scales were also factored. As with the self-reports in Table 2, five factors had eigenvalues greater than 1.0, and after rotation, very similar factors emerged. Coefficients of congruence with the factors in Table 2 ranged from .93 to .97.

Finally, the question remains whether these rated facet scales show differential validity. One simple and instructive way to test this is by examining the pattern of correlations with Form S scales, which have known differential validity (see Table 3). We can see the convergent validity of facet scales in Table 6. The median values for the N, E, O, A, and C domains are .30, .40, .38, .31, and .34, respectively. These values are obtained by correlating each Form R scale with the corresponding Form S scale. If we correlated Form R scales with Form S scales that represent different facets of the same domain—for example, rated Openness to Ideas with self-reported Openness to Feelings—we would expect correlations to be positive

but lower in magnitude than the simple convergent correlations. For each domain there are 30 such "semiconvergent" correlations; the median values are .20, .18, .21, .19, and .21, for N, E, O, A, and C. True discriminant correlations are seen when rated facets from one domain are correlated with self-reported facets from another domain—say, rated Anxiety with self-reported Trust. For each domain there are 288 such correlations; the median absolute discriminant correlations for the five domains are .07, .10, .08, .09, and .08. NEO-PI-R Form R scales thus show reliability, appropriate factor structure, convergent, discriminant, and differential validity. They should be useful in examining stability or change in personality traits.

MEAN LEVEL CHANGES IN PERSONALITY TRAITS WITH AGE

Correlations between ratings and the target's age—given in the last column of Table 6—are a convenient way of summarizing cross-sectional age relations. There are a number of significant relations, the largest of which suggest that increasing age is associated with lower Impulsiveness, Activity, Excitement Seeking, and Openness to Fantasy and Values and with higher Straightforwardness and Modesty. These findings, particularly the lower scores on Activity and Excitement Seeking, are hardly surprising. What is more surprising, at least to those who expect age to exert a pronounced effect on personality, is the very modest magnitude of the correlations, especially because the criterion—age—shows an extreme range (from 29 to 93) and is measured with almost perfect reliability. In any case, these cross-sectional data confound age with birth cohort, and some of the associations may to due to generational differences. The lower scores on Openness to Values, in particular, may reflect socialization in a more traditional society.

As we all know, the usual solution to deconfounding age and date of birth is through the longitudinal study. When the same individuals are measured at two or more times, they can be used as their own controls; in this design not only generational differences but also other possible confounds (such as educational differences or attrition biases) are controlled. Repeated-measures analyses are statistically powerful, so that even small effects can be significant.

Table 7 presents the results of a longitudinal analysis of scales from the NEO-PI. Simple paired *t* tests were used within men and women separately, and no correction was made for multiple statistical tests. This is a statistically liberal approach, designed to let even weak aging effects show themselves. Despite this, only 7 of the 70 comparisons showed significant changes. Men were seen by raters as showing a decline in Activity and in Openness to Feelings and Ideas but an increase in Openness to Fantasy. Women were seen as showing declines in Activity and Openness to Ideas and an increase in total Conscientiousness. None of these changes amounts to more than one-quarter of one standard deviation in magnitude over a period of 7 years. Clearly, the picture these data present is one of predominant stability in the mean levels of personality traits.

No single design or combinations of designs can definitively disentangle aging, cohort, and time-of-measurement effects, but combinations of designs can increase the probability of certain interpretations (Costa & McCrae, 1982). When the results in Table 7 are interpreted in conjunction with the cross-sectional data in Table 6, we see that both cross-sectional and longitudinal findings are consistent with the view that Activity and perhaps Openness to Feelings decline with age. None of the other cross-sectional findings is replicated in the longitudinal analyses. The decline in activity level has been reported before (e.g., Costa & McCrae, 1988b; Douglas & Arenberg, 1978) and is almost certainly an accurate reflection of real age-related change, but it is perhaps the only clear example of a personality trait that does change after age 30.

The third design commonly used in analyses of mean level changes is called a *cross-sequential* design (Schaie, 1977). In this analysis, individuals who were born at the same time but tested at different times are compared. For example, one group of people born in 1900 could be tested in 1983, a second group, also born in 1900, could be tested in 1990. Because the latter group would be 7 years older at testing, differences between the two groups might be due to aging. Differences between them could not be due to birth cohort, because all subjects were born at the same time. In addition, previous exposure to the test could not affect results, because all subjects were tested only once. We employed a version of this design in comparing the 1983 data from our original set of raters with the 1990 data from our new raters.

