A First Large Cohort Study of Personality Trait Stability Over the 40 Years Between Elementary School and Midlife

Sarah E. Hampson
University of Surrey and Oregon Research Institute

Lewis R. Goldberg
Oregon Research Institute

This report provides some initial findings from an investigation of the relations between childhood Big Five personality traits assessed by elementary school teachers and similar traits assessed 40 years later by self-reports at midlife ($N = 799$). Short-term (1–3 years) test–retest reliabilities were lower (.22–.53) in childhood when personality was developing than they were in adulthood (.70–.79) when personality stability should be at its peak. Stability coefficients across the 40-year interval between the childhood assessment and the 2 measures of adulthood personality were higher for Extraversion (e.g., .29) and Conscientiousness (e.g., .25) than for Openness (e.g., .16), Agreeableness (e.g., .08), and Neuroticism (e.g., .00). Construct continuity between childhood and adulthood was evaluated by canonical analysis and by structural equation modeling and indicated continuity at both a broad, two-dimensional level and at the level of the Big Five. The findings are discussed in relation to A. Caspi, B. W. Roberts, and R. L. Shiner's (2005) principles of rank-order personality stability.

Keywords: personality stability, Big Five, construct continuity, longitudinal study

The extent to which child personality is predictive of adult personality has long been a fundamental scientific, philosophical, and even poetic question. Wordsworth's "The child is father of the man" and Milton's "The child shows the man, as morning shows the day" are not unambiguous. Neither poet was necessarily suggesting perfect correspondence and complete predictability from youth to maturity. Similarly, the stability of personality characteristics over the life course remains a controversial empirical issue. Proponents of stability debate the findings with advocates of transformation and change (Lewis, 2001; McCrae et al., 2000).

In this article, we provide some initial findings from an investigation of the relations between childhood personality traits assessed by elementary school teachers and similar traits assessed by self-reports at midlife. Our analyses are based on data from the Hawaii Personality and Health Cohort, a relatively large and culturally diverse sample that was first assessed between 1959 and 1967 when the participants were children in Hawaii. The childhood assessments were designed and supervised by the late John

M. Digman and formed the basis for some of the pioneering work on the five factor structure of personality (e.g., Digman, 1989). The adult assessments were undertaken over a 4-year period (1998–2002) after an interval of 35–43 years (Hampson et al., 2001). Consequently, these data offer an extraordinary opportunity to examine the question of personality stability from childhood to midlife with measures of the five factor model of personality at each time point. We have been unable to find a report of any other longitudinal personality data with a comparable convergence of the key features of our study: the first time of assessment during childhood and the second time in adulthood, spanning a follow-up period of around 40 years between childhood and adulthood, with the same personality factors assessed at both times of measurement.

Two goals motivated these analyses: The first is to contribute to the continuing scientific debate about the stability of personality traits over the lifetime. Our contribution provides a unique addition to this debate because of the portion of the life span bridged by the two assessments. A second reason for studying the stability of personality is in the service of another but related purpose. When a child personality predictor of an adult outcome is identified, the pathway from child trait to adult outcome demands investigation. For example, Friedman and his colleagues (e.g., Friedman et al., 1993, 1995) demonstrated that childhood Conscientiousness is predictive of longevity. This provocative finding stimulated us to embark on a longitudinal study of the Hawaii sample, ultimately to attempt replication and, in the shorter term, to test models about the mechanisms and processes involving personality traits that create the pathways from child predictors to adult outcomes (Hampson, Goldberg, Vogt, & Dubanoski, 2006, in press). An evaluation of the stability of personality traits in the Hawaii sample guides the development of these models.

There are several ways in which personality stability can be defined and studied (Caspi & Roberts, 1999, 2001). Interindividual

stability, called *differential continuity* by Caspi and Roberts (2001), refers to the rank-order stability of members of the same sample over time, and typically it has been evaluated by using correlation coefficients (e.g., Conley, 1984; Roberts & DelVecchio, 2000). Another approach to measuring personality stability has been called *absolute continuity* (Caspi & Roberts, 2001). The mean levels of traits can be compared at different ages, either with different cohorts in cross-sectional designs (e.g., Conley, 1985; Costa & McCrae, 1988) or with the same cohort at different ages in longitudinal studies (e.g., Roberts & Chapman, 2000). Such studies are used to infer the extent to which personality traits increase or decrease with age. The longitudinal data available for the Hawaii cohort lend themselves to the evaluation of rank-order consistency between childhood and adulthood.

