Evaluating the Appropriateness of College-Age Norms for Use With Adolescents on the NEO Personality Inventory–Revised

Alissa Sherry  
University of Texas at Austin

Robin K. Henson  
University of North Texas

Jean G. Lewis  
Austin Peay State University

The NEO Personality Inventory–Revised (NEO PI-R), developed by Costa and McCrae (1992), is based on the five-factor model of personality. The instrument is particularly useful in that it is a personality measure designed to measure normal personality traits. The NEO PI-R Form S is a self-rating scale composed of five broad domains with six, more specific, facets in each domain. Because of the emphasis on normality, neither pole of any domain or facet scale should be considered preferable in terms of mental health or societal standards. The domains, their definitions, and facets are listed below as defined by the NEO PI-R manual (Costa & McCrae, 1992):

- Neuroticism (N): Susceptibility to psychological distress, irrational ideas, poor impulse control. Facets: Anxiety (N1), Angry Hostility (N2), Depression (N3), Self-Consciousness (N4), Impulsiveness (N5), and Vulnerability (N6).
- Extroversion (E): Measures a dimension from sociable, assertive, cheerful, and energetic to shy, reserved, independent, and even paced. Facets: Warmth (E1), Gregariousness (E2), Assertiveness (E3), Activity...
(E4), Excitement Seeking (E5), and Positive Emotions (E6).

Openness to Experience (O): Imagination, intellect, curiosity, willingness to entertain novel and unconventional values. Facets: Fantasy (O1), Aesthetics (O2), Feelings (O3), Actions (O4), Ideas (O5), and Values (O6).


Conscientiousness (C): Control of impulses, planning organizing, ability to carry out a task. Facets: Competence (C1), Order (C2), Dutifulness (C3), Achievement Striving (C4), Self-Discipline (C5), and Deliberation (C6).

There has been extensive research regarding the content validity, convergent validity, and factor analytic loadings of the NEO PI-R (Costa & McCrae, 1992; McCrae & Costa, 1989; Piedmont & Weinstein, 1993). Overall, reviews of the NEO PI-R suggest that it produces moderate to strong correlations in directions predicted by its theoretical constructs (Tinsley, 1994).

In the normative sample, internal consistency reliability estimates ranged from .86 to .95. When internal consistencies were analyzed for facet scale scores, the alpha coefficients were between .56 and .81, with only 13 of the 30 facet scales falling below .70. Retest reliabilities on facet scores after 6 years fell between .66 and .92 (Costa & McCrae, 1992). Retest coefficients for N, E, and O domain scores after 6 years were between .69 and .83, and after 3 years, retest reliabilities for A and C domain scores were .63 and .79, respectively (Costa & McCrae, 1988).

Normative Data

Because the NEO PI-R is a well-recognized and well-used tool for the measurement of normal personality, it has incredible potential to be used with a variety of people in a variety of settings such as in school or university settings. The purpose of this research is to investigate whether the instrument in its current form can be used with age-groups other than those for which it was originally intended, specifically, an adolescent population. It is hoped that results can guide future revisions of the instrument as well as inform professionals of the appropriate usage of the instrument.

One of the distinguishing characteristics of the NEO PI-R that sets it apart from current personality inventories such as the Minnesota Multiphasic Personality Inventory (MMPI) is the fact that it was normed on normal individuals instead of clinical populations. The test authors’ rationale was that most people seeking therapy do so to seek guidance and help in adjusting to life’s everyday hardships and actually experience the same range of emotions and variation in personality as those who do not feel the need to seek therapy. They also argued that if psychologically abnormal profiles are present, they will be seen most clearly when plotted against a normal population (Costa & McCrae, 1992).

The final norm tables were constructed from 500 men and 500 women, 21 years of age and older using a careful process guided by 1995 U.S. census projections. After these norms were established, the authors indicated that although they believed personality traits to be stable in adulthood, the transition from adolescence to adulthood can vary. As a result, college norms were developed for students 17 to 20 years of age. Once testing was completed, the results indicated that college-age individuals scored higher on the N, E, and O domains and lower on the A and C domains than adults. In addition, they scored one standard deviation higher on the Excitement-Seeking facet of Extroversion compared to the adult norms (Costa & McCrae, 1992).

