

Age and Gender Differences in the Content Scales of the Minnesota Multiphasic Personality Inventory

Alan B. Zonderman

Gerontology Research Center, National Institute on
Aging, Baltimore, Maryland

**Ilene C. Siegler, John C. Barefoot, and Redford
B. Williams**

Behavioral Medicine Research Center, Duke University
Medical Center, Durham, North Carolina

Paul T. Costa, Jr.

Gerontology Research Center, National Institute on
Aging, Baltimore, Maryland

We examined time of measurement, gender, and age differences on the nine content scales of the Minnesota Multiphasic Personality Inventory using data collected by three separate studies during the 1950s, 1960s and 1980s. No evidence was found for differences in the content scales due to time of measurement that also could not have been explained by demographic differences. Differences due to gender were found on only one of the nine scales, Masculinity-Femininity, and age differences were found on the Neuroticism, Extraversion, and Agreeableness scales. Younger men and women had significantly higher scores on the Neuroticism and Extraversion scales, and these differences were consistent in both magnitude and direction across sample and gender. Our results suggest that it is likely that openness reaches its lifetime stable level by the time typical adolescents enter college, because we found no significant age differences in intellectual interests. Neuroticism,

We thank Dr. Richard B. Shekelle for providing the Western Electric data and for helpful comments on drafts of this article. We also thank W. Grant Dahlstrom and Stephen J. Almada for their comments. The University of North Carolina at Chapel Hill Alumni Heart Study was supported by Grant AG-09276 from the National Institute on Aging.

Address correspondence to Alan B. Zonderman, Gerontology Research Center, 4940 Eastern Avenue, Baltimore, MD 21224.

extraversion, and agreeableness on the other hand, are likely to show instability throughout, and probably after, adolescence and early adulthood, because we found significant age differences in the content dimensions associated with these factors in separate analyses of three samples.

A growing body of research supports the conclusion of stability in adult personality (e.g., Costa & McCrae, 1988; McCrae & Costa, 1984). The weight of the evidence for personality stability suggests that there are only negligible changes after age 30, but what changes, if any, occur during the late teens and 20s? Does personality change during late adolescence through early adulthood? If so, at what rate, and does it do so at the same rates for young men and young women? Are there particular developmental segues during which personality adapts to new roles, or are there transitions during which adolescents' and young adults' personalities become more clearly differentiated? Does stability begin at some easily characterized moment, or do the various facets of personality coalesce at different rates?

There are few studies with data relevant to these issues. Some researchers have examined college-age students, but there have been few longitudinal studies that begin with adolescence. For example, Zuckerman (1979) found that recklessness and sensation seeking decline after adolescence, and Bachman, O'Malley, and Johnston (1978) found that self-esteem increases in boys from adolescence to adulthood.

Other researchers have attempted follow-up studies of college students but failed to assess personality at the initial time of measurement. Still others have reconstructed cohort studies, which often do not cover the interval of interests with sufficient detail to answer questions about age changes. Even rarer are personality data on age differences unconfounded by the influences of time of measurement and gender differences. Finn (1986) found significant Age \times Cohort interactions for neuroticism, extraversion, and intellectual interests in a study of two male cohorts. Helson and Moane (1987) also found significant age differences between the 20s and 40s using the California Personality Inventory and Gough's Adjective Checklist in a longitudinal study of women.

Longitudinal correlations in a subset of subjects from the University of North Carolina at Chapel Hill (UNC) Alumni Heart Study supported both stability and change from late adolescence through middle age (Siegler et al., 1990). A significant correlation ($r = .39$) was found for the Cook-Medley Hostility scale over 20 years with an estimated stability of

.46. Longitudinal correlations between Minnesota Multiphasic Personality Inventory (MMPI) content scales and NEO Personality Inventory (NEO-PI) factors were about half as large as concurrent correlations, suggesting that about half the variance in the basic dimensions of personality is stable from college age to middle age.

In the present study, we addressed some of these questions in analyses of archival data from three separate studies in which the MMPI was administered. These studies were combined because they represent a sufficiently complete range of conditions for answering questions about personality differences associated with time of measurement, age, and gender effects. Although the MMPI is not necessarily the best choice for examining personality, it provides one of the few sources of archival data for addressing these issues in community-dwelling, healthy subjects.