An issue for all longitudinal studies that span development from childhood to adulthood is that of construct continuity from one time point to another (e.g., Caspi & Silva, 1995). Scientific advances in personality theory and measurement may result in a lack of comparability between the measures selected at the two times of assessment. Moreover, the very process of personality development from childhood to adulthood may result in construct discontinuity (e.g., Kagan, 1980); at the extreme, the evolution and transformation of personality constructs might be such that equivalent measures are not theoretically possible. Even if the behavioral indicators of the construct at each time point are internally consistent and have high short-term test–retest stability, they need not be identical constructs at the two age points (Asendorpf, 1992b). For example, with infants and toddlers the childhood measures have typically focused on temperament variables, for which there may be no directly comparable adult counterparts, and consequently rank-order consistency cannot be evaluated directly.

The potential discontinuity of any given construct has important implications for analyses of personality stability. For example, is it appropriate to attempt to relate measures of the five factor personality trait structure in childhood with measures of the five factor structure at midlife? Recently Measelle et al. (2005) identified Extraversion, Agreeableness, and Conscientiousness in children as young as 5 years of age by using a puppet interview. Other studies suggest that a five factor structure can be derived from teachers' ratings of children as young as 7 years of age but perhaps not from ratings of children aged 4 to 6 years, for whom only four broader factors may be more appropriate (Mervielde, Buyst, & De Fruyt, 1995). Indeed, Shiner and Caspi (2003) advocated a four factor structure for middle childhood. However, a five factor structure corresponding to the adult Big Five (i.e., the "Little Five") was identified by John, Caspi, Robins, Moffitt, and Stouthamer-Loeber (1994) in young adolescents. And Digman (1989; Digman & Inouye, 1986; Digman & Takemoto-Chock, 1981) found clear versions of the Big Five factors by using small portions of the very data here under study; later, Digman and Shmelyov (1996) found a clear five factor structure in teacher ratings of Russian school children in Grades 1, 2, and 3. Together, these studies suggest that the five factor model of personality can be used for the assessment of both children and adults to determine rank-order stability.

In the present study, the childhood assessments were based on variables that have been shown to conform to the five factor model of personality (Goldberg, 2001). The personality assessments in adulthood were also based on the five factor model. The comparability of constructs at the two times of assessment should min-imize construct discontinuity. However, there is still a great deal of difference between the behavioral indicators of the Big Five in children as used by teachers in their assessments versus the indicators used by adults at midlife to arrive at their self-descriptions.

That is, although the personality constructs assessed at each time point in our study might be viewed as broadly equivalent, the method of assessment cannot be viewed as such. The childhood assessments were based on teachers' perceptions of the children in their classrooms. We have argued elsewhere that teacher assessments are the ideal method for measuring the personality traits of children too young to provide self-reports (Goldberg, 2001). In brief, teachers are familiar with the child across a variety of classroom and other settings (e.g., recess), and they are in a perfect position to make normative judgments because of their extensive experience with children of the same age.

In contrast to these teacher assessments, all of the adult measures analyzed in this article were based on self-reports. Although we intend to expand the adult personality battery to include peer and observer descriptions, at this stage of the project we need not apologize for the use of self-report methodology, given the near ubiquity of such measures in the assessment of adult personality traits. The degree of convergence between the differing perspectives on personality provided by observer ratings versus self-report is itself as hotly debated as is the stability of personality over time (e.g., Funder, 1999; John & Robins, 1993; Kenny, 1994). Whatever the eventual outcome of this debate, the difference in perspective at the two times of our assessments may place a limit on the level of stability that can be obtained.