Another perspective argues against creating norm tables for separate populations (Widiger, 1992). This argument notes that profiles should be based on one normal population, and separate norms would underestimate the magnitude of the traits being measured. However, with regard to age-groups, appropriate normative data are essential when researching both groups and individuals. Separate norms can illustrate patterns of responding to obtain a general idea of what to expect from members of the group. Although they cannot predict any individual’s responses specifically, norms give an excellent indication of whether an individual deviates from his or her peer group and, if so, how much (Powell, 1963).

Using the NEO PI-R With Adolescents

When using the NEO PI-R with an adolescent population, it is important to assess whether the presently available norms are sufficient. Of course, the term adolescence must first be defined. Studies of adolescents have used varied age definitions for this developmental period, such as 11 to 14 years of age (Graziano & Ward, 1992), 12 to 16 years of age (Diaz, Belena, & Baguena, 1994), Grades 7 to 9 (Wong & Lau, 1993), 16 to 19 years of age (Gerstein & Briggs, 1993), and 11 to 21 years of age (Hatzichristou & Papadatos, 1993). Obviously, researchers have had a variety of ideas concerning the onset and cessation of adolescence.

Because the NEO PI-R is an instrument concerned with nonpsychiatric personality characteristics, another issue in adolescent research focuses on establishing a baseline
Personality Stability Through the Life Span

The second area to consider in assessing norm appropriateness is whether these adolescent characteristics remain stable over time or if there are specific personality traits that change as one moves into adulthood. The researchers of the NEO PI-R hold the view that many aspects of personality remain highly stable throughout adulthood. They conducted several studies to support this theory. The results appeared to support the premise that personality shows little change across adulthood (Costa & McCrae, 1988) and added that the slight differences that do appear at any given time are a result of measurement error, not personality change (McCrae, 1993).

Many of the studies pertaining to personality stability and change have shown to be useful. In a longitudinal study, two findings were of importance. Using the Bentler Psychological Inventory (Bentler & Newcomb, 1978), participants with a beginning age of 13 to 15 were measured three times: Year 1, Year 5, and Year 9. Supporting the research of Costa and McCrae (1988) and McCrae (1993), personality appeared somewhat stable between the Year 5 and the Year 9 measurements. However, some variability was present between the first measurement and the measurement at Year 5. The authors concluded that there is some fluctuation in personality during adolescence (Stein, Newcomb, & Bentler, 1986). Other studies confirm the existence of both stability and change (Eccles et al., 1989).

With a wealth of new experiences in adolescence such as high school, dating, driving, and increased independence, it is possible these experiences may have an effect on personality stability. Costa and McCrae (1992) indicated that there are some changes during the life span by the creation of separate college norm tables on the NEO PI-R and clearly stated they believe there to be significant changes between adolescence and early adulthood. Before the NEO PI-R should be used in clinical or research applications with adolescents, a determination needs to be made regarding its generalizability to this population.

The purpose of this research was to collect data from adolescent and college-age samples and compare the two groups’ domain and facet scores of the NEO PI-R. Differences may indicate the need for more extensive research and possibly the development of adolescent norm tables for the NEO PI-R. On the basis of the research suggesting some personality instability during adolescence, we hypothesized differences between adolescent and college groups on both the facet and domain scales, although we did not predict the direction of these differences.

METHOD

Participants

Adolescent participants consisted of 79 individuals (38 girls and 41 boys, mean age 13.7 years, range 13-15) from a middle school in Tennessee. College-age participants included 80 undergraduate volunteers (53 women and 27 men, mean age 19 years, range 18-20) from psychology (n = 57) and leadership (n = 23) courses at a state university in Tennessee. Demographically, this university was believed to be comparable to the middle school population except for age and education level. Ethnicity was reasonably comparable between the groups. Of the adolescents, 75 were White and 4 were Black. Of the college-age participants, 61 were White, 13 were Black, 2 were Asian American, 2 were Latin American, and 2 were Native American.