Table 7

Seven-Year Longitudinal Changes and Stability Coefficients for Peer Ratings of Men and Women on the NEO Personality Inventory

NEO-PI-R Scale	Men 1983 Mean	SD	1990 Mean	SD	Stability Coefficient	Women 1983 Mean	SD	1990 Mean	SD	Stability Coefficient
Domains										
Neuroticism	63.3	18.8	63.6	17.9	.76	75.8	23.0	74.9	21.4	.67
Extraversion	112.0	18.5	111.1	19.2	.78	115.0	19.0	112.5	19.3	.81
Openness	108.2	17.9	108.1	15.9	.76	107.9	17.2	106.1	15.6	.84
Agreeableness	49.9	8.0	49.8	8.4	.75	49.5	8.8	48.8	9.9	.63
Conscientiousness	53.7	7.7	54.0	8.7	.74	53.4	10.3	56.0[b]	10.1	.78
N facets										
Anxiety	11.9	4.8	12.3	4.4	.71	14.8	5.1	14.5	5.4	.51
Hostility	9.9	5.3	10.1	5.2	.74	11.3	5.3	12.2	6.0	.54
Depression	10.4	4.3	9.8	3.8	.69	12.6	5.7	11.8	5.0	.71
Self-Consciousness	10.9	3.8	11.3	3.6	.76	13.0	5.1	12.8	4.1	.72
Impulsiveness	12.5	4.2	12.2	4.1	.66	15.0	5.1	14.4	4.3	.70
Vulnerability	7.7	3.4	7.9	3.6	.66	9.2	4.4	9.1	3.9	.69
E facets										
Warmth	24.6	4.3	25.0	4.2	.71	24.8	4.1	24.9	4.1	.62
Gregariousness	17.2	4.1	17.5	4.5	.67	18.7	4.4	17.9	5.2	.76
Assertiveness	19.4	5.3	18.8	4.8	.78	19.7	6.1	19.5	6.0	.79
Activity	18.0	5.1	16.9[b]	4.9	.73	18.8	5.6	17.4[b]	5.5	.82
Excitement Seeking	14.4	4.5	14.1	4.8	.79	13.9	4.4	13.7	4.2	.67
Positive Emotions	18.4	4.8	18.7	5.0	.62	19.2	4.4	19.1	4.7	.65
O facets										
Fantasy	13.5	4.3	14.6[b]	3.6	.67	14.5	3.2	15.0	3.7	.63
Aesthetics	17.0	4.4	17.0	4.6	.69	17.8	4.7	17.5	4.4	.80
Feelings	20.6	3.8	19.9[a]	3.5	.58	21.1	4.3	20.7	4.2	.77
Actions	14.4	3.9	14.7	3.8	.71	16.1	4.0	15.9	4.1	.61
Ideas	22.9	5.4	22.0[b]	4.9	.71	18.9	5.9	17.7[a]	5.3	.71
Values	19.7	4.6	19.8	3.9	.68	19.5	4.1	19.4	4.2	.64

Note. Ns = 97 ratings of 54 men, 60 ratings of 37 women. All stability coefficients are significant at $p < .001$.
[a] Change from 1983 to 1990 is significant at $p < .05$
[b] Change from 1983 to 1990 is significant at $p < .01$

Cross-sequential comparisons make the assumption that the two groups are similar with regard to all relevant variables except the time at which they are assessed; unless one begins with a design that randomly assigns subjects to "test now" and "test later" conditions, this assumption is likely to be violated to some degree, but we believe it is worthwhile to examine the data even though the experimental design is less than perfect. Cross-sequential analyses hold one more piece of a complex puzzle, and our application may be a

useful illustration of how cross-sequential analyses can be approximated in many data sets.

A first concern, not encountered in cross-sequential studies of self-reports, is the comparability of raters. We are interested in differences due to the subjects' aging, so any differences between the two sets of raters would confound interpretation of the data. A background sheet completed by raters suggests that the two groups are similar. For example, in the original sample (McCrae & Costa, 1987) the age range of the raters was 19 to 87; in the present sample it is 26 to 91. Most of the raters in both samples (57% and 64%) had college degrees. In the earlier study, the mean length of acquaintance had been 18.3 years; in the present study it was 23 years. In both samples, 75% of the raters described themselves as close personal friends of the target persons. Most raters (57%) in the first study reported seeing or talking to the targets at least weekly; the corresponding figure was 56% in the current study. A more detailed description of the raters is given in our earlier report (McCrae & Costa, 1987); in general, the figures given there can be taken to apply to the present set of raters as well.