We examined time of measurement, gender, and age differences in nine MMPI content dimensions using data collected in three separate studies during the 1950s, 1960s and 1980s. Time of measurement was examined by comparing men tested in the 1950s with men tested in the 1980s. Gender effects were examined by making separate comparisons of men and women tested contemporaneously in the 1960s and 1980s. Age effects were examined by comparing college-age men tested in the 1960s with adult men tested in the 1950s.

These analyses were designed to contrast specific effects by comparing samples selected so as to eliminate confounding. Because differences due to age could be associated with differences due to time of measurement, we examined time of measurement by comparing adult men measured in different decades. Similarly, we eliminated the potentially confounding effects of time of measurement and age from the gender analyses by comparing men and women of approximately the same ages who were tested in the same decade.

METHOD

Subjects

Subjects from three separate samples differing in time of measurement and age were examined. The general characteristics of these samples are summarized in Table 1. The first sample consisted of participants in the Baltimore Longitudinal Study of Aging (BLSA), a community-dwelling, generally healthy group of volunteers who have agreed to return for medical and psychological testing at fixed intervals (Shock et al., 1984). The sample has been recruited continuously since 1958, and the majority are white; work in or are retired from scientific, professional, or managerial positions; graduated from high school or college; and are married.

Table 1. Characteristics of the Three Samples

Sample	No. of subjects		Status		Description
	Men	Women	Time of 1st Measurement	Age	
BLSA	318	163	1981-1988	30-89	Volunteer participants in a longitudinal study
UNC	3,327	395	1964-1967	17-21	First year students entering UNC
WE	2,050	none	1957-1958	40-56	Participants in a prospective study of risk factors for coronary heart disease in WE employees

Note. BLSA = Baltimore Longitudinal Study of Aging, UNC = University of North Carolina at Chapel Hill, WE = Western Electric.

Although not representative of the general population in education or occupation, comparison with a national probability sample (National Health and Nutrition Examination Survey Epidemiological Follow-up Study) suggests that the BLSA sample does not differ markedly on the personality dimensions Neuroticism, Extraversion, and Openness to Experience (Costa et al., 1986). Because studies of women were introduced in 1978, fewer women were participants when these data were collected. For the purposes of the present study, data were available on 481 participants (318 men and 163 women) tested between 1981 and 1988. Subjects ranged in age from 30 to 89 ($M = 58.2$, $SD = 14.8$). The age range for men was 30–89 ($M = 59.0$, $SD = 14.6$), and that for women was 30–86 ($M = 56.8$, $SD = 15.1$).

The second sample consisted of entering freshman at the University of North Carolina at Chapel Hill (UNC), who were examined as part of a routine normative study of the MMPI during the 1960s (Siegler et al., 1992). For the present purposes, data were available on 3,722 students (3,327 men and 395 women) tested between 1964 and 1967. Subjects ranged in age from 17 to 21 ($M = 18.3$, $SD = 0.4$). The average age for men was 18.4 ($SD = 0.4$), and the average age for women was 18.3 ($SD = 0.5$). This sample closely matched the demographic characteristics of the BLSA in that it was generally white (99%) and was made up largely of college graduates (96%).

The third sample consisted of randomly selected male employees of the Western Electric Company's (WE) Hawthorne Works who were examined during the late 1950s as part of a prospective study of risk factors for coronary heart disease (Paul et al., 1963). The sample included men from occupations related to the manufacture of telephone equipment: 32% were in professional, technical, managerial, or clerical positions; 42% were foremen or skilled craftsmen; and 26% were service workers or laborers. The details of the sampling procedure and test administration were described by Paul et al. (1963). For the present study, data were available on 2,050 men ranging in age from 40 to 56 ($M = 47.5$, $SD = 4.4$) tested between 1957 and 1958.

Measures

The booklet form of the MMPI (Hathaway & McKinley, 1940) was administered to subjects from the UNC and WE samples. The BLSA participants were administered the MMPI using a computer-assisted testing program. Scores for the nine content dimensions (Costa, Zonderman, McCrae, & Williams, 1985) were computed from their scoring keys. These dimensions summarize the psychological content represented by the complete MMPI and include