Because we sought to evaluate the appropriateness of the NEO PI-R college norms for use with adolescents, it is also relevant to evaluate comparability of our current sample of college students with the college normative sample.
used for the instrument. The two samples were comparable on both gender (current sample: 66.3% female, 33.8% male; normative sample: 62.0% female, 33.0% male) and ethnicity (current sample: 76.3% White, 16.3% Black, 7.5% other; normative sample: 85.4% White, 11.2% Black, 3.4% other). No other descriptive information was given in the manual for the normative college sample (Costa & McCrae, 1992); therefore, other comparisons were not possible. Nevertheless, these comparisons suggest that our current sample reasonably mirrored the normative sample demographically.

**Instrument**

The NEO PI-R Form S (Costa & McCrae, 1992) was used to assess differences in normal personality between the two groups. Scores on each of the facets range from 0 to 32. Facet scores can be aggregated to yield domain (N, E, O, A, C) scores ranging from 0 to 192. Higher scores indicated more of the personality characteristic being measured. The instrument has a total of 240 statements. Reading level for the instrument is stated in the manual as sixth grade. Participants rate the statements on a 5-point Likert-type scale, anchored at strongly disagree, disagree, neutral, agree, and strongly agree.

**Procedures**

Data were collected in group format. The first author was available during the sessions to answer any questions concerning definitions or wording. Any protocol with missing data resulted in elimination of that participant’s scores from analysis.

**RESULTS**

Domain-level coefficient alphas were computed using the facet-level responses for the adolescent, college, and combined groups. Table 1 presents the reliability estimates. Coefficient alphas for the normative sample were found to be .92, .89, .87, .86, and .90 for N, E, O, A, and C, respectively (Costa & McCrae, 1992).

Beyond demographics, we also compared our college sample means with the normative means. If these differences were substantial, then any observed differences between adolescents and college-age individuals in the current study may be more sample related rather than age related. To make these comparisons (n = 35—across all domains and facets within each domain), we calculated the Z scores associated with our college sample means using the normative means and standard deviations combined across genders in Table B-3 from the NEO PI-R manual (Costa & McCrae, 1992, p. 77). These Z scores were then converted to T score form. Standardized mean difference effect sizes were then calculated in T score metric (i.e., [sample T – 50] / 10) to yield the differences between the current sample means and the college norms in standard deviation units (where \( SD = 10 \)). The average of these effects was near zero with minimal variation (\( M = .02, SD = .17 \)). Furthermore, the only substantial effect (–.48) was for Values (O6), which had minimal contribution to the current adolescent and college mean differences observed later in this study. These results again suggest reasonable comparability between our college sample and the normative sample given in the NEO PI-R manual.

To examine differences between the adolescent and college samples on the NEO PI-R facet and domain scores, a series of descriptive discriminant analyses (DDAs) were conducted (Huberty, 1994; Klecka, 1980). DDA allows the researcher to define the degree of group distributional separation within a multivariate context. Because the personality domains and facets measured by the NEO PI-R can be conceived of as existing simultaneously within individuals as a matter of degree, we chose to model this assumption with multivariate analysis rather than univariate approaches that would treat each facet and domain separately. DDA is analogous to MANOVA (and other general linear model analyses) and is used here due to DDA’s ease of interpreting variable importance via examination of the standardized discriminant function coefficients and structure coefficients (cf. Courville & Thompson, 2001; Huberty, 1994; Thompson & Borrello, 1985). DDAs were conducted for each of five sets of facet scores. Another DDA was conducted using the aggregated domain scores as the dependent variables.

Table 2 presents descriptive statistics for the adolescent and college samples’ scores on the six facets across each domain. Table 2 also presents descriptive statistics for the aggregated domain-level scores. On average, the college sample tended to score higher than the adolescent sample, particularly on the E, A, and C facet and domain scales, with higher means in 28 of 35 total comparisons. On the A
and C scales, the adolescent means were uniformly lower than the college sample means.