With regard to the subjects themselves, one difference between those who were rated on the NEO-PI in 1983 and those who were first rated in 1990 is that all members of the latter group were alive and willing participants in the longitudinal study in 1990; some of the former had died, and others had become incapacitated or had for other reasons dropped out of the study. If personality traits are systematically related to the probability of death, disability, or attrition, then it would be misleading to compare the 1983 ratings with the 1990 ratings. This problem can easily be solved by restricting the cross-sequential analyses to those individuals rated in 1983 who were also rated in 1990—the same subjects on whom repeated-measures analyses were performed.

But even after this restriction, the two groups are different in at least two important ways. Recruitment patterns in the BLSA have changed over the years, and as we noted earlier, the new sample has a higher proportion of women than men and an earlier date of birth than the original sample. We therefore used a matching strategy. Ratings from 1983 in the original sample were matched by sex and date of birth (within one year) with ratings from 1990 in the new sample. The resulting sample consisted of 45 pairs of individuals,

born between 1902 and 1950 (M = 1921). Although the two groups did not differ in date of birth, they did differ by 7 years in the date at which they were rated. If personality traits increase or decrease with age, the effects should be seen when comparing the two sets of ratings.

There were no significant differences between the two groups on any of the five domain scales, and only two of the 18 N, E, and O facet scales showed an effect ($p < .05$): the new sample was rated somewhat lower in Depression and higher in Warmth than the original sample. These differences replicate neither the cross-sectional nor the longitudinal analyses and are probably best attributed to chance. Overall, the cross-sequential analyses confirm the general view that the mean levels of most traits change little in adults.

THE STABILITY OF INDIVIDUAL DIFFERENCES

As we noted previously, there are two distinct meanings of the term *stability*. The stability of individual differences refers to the extent to which individuals maintain their rank order in the distribution of a trait: Do introverts remain introverts, or do they become extraverts? Cross-sectional and cross-sequential designs tell us nothing about this question; repeated measurements of the same individual over time are essential. The retest correlation of two sets of scores, which is interpreted as a measure of reliability when the retest interval is short, can be interpreted as a stability coefficient when the interval is substantial.

Table 7 gives 7-year stability coefficients for ratings of men and women separately. These values range from .51 to .84, with median values of .70 for women and .71 for men. By comparison, the median values over a 6-year interval for self-reports from women and men initially aged 57 to 84 were .69 and .75; the median values for 6-year retests of spouse ratings for the 21 N, E, and O scales were .73 and .74 (Costa & McCrae, 1988b). All three sources concur in showing levels of retest stability that are not far from the short-term retest reliability of the scales.

Those familiar with the literature on the longitudinal stability of personality as seen in observer ratings may be puzzled by the magnitude of these correlations. Field and Millsap (1991), for example,

report 14-year stability coefficients of .09 to .53. Block's (1971) California Q-Set (CQS) items show median correlations between senior high school and the mid-30s of .26 for men and .25 for women. Why are the correlations in Table 7 so much higher?

There are probably several reasons. In the Block study, subjects were adolescents at the time of the first assessment, and there is considerable evidence that there are true changes in personality between adolescence and middle adulthood (e.g., Haan, Millsap, & Hartka, 1986; Siegler et al., 1990). In the Block study, single items were analyzed rather than multi-item scales; it is a simple fact of psychometrics that aggregating similar items yields more reliable data. Field and Millsap (1991) found higher stability for component scores than for individual items.

But perhaps more important than either of these features is that in most previous research, different raters have been used to assess personality at different ages, and different raters have different views of the individual and different styles of responding to assessment instruments. There is thus an inherent confounding of change in the target person with differences in the raters—and these latter differences are by no means trivial. Table 6 gives the interrater agreement for pairs of peer raters; the median value is .36 *when ratings are made at the same time*. It is hardly surprising that correlations *across time* rarely exceed .4 when they are made by different raters.