- Neuroticism—a measure of worry, emotional instability, and psychological depression;
- Psychoticism/Infrequency—a measure of bizarre thinking, paranoid ideation, and items infrequently endorsed (e.g., “Evil spirits possess me at times” and “I feel uneasy indoors”);
- Religious Orthodoxy—a measure of fundamentalist beliefs, strict adherence to religious and moral rules, and attendance at religious services;
- Extraversion—a measure of wanting the company of others, preferring not to be alone, and enjoying crowds and parties;
- Cynicism—a measure of distrusting and disparaging attitudes toward the motives of others and a belief in the selfishness of human nature;
- Somatic Complaints—a measure of tiredness, stomach trouble, and headaches;
- Inadequacy—a measure of shyness and feelings of incompetence;
- Intellectual Interests—a measure of interests in poetry, dramatics, and science; and
- Masculinity–Femininity—a measure of stereotypical masculine interests (e.g., mechanics and sports) versus stereotypical feminine interests (e.g., dolls, dressmaking, and nursing).

All subjects omitted fewer than 20 of the 395 scored items. The BLSA participants had no missing items because the computer-assisted testing program would not permit missing responses. Less than 4% of the UNC and 2% of the WE subjects missed 11 or more items, 51% of the UNC and 33% of the WE subjects missed 1–10 items, and 45% of the UNC and 65% of the WE subjects had no missing items. Missing items were replaced by item means computed separately by gender for each sample.

Analyses

Three separate sets of analyses were performed on the nine content scales to examine differences due to time of measurement, gender, and age. Time-of-measurement effects were examined by comparing BLSA men (tested in the 1980s) with WE men (tested in the 1950s). This analysis was not confounded by the effects of gender or by the effects of age because both samples consisted of only men and the age means and ranges in both samples were nearly identical.

Gender differences were examined in two separate analyses by comparing men and women within the UNC sample and by comparing men and women within the BLSA sample. This analysis was not confounded by the effects of time of measurement or by the effects of age because men and women within each sample were roughly the same age and were tested at the same times. Analyses were performed in both samples sep-

arately to ensure that gender differences were not specific to a particular cohort at a particular time of measurement.

Age differences were examined by comparing the WE men (ages 40–56) with the UNC men (ages 17–21). This analysis was not confounded by the effects of gender and, in the absence of cohort differences, was not confounded by the effects of time of measurement. Socioeconomic differences between the WE and UNC cohorts were impossible to eliminate, so additional age difference analyses were performed. Separate comparisons were made between the BLSA men (ages 30–89) and UNC men (ages 17–21), and between the BLSA women (ages 30–86) and UNC women (ages 17–21) to ensure that age differences, if any, were not gender specific. The extent to which the results of the age difference analyses agreed was interpreted as indicating that cohort differences between the WE and UNC men were unimportant in the age difference findings.

A multivariate analysis of variance (MANOVA) was used because the nine content scales were correlated. For each analysis, significant differences on each of the nine scales were not interpreted unless the overall multivariate test statistic was significant. Because the sample sizes for some of the analyses were so large, a p of less than .01 was required for significance. Power analyses (Cohen, 1969; Muller & Peterson, 1984) indicated that the sample sizes in the present study were large enough to ensure statistically reliable findings using $p < .01$ as the cutoff for statistical significance. Power exceeded 0.99 in all of our analyses, indicating that our sample sizes were sufficiently large to detect small effects if they were present.

RESULTS

Time-of-Measurement Effects

Table 2 shows the within-sex scale means for each sample as well as the overall sample standard deviations. Differences in the content scales due to time of measurement were examined by a one-way MANOVA comparing the BLSA sample (tested in the 1980s) with the WE sample (tested in the 1950s). This contrast was not confounded by effects due to age differences, because both samples were restricted to adults over 30; nor was it confounded by gender differences, because both samples consisted of only men.

The overall one-way MANOVA was significant, $F(9, 2358) = 17.05$, $p < .001$, and the multivariate effect size, η^2 (Camp & Maxwell, 1983; Cohen, 1969; Haase, 1983), was 0.06. The significance of the multivariate test indicated that we could legitimately examine differences on each

Table 2. Scale Means by Sex and Sample and Complete Sample Standard Deviations

Scale	Women		Men		SD	
	BLSA (N = 163)	UNC (N = 395)	BLSA (N = 318)	UNC (N = 3327)		WE (N = 2050)
Neuroticism	21.94	29.57	18.79	25.53	17.71	12.7
Psychoticism/Infrequency	7.12	11.35	6.93	10.41	7.15	9.2
Religious Orthodoxy	11.45	12.29	10.26	10.45	11.47	4.5
Extraversion	13.28	15.59	13.70	16.24	13.23	4.6
Cynicism	11.72	12.57	13.29	15.48	14.26	6.8
Somatic Complaints	6.75	7.06	5.30	5.45	5.97	4.8
Inadequacy	10.31	11.72	9.23	9.87	8.79	5.8
Intellectual Interests	8.30	7.72	8.24	7.11	7.26	2.3
Masculinity-Femininity	17.59	18.03	27.36	28.48	28.67	5.0

Note. BLSA = Baltimore Longitudinal Study of Aging sample, UNC = University of North Carolina at Chapel Hill sample, WE = Western Electric sample.