Five DDAs were conducted to determine whether the six facet scores for each of the five domains could define adolescent and college group differences. Another DDA was conducted to determine whether the five aggregate domain scores could account for group differences. Table 3 presents results from these analyses. Using a .05 criterion, all analyses indicated statistically significant group differences. Squared canonical correlation effect sizes also suggested practical differences for each of the analyses, with the largest effects observed for the N ($R^2_c = 21.72\%$) and C ($R^2_c = 16.83\%$) facets and the domain-level ($R^2_c = 15.29\%$) analysis.

Standardized discriminant function coefficients and structure coefficients were examined to determine what variables contributed to the observed group differences. Function coefficients are directly analogous to beta weights in the regression context. However, as noted by Courville and Thompson (2001), structure coefficients are also critical when interpreting variable importance due to the impact of multicollinearity on function coefficients. Structure coefficients represent the observed variable’s correlation with the latent discriminant function variable.

Table 4 presents both sets of coefficients for all analyses. For the N facets, N2 and N6 were primarily responsible for the group differences. N3 had a large function coefficient (−.741), but its structure coefficient (−.145) indicated the facet was poorly correlated with the synthetic function variable. Adolescent and college group differences were largely defined by E5 and E6 for the E facets and O1 and O3 for the O facets. For A and C, group differences were more diversely attributable across the facets, with A3, A4, and A5, and C1, C3, C4, and C5 as primary contributors for the A and C facets, respectively. These results suggest that the original N, E, and O domains may have some facets that are problematic with regard to yielding similar results for college and adolescent samples. However, these group differences are more uniform (and thusly more problematic) across the A and C facets.

This conclusion was further evidenced by the DDA results for the aggregated domain-level scales. As expected based on the facet-level analyses, the A and C domains were clearly the primary contributors to the domain-level adolescent and college group differences, with structure coefficients of .721 and .816, respectively. The other structure coefficients were .186 or less. Taken together, these results suggest that the adolescent and college samples tended to score differently on the NEO PI-R facets and domains and that these differences may be most pronounced on the A and C facet and domains scores.
DISCUSSION

The present results appear to support the position that there are fundamental differences between the college and adolescent samples observed with regard to personality traits measured by the NEO PI-R. This study suggests that knowledge of group membership can predict scores at the facet and domain levels. Specifically, regarding each of the facet scores’ ability to account for these group differences, adolescents scored higher on the facets Angry Hostility (N2) and Vulnerability (N6) on the N domain; higher on Excitement Seeking (E5), but lower on Positive Emotions (E6) on the E domain; and higher on the Fantasy (O1), but lower on the Feelings (O3) facets of the O domain. The adolescents scored uniformly lower on all the A and C facets, but the observed effects were primarily due to lower scores on the Altruism (A3), Compliance (A4), and Modesty (A5) facets of the A domain and lower on the Competence (C1), Dutifulness (C3), Achievement Striving (C4), and Self-Discipline (C5) facets of the C domain. Overall, the A and C domains were the largest predictors in domain differences between the two groups.

Earlier research conducted by Costa and McCrae (1992) found that adults scored higher than college-age individuals on the Agreeableness and Conscientiousness domain scales. However, the authors have cautiously suggested the possible use of the NEO PI-R in high school and junior high samples (Costa & McCrae, 1992). Given the current findings, the test authors’ caution appears warranted, and use of college-age norms with adolescents is questionable as group differences existed on all domains. The development of normative data for adolescent populations appears necessary.

More important, all domains exhibited differences, but the facets on the A and C domains were uniformly lower. Several issues should be noted when trying to account for the group differences on the A and C domains. Within the debate of stability in personality, it is possible that characteristics measured by the A and C domains may be less stable in personality development than characteristics measured by the N, E, and O domains. Costa and McCrae (1992) noted that the college-age sample scored lower than adults on both the A and C domains in their normative data. In the present study, the adolescent sample scored even lower than college-age participants on all of the A and C facets. Taken collectively, these findings suggest that the A and C domains may be less stable than the N, E, and O domains across the life span. Furthermore, there is a consistent pattern of A and C domain growth, as opposed to unpredictable instability, from adolescence to adulthood. As an additional point concerning the E domain, Costa and McCrae (1992) also observed that their college sample scored higher than adults on the Excitement Seeking facet. The adolescents in the present study scored higher than the college sample, again suggesting a possible trend in personality change across time on this facet.