The data in the present study allow us to deconfound raters and time of ratings by using a somewhat different analysis. We can approximate the designs of Block and of Field and Millsap by looking at the correlations between one rater at one time and a different rater at another; in fact, we can get the best estimate of this value by correlating all possible pairs ($N = 166$) of ratings by different raters of the same target at different times: a cross-lagged intraclass correlation. For the five NEO-PI domain scales N, E, O, A, and C, these values are .36, .39, .44, .32, and .21, respectively. Had we been unable to make contact with the original peer raters again, we could have carried out a peer-rating study by recruiting a new set of raters, and we would probably have seen "stability coefficients" in this .2 to .4 range—leading perhaps to a very different view of the stability of personality.

Even the much larger values seen in Table 7 may in fact underestimate the stability of rated personality, because individual raters

are not perfectly reliable. Correcting for attenuation due to unreliability is one approach. Block (1971) reported disattenuated stability coefficients correcting for interrater unreliability, and Conley (1984) used internal consistency to correct stability coefficients. When the same rater and the same instrument are used on the two occasions, short-term test-retest reliability can be used to estimate true stability (Costa & McCrae, 1988b).

At present there are no data available on the short-term retest reliability of Form R NEO-PI scales. The longitudinal stability coefficients are a lower-bound estimate and suffice for most purposes to show the reliability of the scales. However, it is possible to estimate the true stability of rated personality by the use of path analysis (cf. Siegler et al., 1990), as Figure 2 shows: the ratio of the cross-lagged intraclass correlation to the concurrent intraclass correlation estimates true score stability. Intuitively, this makes sense. The concurrent intraclass correlations show how much of the variance in trait ratings is due to the trait itself at one time; the cross-lagged correlations show how much variance is due to that part of the trait itself that is stable over the interval. The ratio of these is the portion of the trait that is stable.

We have already reported the cross-lagged intraclass correlations for the NEO-PI domain scales. The concurrent intraclass correlations—based on all possible pairs of different raters of the same individual at the same time, either 1983 or 1990—are .40, .42, .51, .31, and .25 for N, E, O, A, and C, respectively. The ratio of the two correlations yield estimates of 7-year true score stability of .90, .93, .86, 1.03, and .84. The value of 1.03, of course, is impossible and serves as a reminder that these are only estimates. However, they do illustrate that even the high retest correlations seen in Table 7 underestimate the true stability of rated personality. Personality as seen by peer raters is extremely stable in adulthood.

Limitations to Generalizations About Stability

The data presented here and reviewed elsewhere (e.g., Kogan, 1990) clearly make the point that personality traits are, in general, stable. In the face of such evidence, it is perhaps wise to spell out the limita-
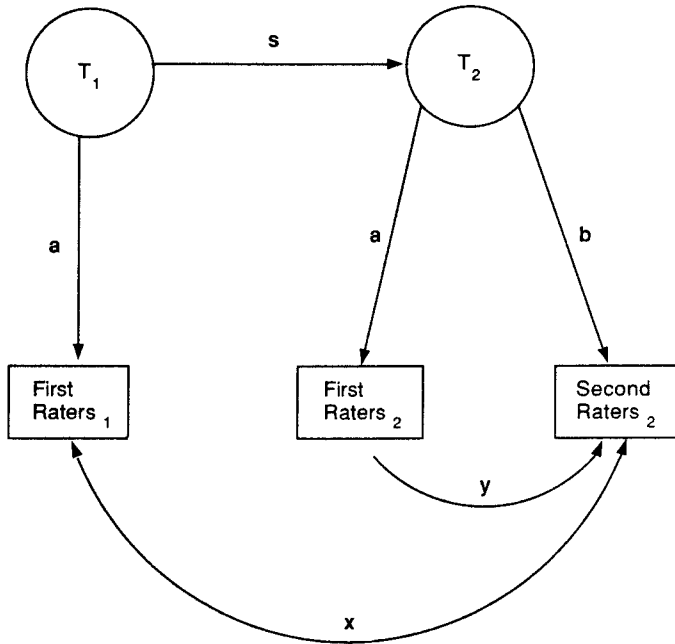
Figure 2. A path-analytic diagram for estimating the stability of true scores. In this illustration, assessments of trait T are given by two groups of raters at two times. Let a represent the correlation of the true scores T1 and T2 with the scores from the First Raters at the two times. Let b represent the correlation of T2 with scores from the Second Raters at the second time, and s the stability coefficient for T. The observed cross-lagged and concurrent intraclass correlations are x and y, respectively. From the principles of path analysis, $x = asb$, and $y = ab$. The ratio of x to y, x/y, is thus asb/ab, or s, the stability coefficient.

tions to this otherwise robust generalization. There are several, known and suspected.