Table 3. Mean Differences in Standard Deviation Units, Univariate Significance Levels (p), and Effect Sizes (η^2) for the Time-of-Measurement Contrast Between the WE and BLSA Samples

Scale	Mean SD difference ^a	η^2
Neuroticism	0.1	0.00
Psychoticism/Infrequency	0.0	0.00
Religious Orthodoxy	-0.3*	0.01
Extraversion	0.1	0.00
Cynicism	-0.1	0.00
Somatic Complaints	-0.1	0.00
Inadequacy	0.1	0.00
Intellectual Interests	0.4*	0.02
Masculinity-Femininity	-0.3*	0.01

Note. WE = Western Electric, BLSA = Baltimore Longitudinal Study of Aging.

^aThe difference is the BLSA standardized mean minus the WE standardized mean, so that positive differences indicate the 1980s cohort's (BLSA) mean was greater than the 1960s cohort's (WE) mean and negative differences indicate the 1960s cohort's mean was greater than the 1980s cohort's mean.

* $p < .01$.

of the nine content scales. Table 3 shows the mean differences between the samples in standard deviation units and univariate effect sizes (η^2) from separate one-way analyses of each of the nine content scales. The WE cohort had significantly higher scores on the Religious Orthodoxy and Masculinity-Femininity scales. The BLSA cohort scored significantly higher on the Intellectual Interests scale.

Although three of the nine content scales showed significant ($p < .01$) differences between the two samples, the effect sizes were quite small and none exceeded 0.02. Socioeconomic and demographic discrepancies between the two samples probably account for the differences that were found. The BLSA participants are highly educated, and many have been making or have made careers pursuing their intellectual interests, whereas subjects in the WE sample have on average a high school education. This probably accounts for the difference on the Intellectual Interests scale.

There were also socioeconomic and geographic differences between the WE and BLSA samples that may account for the differences in the Religious Orthodoxy and Masculinity-Femininity scales. Subjects in the BLSA sample are largely professionals from the Mid-Atlantic region, whereas the WE sample consists largely of blue-collar workers from the Chicago area. The BLSA participants are less likely to have traditional sex role interests and are less likely to hold orthodox religious views than subjects in the WE sample. However, these relatively small differences

may also represent changes from the 1950s to the 1980s in attitudes toward traditional masculine roles and the importance of religious practices. In either case, the magnitudes of the differences between the WE and BLSA samples were quite small.

The time-of-measurement analyses were repeated using only BLSA data from the 141 men ages 30–58, the age range of the WE participants. This reanalysis was performed to determine whether the absence of large differences between the WE and BLSA samples was due to the greater age range in the BLSA sample. The results were unchanged in the reanalysis. This suggests that differences that were unrelated to socioeconomic effects between adult men tested in the 1950s and adult men tested in the 1980s were not an important source of variance in the content scales.

Gender Effects

Gender differences were examined by two separate one-way MANOVAs comparing men and women within the UNC sample and men and women within the BLSA sample. These contrasts were not confounded by age differences, because the men and women were roughly the same ages within each sample. These contrasts were also not confounded by time of measurement, because men and women within each sample were tested at the same times. The overall multivariate differences for gender were significant in both one-way MANOVAs: For the UNC sample, $F(9, 3712) = 338.55, p < .001, \eta^2 = 0.45$; for the BLSA sample, $F(9, 471) = 70.44, p < .001, \eta^2 = 0.57$.

Table 4 shows the mean differences between men and women within each sample and the effect sizes from separate one-way tests on each of the content scales. In both samples, women had significantly higher scores on Neuroticism and Somatic Complaints, and men scored significantly higher on Masculinity–Femininity. In the younger sample (UNC), women had significantly higher scores on Religious Orthodoxy, Inadequacy, and Intellectual Interests, and men had significantly higher Cynicism scores. Masculinity–Femininity was the only scale in both analyses that had a large effect size, reflecting the stereotypic masculine interests that it assesses. Because of these gender differences, in subsequent analyses men and women were examined separately.