Furthermore, the A and C domains were added later in the construction of the NEO Personality Inventory. With the authors theorizing personality stability in the construction of their instrument (Costa & McCrae, 1988; McCrae, 1993), it is possible that these two additions simply are not as readily amenable to this theoretical expectation as the N, E, and O domains. Further operationalization and development of the A and C facet scales may be necessary to achieve the stability goal. Piedmont and Weinstein (1993) noted that the A and C domains had no substantial cross loading in their factor analysis of NEO PI-R scores, suggesting that the A and C domains may be independent of, and different from, the other domains. Although this finding only indicates factorial independence, it does leave open the possibility of a different theoretical conceptualization for the A and C domains.

TABLE 4

<table>
<thead>
<tr>
<th>Domain Coefficients</th>
<th>N</th>
<th>E</th>
<th>O</th>
<th>A</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.451</td>
<td>-.045</td>
<td>-.071</td>
<td>.186</td>
<td>.042</td>
</tr>
<tr>
<td>2</td>
<td>.517</td>
<td>.721</td>
<td>.815</td>
<td>.816</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness. Domain coefficients represent results from discriminant analysis using the total domain scores as the continuous variables.
It is interesting to look at the specific facet scales that contributed to the adolescent and college differences. Many of these differences, once noted, may appear to be logical. For example, adolescents scored higher on Vulnerability and Excitement Seeking and lower on Compliance and Competence. In general, one might expect adolescents to have lower scores on the A and C facets as many of these facets measure traits that may develop as persons move into adulthood. This interpretation is consistent with the noted findings in which adolescents score lower than college-age samples who, in turn, score lower than adults. Of course, longitudinal personality research is necessary to fully examine these issues more definitively.

If a downward extension of the NEO PI-R to adolescents is considered, a slight revision of the instrument is recommended. Many of the adolescents tested had problems defining words such as lackadaisical, prudence, panhandlers, and conscientiously. They also had difficulty with the idea of “new morality” and some of the phrasing of the statements. Overall, the researcher addressed an average of 10 questions per adolescent administration session (two sessions total) compared with only two questions in all of the college-age individuals with the same words generally being the target of their inquiries. None of the words with which individuals had difficulties were in the same facet.

An obvious limit to this study is that the NEO PI-R is a self-report instrument and as such is influenced by honesty, social desirability, and self-awareness of the respondent. Although the instrument has three validity scales separate from the test questions to address accuracy and honesty, these are again dependent on self-report. In addition, because of the difficulty in working and getting approval in the school system, 16- and 17-year-old students were not included in data collection. It is unclear how these ages may have performed on the scales.

One of the strengths of the NEO PI-R lies in its attempt to assess normal personality. Future research into normal adolescent personality seems essential to fully understand this population. The current study strongly suggests the need for separate normative data for adolescents administered the NEO PI-R as a starting point in future investigations. If research is conducted toward the development of these norms, the NEO PI-R could be a valuable tool in schools and in more extensive research of this population. Using larger or more diverse samples, the domain and facet scales may be more closely scrutinized and possibly gender differences may also be addressed. Furthermore, to separate cohort differences, age differences, and time of measurement confounds, a sequential research design is recommended.

REFERENCES


Alissa Sherry, Ph.D., is an assistant professor at the University of Texas at Austin. Her research interests include Substance abuse issues, adult attachment theory, personality development, ethics, constructivist applications.

Robin K. Henson, Ph.D., is an assistant professor at the University of North Texas. Her research focuses on applied statistics, measurement and assessment, and self-efficacy theory.

Jean G. Lewis, Ed.D., is a professor at Austin Peay State University. Her research interests are testing and assessment, and aging.