1. There are both continuity and change between adolescence and adulthood. College-age students, in particular, are somewhat higher in N, E, and O, and lower in A and C, than are older adults (Costa & McCrae, 1989a). As Zuckerman (1979) pointed out, excitement seeking declines dramatically between turbulent adolescence and staid middle age. Similarly, there are changes in individual rank order in this same period (Siegler et al., 1990).

2. Even among adults over age 30 there are small changes in the mean levels of some traits, such as activity level. There is also a very gradual decay in the predictability of individual differences over

time. Perhaps as much as 40% of true score variance changes over a 50-year adult life span.

3. Personality changes may occur as the result of psychiatric disorders. Certainly the scores that clinically depressed individuals receive on personality inventories change between periods of depression and periods of remission (Hirschfeld et al., 1983), although this effect is seen primarily on measures of N, not of E or O. Whether this represents true, but temporary, change in personality or whether it reflects merely a distortion of the personality assessment process remains to be seen.

4. Personality may change as a result of psychotherapy. We have argued elsewhere that dramatic changes in personality are not to be expected as an outcome of psychotherapy (Costa & McCrae, 1986), but enduring smaller changes might be seen. Other experiences, such as religious conversion, might also affect personality. These are important topics for future research.

5. Personality may change as a result of catastrophic stressors. We know that normal life stressors (including retirement, bereavement, and many chronic illnesses) do not have major effects on personality, but there may be certain types or degrees of stress that do. Research on posttraumatic stress syndrome suggests that events can have permanent effects on basic dispositions; longitudinal studies that include both pre- and post-measures are needed here.

6. Personality unfortunately does change in response to dementias such as Alzheimer's disease. Retrospective research using spouse ratings of what the patient was like before onset of the dementia and what the patient is like now suggests profound changes in a number of aspects of personality, especially decreased Conscientiousness (personal communication, I. C. Siegler, 1990). Given progressive cognitive deterioration, it is hardly surprising that patients become lower in competence, order, and achievement striving. Declines in C might be useful early markers of Alzheimer's disease.

Adult Development and Trait Psychology

This chapter has been largely—perhaps excessively—psychometric. Cross-lagged intraclass correlation coefficients may provide unique information on the stability of rated personality, but they

probably have little appeal for those who want to study the development of life structures (Levinson, 1986) or to understand the personality of a given individual in a unique historical time and place (Runyan, 1990). How should we deal with such alternative approaches to the study of personality in adults?

We do not think it is scientifically acceptable to avoid the confrontation. The eclecticism that tolerates many different perspectives on a topic is commendable as an approach to generating ideas. But in science, ideas must also be tested: different theories must at some point be pitted against each other. Trait psychology offers a set of widely replicated observations about the long-term stability of recurrent personality factors that can be assessed in a variety of ways. Further, these traits are important in the lives of individuals, influencing everything from choice of an occupation to the development of psychopathology. We expect these data to be taken seriously by anyone who wishes to construct a theory of adult development.

For example, theories of a midlife crisis that predict a universal period of emotional upheaval around age 40 are simply wrong (Farrell & Rosenberg, 1981; McCrae & Costa, 1990). Jungian notions about the balancing of functions in old age are equally untenable (McCrae & Costa, 1989), and Erikson's venerable stages of psychosocial development may require rethinking. Theories of adult development need to be consistent with the facts about trait stability.

This does not mean that adult development is a myth, however, or that alternative approaches to understanding personality have nothing to offer. It is a sociological fact that individuals' life structures change as they move from adolescence to maturity to retirement, even though their dispositions may remain much the same. It is a historical truism that the open extravert who lived in the Middle Ages faced a much different world than the open extravert who lives in the 1990s. There is a great deal to be learned by integrating trait psychology with the insights of biography, anthropology, and neuropsychology.

Elsewhere (McCrae & Costa, 1990, chap. 9) we have discussed the ways stable dispositions can help explain both the continuities and the changes seen in adult lives. Personality traits affect our attitudes and opinions, the social roles we select and the ways we interpret those roles, our closest interpersonal relationships, even the stories we tell ourselves about our lives. They are certainly not the

only influence on our lives—intelligence and physical attractiveness, health and wealth, race, sex, and religion are also important determinants, as are the larger social contexts of family, community, and nation. But each of the five factors—Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness—is essential for understanding the course of life, and a systematic study of their effects on life choices and life outcomes could form the basis for a new view of adulthood.