Age Effects

Age differences in the nine content dimensions were examined in three separate one-way MANOVAs. The first analysis compared men from the WE (older) sample with men from the UNC (younger) sample.

Table 4. Mean Differences in Standard Deviation Units, Univariate Significance Levels (p), and Effect Sizes (η^2) for Separate Gender Contrasts in the BLSA and UNC Samples

Scale	BLSA ^a		UNC ^b	
	Mean <i>SD</i> difference ^c	η^2	Mean <i>SD</i> difference ^c	η^2
Neuroticism	-0.2*	0.02	-0.3*	0.01
Psychoticism/Infrequency	0.0	0.00	-0.1	0.00
Religious Orthodoxy	-0.3	0.01	-0.5*	0.02
Extraversion	0.1	0.00	0.1	0.00
Cynicism	0.2	0.01	0.4*	0.02
Somatic Complaints	-0.3*	0.02	-0.3*	0.01
Inadequacy	-0.2	0.01	-0.3*	0.01
Intellectual Interests	0.0	0.00	-0.2*	0.01
Masculinity-Femininity	1.9*	0.54	2.2*	0.42

^aBaltimore Longitudinal Study of Aging sample.

^bUniversity of North Carolina at Chapel Hill sample.

^cDifference is the men's standardized mean minus the women's standardized mean, so that positive differences indicate that the men's mean was greater than the women's mean and negative differences indicate that the women's mean was greater than the men's mean.

* $p < .01$.

In the second two analyses, we separately compared men and women from the BLSA (older) sample with men and women from the UNC (younger) sample. Neither of the two comparisons was confounded by gender differences, but they could have been confounded by time-of-measurement differences.

For men, the overall multivariate age differences were significant in both one-way MANOVAs: For WE vs UNC, $F(9, 5367) = 256.91, p < .001, \eta^2 = 0.30$; and for BLSA vs. UNC, $F(9, 3635) = 48.77, p < .001, \eta^2 = 0.11$. These significant multivariate effects indicated that we could legitimately examine differences on each of the nine content scales. Table 5 shows the mean differences in standard deviation units for both of the men's analyses. In both analyses, the younger sample (UNC) scored significantly higher on the Neuroticism, Psychoticism/Infrequency, Extraversion, and Cynicism scales. The younger sample scores significantly higher on the Inadequacy scale only in the WE and UNC comparison. The older subjects scored significantly higher on Religious Orthodoxy and Somatic Complaints only in WE and UNC comparison. Older subjects scored significantly higher on the Intellectual Interests scale and younger subjects scored significantly higher on the Masculinity-Femininity scale in the BLSA and UNC comparison. Large effect sizes

were found for the age differences between the WE and UNC samples on Neuroticism ($\eta^2 = 0.09$), Psychoticism/Infrequency ($\eta^2 = 0.4$), and Extraversion ($\eta^2 = 0.11$).

The overall multivariate age difference was also significant for the comparison between women in the BLSA (older) and UNC (younger) samples, $F(9, 548) = 21.78, p < .001, \eta^2 = 0.26$. As shown in Table 6, the younger (UNC) women scored significantly higher on the Neuroticism, Psychoticism/Infrequency, Religious Orthodoxy, and Extraversion scales than did older (BLSA) women; the older women scored significantly higher on the Intellectual Interests scale. Large effect sizes were found for the differences on the Neuroticism ($\eta^2 = 0.07$), Psychoticism/Infrequency ($\eta^2 = 0.05$), and Extraversion ($\eta^2 = 0.06$) scales, the three scales on which younger men had significantly larger scores than older men.