Levinson (1986) recently noted that "the study of adult development is . . . in its infancy" (p. 3). By comparison, trait psychology is a mature science, with established methods and findings. Progress in understanding adult development will come in part from an appreciation of the role of enduring dispositions in shaping the life course.

REFERENCES

Block, J. (1971). *Lives through time*. Berkeley, CA: Bancroft Books.

Block, J. (1977). Advancing the psychology of personality: Paradigmatic shift or improving the quality of research? In D. Magnusson & N. S. Endler (Eds.), *Personality at the cross-roads: Current issues in interactional psychology* (pp. 37–64). Hillsdale, NJ: Erlbaum.

Buss, A. H., & Durkee, A. (1957). An inventory for assessing different kinds of hostility. *Journal of Consulting Psychology, 21,* 343–348.

Caspi, A., Elder, G. H., Jr., & Bem, D. J. (1987). Moving against the world: Life-course patterns of explosive children. *Developmental Psychology, 23,* 308–313.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *The handbook for the Sixteen Personality Factor Questionnaire.* Champaign, IL: Institute for Personality and Ability Testing.

Conley, J. J. (1984). 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, 5,* 11–26.

Conley, J. J. (1985). Longitudinal stability of personality traits: A multitrait-multimethod-multioccasion analysis. *Journal of Personality and Social Psychology, 49,* 1266–1282.

Converse, P. E., & Markus, G. B. (1979). Plus ça change . . .: The new CPS Election Study panel. *American Political Science Review, 73,* 32–49.

Costa, P. T., Jr., Busch, C. M., Zonderman, A. B., & McCrae, R. R. (1986). Correlations of MMPI factor scales with measures of the five-factor model of personality. *Journal of Personality Assessment, 50,* 640–570.

Costa, P. T., Jr., & McCrae, R. R. (1976). Age differences in personality structure: A cluster analytic approach. *Journal of Gerontology, 31*, 564–570.

Costa, P. T., Jr., & McCrae, R. R. (1977). Age differences in personality structure revisited: Studies in validity, stability, and change. *International Journal of Aging and Human Development, 8*, 261–275.

Costa, P. T., Jr., & McCrae, R. R. (1980). Still stable after all these years: Personality as a key to some issues in adulthood and old age. In P. B. Baltes & O. G. Brim, Jr. (Eds.), *Life span development and behavior* (Vol. 3, pp. 65–102). New York: Academic Press.

Costa, P. T., Jr., & McCrae, R. R. (1982). An approach to the attribution of age, period, and cohort effects. *Psychological Bulletin, 92*, 238–250.

Costa, P. T., Jr., & McCrae, R. R. (1985a). Concurrent validation after 20 years: Implicatons of personality stability for its assessment. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (Vol. 4, pp. 31–54). Hillsdale, NJ: Erlbaum.

Costa, P. T., Jr., & McCrae, R. R. (1985b). *The NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1986). Personality stability and its implications for clinical psychology. *Clinical Psychology Review, 6*, 407–423.

Costa, P. T., Jr., & McCrae, R. R. (1988a). From catalog to classification: Murray's needs and the five-factor model. *Journal of Personality and Social Psychology, 55*, 258–265.

Costa, P. T., Jr., & McCrae, R. R. (1988b). 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, 54*, 853–863.

Costa, P. T., Jr., & McCrae, R. R. (1989a). *The NEO-PI/NEO-FFI manual supplement*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1989b, November). What lies beneath the Big Five? Facet scales for Agreeableness and Conscientiousness. In O. P. John (Chair), *The Big Five: Historical perspective and current research.* Symposium conducted at the Annual Meeting of the Society for Multivariate Experimental Psychology, Honolulu.

Costa, P. T., Jr., & McCrae, R. R. (in press). Multiple uses for longitudinal personality data. *European Journal of Personality*.

Costa, P. T., Jr., McCrae, R. R., & Arenberg, D. (1980). Enduring dispositions in adult males. *Journal of Personality and Social Psychology, 38*, 793–800.

Costa, P. T., Jr., McCrae, R. R., & Dye, D. A. (1991). Facet scales for Agreeableness and Conscientiousness: A revision of the NEO Personality Inventory. *Personality and Individual Differences, 12*, 887–898.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*, 417–440.

Douglas, K., & Arenberg, D. (1978). Age changes, cohort differences, and cultural change on the Guilford-Zimmerman Temperament Survey. *Journal of Gerontology, 33*, 737–747.

Edwards, A. L., & Klockars, A. J. (1981). Significant others and self-evalua-
tion: Relationships between perceived and actual evaluations. *Personality
and Social Psychology Bulletin, 7*, 244–251.

Eichorn, D. H., Clausen, J. A., Haan, N., Honzik, M. P., & Mussen, P. H.
(Eds.). (1981). *Present and past in middle life*. New York: Academic Press.

Farrell, M. P., & Rosenberg, S. D. (1981). *Men at midlife*. Boston: Auburn
House.

Field, D., & Millsap, R. E. (1991). Personality in advanced old age: Continu-
ity or change? *Journals of Gerontology: Psychological Sciences, 46*, P299–
P308.

Finn, S. E. (1986). Stability of personality self-ratings over 30 years: Evidence
for an age/cohort interaction. *Journal of Personality and Social Psychology,
50*, 813–818.

Funder, D. C. (1991). Global traits: A Neo-Allportian approach to person-
ality. *Psychological Science, 2*, 31–39.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-
Five factor structure. *Journal of Personality and Social Psychology, 59*, 1216–
1229.

Gough, H. G. (1987). *California Psychological Inventory administrator's guide*.
Palo Alto, CA: Consulting Psychologists Press.

Guilford, J. S., Zimmerman, W. S., & Guilford, J. P. (1976). *The Guilford-Zim-
merman Temperament Survey handbook: Twenty-five years of research and ap-
plication*. San Diego, CA: EdITS Publishers.

Haan, N., Millsap, R., & Hartka, E. (1986). As time goes by: Change and sta-
bility in personality over fifty years. *Psychology and Aging, 1*, 220–232.

Heise, D. R. (1969). Separating reliability and stability in test-retest correla-
tion. *American Sociological Review, 34*, 93–101.

Hirschfeld, R. M. A., Klerman, G. L., Clayton, P., Keller, M. B., McDonald-
Scott, P., & Larkin, B. (1983). Assessing personality: Effects of depressive
state on trait measurement. *American Journal of Psychiatry, 140*, 695–699.

Holland, J. L. (1985). *Self-Directed Search—1985 edition*. Odessa, FL: Psycho-
logical Assessment Resources.

Jackson, D. N. (1984). *Personality Research Form manual* (3rd ed.). Port Huron,
MI: Research Psychologists Press.

John, O. P. (1990). The "Big Five" factor taxonomy: Dimensions of person-
ality in the natural language and in questionnaires. In L. Pervin (Ed.),
Handbook of personality theory and research. New York: Guilford.

John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to
personality: A historical review of trait taxonomic research. *European
Journal of Personality, 2*, 171–203.

Johnson, J. H., Butcher, J. N., Null, C., & Johnson, K. N. (1984). Replicated
item level factor analysis of the full MMPI. *Journal of Personality and Social
Psychology, 47*, 105–114.

Kammann, R., Smith, R., Martin, C., & McQueen, M. (1984). Low accuracy
in judgments of others' psychological well-being as seen from a phenom-
enological perspective. *Journal of Personality, 52*, 107–123.

Kelly, E. L. (1955). Consistency of the adult personality. *American Psychologist, 10,* 659–681.

Kogan, N. (1990). Personality and aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 330–346). San Diego, CA: Academic Press.

Leon, G. R., Gillum, B., Gillum, R., & Gouze, M. (1979). Personality stability and change over a 30 year period—middle age to old age. *Journal of Consulting and Clinical Psychology, 47,* 517–524.

Levinson, D. J. (1986). A conception of adult development. *American Psychologist, 41,* 3–13.

Lorr, M. (1986). *Interpersonal Style Inventory (ISI) manual.* Los Angeles: Western Psychological Services.

McAdams, D. P. (1990). Unity and purpose in human lives: The emergence of identity as a life story. In A. I. Rabin, R. A. Zucker, R. A. Emmons, & S. Frank (Eds.), *Studying persons and lives* (pp. 148–200). New York: Springer.

McCrae, R. R. (1989). Why I advocate the five-factor model: Joint analyses of the NEO-PI and other instruments. In D. M. Buss & N. Cantor (Eds.), *Personality psychology: Recent trends and emerging directions* (pp. 237–245). New York: Springer-Verlag.

McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52,* 81–90.