Supplementary Analyses

An additional 2,103 UNC students ages 17–21 (432 women and 1,671 men) were administered MMPI inventory forms missing at least 85 of the 566 MMPI items. These subjects were eliminated from our analyses

Table 5. Mean Differences in Standard Deviation Units, Univariate Significance Levels (p), and Effect Sizes (η^2) for the Separate Age Contrasts Between Men in the WE and UNC Samples and Between Men in the BLSA and UNC Samples

Scale	WE ^a vs. UNC ^b		BLSA ^c vs. UNC	
	Mean <i>SD</i> difference ^d	η^2	Mean <i>SD</i> difference ^d	η^2
Neuroticism	-0.6*	0.09	-0.5*	0.02
Psychoticism/Infrequency	-0.4*	0.04	-0.4*	0.01
Religious Orthodoxy	0.2*	0.01	0.0	0.00
Extraversion	-0.7*	0.11	-0.6*	0.02
Cynicism	-0.2*	0.01	-0.3*	0.01
Somatic Complaints	0.1*	0.00	0.0	0.00
Inadequacy	-0.2*	0.01	-0.1	0.00
Intellectual Interests	0.1	0.00	0.5*	0.02
Masculinity-Femininity	0.0	0.00	-0.2*	0.01

^aWestern Electric sample.

^bUniversity of North Carolina at Chapel Hill sample.

^cBaltimore Longitudinal Study of Aging Sample.

^dThe difference is the older standardized mean minus younger standardized mean, so that positive differences indicate that the mean for the older sample was greater than that for the younger sample and negative differences indicate that the mean for the younger sample was greater than that for the older sample.

* $p < .01$.

Table 6. Mean Differences in Standard Deviation Units, Univariate Significance Levels (p), and Effect Sizes (η^2) for the Age Contrast Between Women in the BLSA and UNC Samples

Scale	Mean SD difference ^a	η^2
Neuroticism	-0.6*	0.07
Psychoticism/Infrequency	-0.5*	0.05
Religious Orthodoxy	-0.3*	0.02
Extraversion	-0.6*	0.06
Cynicism	-0.1	0.00
Somatic Complaints	0.0	0.00
Inadequacy	-0.2	0.01
Intellectual Interests	0.3*	0.02
Masculinity-Femininity	0.0	0.00

Note. BLSA = Baltimore Longitudinal Study of Aging, UNC = University of North Carolina at Chapel Hill

^aThe difference is the older women's standardized mean minus the younger women's standardized mean, so that positive differences indicate that the mean for the older sample was greater than that for the younger sample and negative differences indicate that the mean for the younger sample was greater than that for the older sample.

* $p < .01$.

because 64 of the 395 content dimension items were among the missing 85 items. Shortened versions of the content scales were computed using the 331 items that were administered to all UNC students, and the data were reanalyzed. The correlations between the complete scales and their shortened versions were all greater than 0.95. All of the MANOVAs yielded results that replicated our findings on the complete scales. Although the mean values of the shortened scales were different from those of the complete scales, the standardized mean differences were nearly indistinguishable. All of the significant differences reported in Tables 3-6 were replicated by the analyses of the shortened scales, as were the magnitudes of the effect sizes. These results suggest that for contrasts that involved the UNC sample, none of the results were nonsignificant as a result of too few subjects or because there was a restricted range of variation. These results also suggest that the shortened versions could be used as adequate representatives of the content dimension scales when only an incomplete MMPI is available.

DISCUSSION

Analyses of the MMPI by Costa et al. (1985) indicated that it comprises nine content dimensions, seven of which have been classified in terms of the five-factor model of personality (Costa, Busch, Zonderman, & McCrae, 1986). Derived from theoretical and empirical taxonomies based

on natural language trait terms, the five-factor model provides an exhaustive framework for organizing the dimensions of individual differences in personality. These factors have been found in a variety of studies of ratings and inventories in children and adults (Digman, 1989). Although different investigators may disagree on their exact names and interpretations, these factors have been identified in a variety of personality instruments (Digman & Takemoto-Chock, 1981; McCrae & Costa, 1987). Few investigators disagree, however, that the five-factor model is a comprehensive model for describing individual differences in personality in systematic analyses over the life span.

Of the seven content scales that have been classified in terms of the five-factor model of personality (Costa, Busch, Zonderman, & McCrae, 1986), four (Neuroticism, Psychoticism/Infrequency, Somatic Complaints, and Inadequacy) correspond to the Neuroticism factor, the Extraversion scale corresponds to the Extraversion factor, the Intellectual Interests scale corresponds to the Openness factor, and the Cynicism scale corresponds to the Neuroticism and Agreeableness factors. Religious Orthodoxy and Masculinity–Femininity have no counterparts in the five-factor model and are unclassified. None of the nine content dimensions corresponds to the Conscientiousness factor.