McCrae, R. R., & Costa, P. T., Jr. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality. *Journal of Personality, 57,* 17–40.

McCrae, R. R., & Costa, P. T., Jr. (1990). *Personality in adulthood.* New York: Guilford.

McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *EdITS manual for the Profile of Mood States.* San Diego: EdITS.

Mischel, W. (1968). *Personality and assessment.* New York: Wiley.

Morey, L. C., Waugh, M. H., & Blashfield, R. K. (1985). MMPI scales for DSM-III personality disorders: Their derivation and correlates. *Journal of Personality Assessment, 49,* 245–251.

Murray, H. A. (1938). *Explorations in personality.* New York: Oxford University Press.

Mussen, P., Eichhorn, D. H., Honzik, M. P., Bieber, S. L., & Meredith, W. M. (1980). Continuity and change in women's characteristics over four decades. *International Journal of Behavioral Development, 3,* 333–347.

Myers, I. B., & McCaulley, M. H. (1985). *Manual: A guide to the development and use of the Myers-Briggs Type Indicator.* Palo Alto, CA: Consulting Psychologists Press.

Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology, 4,* 681–691.

Ostendorf, F. (1990). *Sprache und Persönlichkeitsstruktur: Zur Validität des*

Fünf-Faktoren-Modells der Persönlichkeit [Language and personality structure: Toward the validation of the five-factor model of personality]. Regensburg: S. Roderer.

Runyan, W. McK. (1990). Individual lives and the structure of personality psychology. In A. I. Rabin, R. A. Zucker, R. A. Emmons, & S. Frank (Eds.), *Studying persons and lives* (pp. 10–40). New York: Springer.

Schaie, K. W. (1977). Quasi-experimental research designs in the psychology of aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (1st ed., pp. 39–69). New York: Van Nostrand Reinhold.

Schuerger, J. M., Zarrella, K. L., & Hotz, A. S. (1989). Factors that influence the temporal stability of personality by questionnaire. *Journal of Personality and Social Psychology, 56*, 777–783.

Sears, R. R., & Sears, P. S. (1982). Lives in Berkeley [review of *Present and past in middle life*]. *Contemporary Psychology, 27*, 925–927.

Shock, N. W., Greulich, R. C., Andres, R., Arenberg, D., Costa, P. T., Jr., Lakatta, E. G., & Tobin, J. D. (1984). *Normal human aging: The Baltimore Longitudinal Study of Aging* (NIH Publication No. 84-2450). Bethesda, MD: National Institutes of Health.

Siegler, I. C., George, L. K., & Okun, M. A. (1979). Cross-sequential analysis of adult personality. *Developmental Psychology, 15*, 350–351.

Siegler, I. C., Zonderman, A. B., Barefoot, J. C., Williams, R. B., Jr., Costa, P. T., Jr., & McCrae, R. R. (1990). Predicting personality in adulthood from college MMPI scores: Implications for follow-up studies in psychosomatic medicine. *Psychosomatic Medicine, 52*, 644–652.

Spielberger, C. D., Jacobs, G., Crane, R., Russell, S., Westberry, L., Barker, L., Johnson, E., Knight, J., & Marks, E. (1979). *Preliminary manual for the State-Trait Personality Inventory (STPI)*. Tampa: University of Florida Human Resources Institute.

Tupes, E. C., & Christal, R. E. (1961). *Recurrent personality factors based on trait ratings* (USAF ASD Technical Report No. 61-97). Lackland Air Force Base, TX: U.S. Air Force.

Vaillant, G. E., & Vaillant, C. O. (1990). Natural history of male psychological health, XII: A 45-year study of predictors of successful aging at age 65. *American Journal of Psychiatry, 147*, 31–37.

White, R. W. (Ed.). (1964). *The study of lives: Essays on personality in honor of Henry A. Murray*. New York: Atherton.

Wiggins, J. S., & Trapnell, P. D. (in press). Personality structure: The return of the Big Five. In S. R. Briggs, R. Hogan, & W. H. Jones (Eds.), *Handbook of personality psychology*. New York: Academic Press.

Wiggins, J. S., Trapnell, P., & Phillips, N. (1988). Psychometric and geometric characteristics of the Revised Interpersonal Adjective Scales (IAS-R). *Multivariate Behavioral Research, 23*, 119–134.

Zuckerman, M. (1979). *Sensation seeking: Beyond the optimal level of arousal*. Hillsdale, NJ: Erlbaum.