This conceptual reorganization of the nine content dimensions provides a useful way to understand our results. We found little evidence for differences in the content dimensions due to time of measurement that could not also have been due to demographic or socioeconomic differences. Furthermore, the two scales on which we found significant, albeit small, differences due to time of measurement (Religious Orthodoxy and Masculinity–Femininity) are unclassified in the five-factor model of personality. These results suggest that time of measurement by itself is probably unimportant, particularly for scales that measure one of the dimensions of the five-factor model of personality.

We found consistent evidence for gender differences in two samples on three of the nine content scales, Neuroticism, Psychoticism/Infrequency, and Masculinity–Femininity. In the younger sample, we also found gender differences on the Religious Orthodoxy, Cynicism, Inadequacy, and Intellectual Interests scales, although the magnitudes of the differences were small. Gender differences on the Cynicism scale were consistent with previous findings (Barefoot, Beckham, Haney, Siegler, & Lipkus, 1993; Barefoot *et al.*, 1991). These results suggest that there are gender differences on Neuroticism in both young and old samples, but, except for Extraversion, differences in the other dimensions of the five-factor model may be reserved for younger ages.

Masculinity–Femininity was the only scale on which there was a gender difference with a large effect size. This result is uninteresting in the

present context because it only confirms the discriminant validity of a scale that is unclassified in terms of the five-factor model. However, the pervasive gender differences, although small in most cases, suggest that each gender should be examined separately in future analyses.

We also found evidence for age differences in the content scales that correspond to the Neuroticism and Extraversion factors in the five-factor model of personality. Younger men and women had significantly higher scores on the Neuroticism and Extraversion scales, and these differences were consistent in both magnitude and direction across sample and across sex. These age differences are consistent with the small, although significant, age differences found in a nationally representative sample (Costa et al., 1986). In that sample, younger subjects (in their 30s) also had significantly higher Neuroticism and Extraversion scores. The present results replicate and extend those results to younger ages (17–21). They also suggest that change in neuroticism and extraversion probably decelerates steadily during the 20s.

In a longitudinal study of two cohorts of men, Finn (1986) also found evidence for age differences in the MMPI. Although his samples were considerably smaller than ours and not as well differentiated by age, he found interactions between age and cohort in several scales related to neuroticism and extraversion. Contrary to our findings, he also found a significant Age \times Cohort interaction for intellectual interests. It is difficult to draw exact analogies between our study and Finn's because he was interested primarily in differences in stability across cohorts. Because his cohorts overlapped in age, and because one cohort represented college students and the other represented businessmen, it is likely that the Age \times Cohort interaction for intellectual interests was due to differences in the cohorts, and not differences in their ages. In this way, his results parallel the differences we found between the BLSA (older) and the UNC (younger) samples.

Using the California Personality Inventory and Gough's Adjective Check List in a longitudinal study of women, Helson and Moane (1987) also found significant age differences between the 20s and 40s. Interestingly, both Finn (1986) and Helson and Moane found significant age differences as well as significant stability over time. These seemingly contradictory findings suggest that there are consistent changes in mean levels over time that also preserve the relative rank orderings of individuals within age groups. From these data, it appears that predictable patterns of personality change occur with aging and that the magnitudes of these differences are relatively small.

The finding of higher Cynicism scores in the UNC sample of young adults is consistent with findings in previous research for measures reflecting cynicism and aggressiveness (Barefoot et al., 1991). Much of

this work has been conducted with a hostility scale from the MMPI that is related to the Cynicism content dimension (Cook & Medley, 1954). In a test of age differences in hostility, this scale was administered to 4,710 members of the UNC sample in 1987–1990, approximately 22 years after the initial administration of the MMPI (Siegler et al., 1992). A decrease in mean level of hostility (approximately half of a standard deviation) was observed in this longitudinal analysis.

Our results suggest that openness reaches its lifetime stable level by the time typical adolescents enter college, because we found no significant age differences in intellectual interests. Neuroticism, extraversion, and agreeableness, on the other hand, are likely to show instability throughout, and probably after, adolescence and early adulthood, because we found significant age differences in the content dimensions associated with these factors in separate analyses of three samples.

Unfortunately, we cannot use our data to determine when stability in these personality dimensions is achieved, and no conclusions can be drawn regarding conscientiousness. At the very least, we need a sufficiently large sample in their 20s to compare groups with smaller age ranges. It would be even better to have longitudinal personality data starting with adolescence, especially if assessments were made using a personality inventory designed to measure all of the dimensions of the five-factor model.

REFERENCES

- Bachman, J. G., O'Malley, P. M., & Johnston, J. (1978). *Adolescence to adulthood: Change and stability in the lives of young men*. Ann Arbor, MI: Institute for Social Research.
- Barefoot, J. C., Beckham, J. C., Haney, T. L., Siegler, I. C., & Lipkus, I. M. (1993). Age differences in hostility among middle-aged and older adults. *Psychology and Aging, 8*, 3–9.
- Barefoot, J. C., Peterson, B. L., Dahlstrom, W. G., Siegler, I. C., Anderson, N. B., & Williams, R. B. (1991). Hostility patterns and health implications: Correlates of Cook-Medley hostility scale scores in a national survey. *Health Psychology, 10*, 18–24.
- Camp, C. J., & Maxwell, S. E. (1983). A comparison of various strength of association measures commonly used in gerontological research. *Journal of Gerontology, 38*, 3–7.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. San Diego, CA: Academic Press.
- Cook, W. W., & Medley, D. M. (1954). Proposed hostility and pharisaic-virtue scales for the MMPI. *Journal of Applied Psychology, 38*, 414–418.
- Costa, P. T., McCrae, R. R., Zonderman, A. B., Barbano, H. E., Lebowitz, B., & Larson, D. M. (1986). Cross-sectional studies of personality in a national sample: 2. Stability in neuroticism, extraversion, and openness. *Psychological Aging, 1*, 144–149.

- Costa, P. T., Busch, C. M., Zonderman, A. B., & McCrae, R. R. (1986). Correlations of MMPI factor scores with measures of the five factor model of personality. *Journal of Personality Assessment*, *50*, 640-650.
- Costa, P. T., & McCrae, R. R. (1988). 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., Zonderman, A. B., McCrae, R. R., & Williams, R. B. (1985). Content and comprehensiveness in the MMPI: An item factor analysis in a normal adult sample. *Journal of Personality and Social Psychology*, *48*, 925-933.
- Digman, J. M. (1989). Five robust trait dimensions: Development, stability, and utility. *Journal of Personality*, *57*, 195-214.
- Digman, J. M., & Takemoto-Chock, N. K. (1981). Factors in the natural language of personality: Re-analysis, comparison and interpretation of six major studies. *Multivariate Behavioral Research*, *16*, 149-170.
- Finn, S. E. (1986). Stability for personality self-ratings over 30 years: Evidence for an age/cohort interaction. *Journal of Personality and Social Psychology*, *50*, 813-818.
- Haase, R. F. (1983). Classical and partial eta square in multifactor ANOVA designs. *Educational and Psychological Measurement*, *41*, 35-39.
- Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology*, *10*, 249-254.
- Helson, R., & Moane, G. (1987). Personality change in women from college to midlife. *Journal of Personality and Social Psychology*, *53*, 81-90.
- McCrae, R. R., & Costa, P. T., Jr. (1984). *Emerging lives, enduring dispositions*. Boston: Little, Brown.
- McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*, 81-90.
- Muller, K. E., & Peterson, B. L. (1984). Practical methods for computing power in testing the multivariate general linear hypothesis. *Computational Statistics and Data Analysis*, *2*, 143-158.
- Paul, O., Lepper, M. H., Phelan, W. H., Dupertuis, G. W., MacMillan, A., McKean, H., & Park, H. (1963). A longitudinal study of coronary heart disease. *Circulation*, *28*, 20-31.
- Shock, N. W., Greulich, R. C., Andres, R., Arenberg, D., Costa, P. T., Lakatta, E. G., & Tobin, J. D. (1984). *Normal Human Aging: The Baltimore Longitudinal Study of Aging* (NIH Publication No. 84-2450). Washington, DC: U.S. Government Printing Office.
- Siegler, I. C., Peterson, B. L., Barefoot, J. C., Harvin, S. H., Dahlstrom, W. G., Kaplan, B. H., Costa, P. T., Jr., & Williams, R. B., Jr. (1992). Using college alumni populations in epidemiologic research: The UNC Alumni Heart Study. *Journal of Clinical Epidemiology*, *45*, 1243-1250.
- 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.
- Zuckerman, M. (1979). *Sensation seeking: Beyond the optimal level of arousal*. Hillsdale, NJ: Erlbaum.