Given that this may be the first large-scale long-term study of rank-order personality stability, we must turn to studies of shorter age ranges to set our initial expectations (e.g., Roberts & DelVecchio, 2000). An examination of the relevant studies of rank-order stability reviewed by Roberts and DelVecchio (2000, Table 1) provides context for the present investigation. Those authors identified 34 studies in which the age at first assessment was between 6 and 12 years, which corresponds to the ages of our child participants at their first assessment. In 24 of these studies, the method of personality assessment was observer ratings; however, there were few studies (not including that of Digman, 1989, which is based on a portion of the data reported here) in which teacher assessments were used. Rubin, Hymel, and Mills (1989) used both parent and teacher ratings of sociability and social withdrawal in young children; social withdrawal observed at kindergarten and at Grade 2 ($N = 52$) were significantly correlated ($r = .37$, $p < .01$), but the correlation (not given) for sociability was not significant. Backteman and Magnusson (1981) used teachers' ratings of aggressiveness, motor disturbance, timidity, disharmony, distraction, and lack of school motivation in their study of 858 children. They examined stability from age 10 to 13 years and obtained a mean stability coefficient for boys of .52 and for girls of .48. In both of these studies, stability was examined over a brief interval within childhood, which gives an indication of the level of test–retest stability to be expected within childhood for our cohort. In this article, we provide another and somewhat richer examination of the longitudinal stability of teacher assessments.

Although there are no comparable longitudinal studies of rank-order stability that bridge the entire interval between middle childhood and middle age, there are studies that evaluate personality trait continuity over intervals within adulthood as long or longer

than the roughly 40-year time span of our present investigation. Two such studies (Haan, Millsap, & Hartka, 1986; Soldz & Vaillant, 1999) are remarkable for the time span that they encompassed. For 118 participants in the Oakland Growth Study, from adolescence to mid–late adulthood, a period of approximately 50 years, Haan et al. (1986) reported the following rank-order correlations for personality constructs assessed by observer Q-sorts: outgoing versus aloof = .37, cognitively committed = .34, self-confident vs. victimized = .26, dependable = .25, assertive versus submissive = .24, and warm versus hostile = .14. To the extent that the first two variables are associated with Extraversion and Intellect, respectively, this suggests that those factors may be more stable than the other three of the Big Five.

Soldz and Vaillant (1999) examined the stability of the dimensions of the five factor model of personality in their analysis of 163 participants in the Grant Study. These men were rated at the end of college on 25 personality traits by a psychiatrist, and they completed the NEO personality inventory at age 67–68 years. Scores on the five personality dimensions assessed by the NEO were estimated from the college trait ratings with the help of experts who assessed the relevance of each of the 25 traits to each of the five factors. The stability coefficients across the 30-year interval for the five domains were: Openness to Experience = .38, Neuroticism = .20, Extraversion = .19, Conscientiousness = .12, and Agreeableness = .07. The unusually high stability correlation for Openness may reflect its association with intellect/imagination and the possibility that cognitive variables may be more stable than other types of personality traits.

More generally, in using previous studies to guide our predictions regarding personality stability in the Hawaii cohort, there are several replicated findings across studies of rank-order personality trait stability that have now become enshrined as general principles (Asendorpf, 1992a; Fraley & Roberts, 2005; Roberts & DelVecchio, 2000). The following six such principles were identified recently by Caspi, Roberts, and Shiner (2005):

1. As a fundamental rule, the longer the interval between assessments, the lower the level of rank-order stability.

2. More specifically, rank-order stability increases with age.

3. Rank-order stability reaches a plateau between the ages of 50 to 70 years.

4. Rank-order stability does not seem to vary markedly by assessment method.

5. Rank-order stability does not seem to vary markedly by gender.

6. Rank-order stability does not seem to vary markedly among the Big Five personality factors.

On the basis of the first principle, higher short-term stability was expected when examining relatively brief test–retest intervals within childhood and within adulthood than when examining long-term stability from childhood to adulthood. Moreover, we predicted that the rank-order stability of traits from childhood to midlife would be lower than that found in previous studies such as Haan et al. (1986) and Soldz and Vaillant (1999). This is because

for the Hawaii cohort, personality trait assessments were available when the participants were in the midst of their childhood development, *middle childhood* (Shiner, 1998), and subsequently when they were in their 4th or 5th decade, when trait stability should be maximized (Principles 2 and 3). On the basis of Principles 5 and 6, we also predicted that rank-order stability would not differ substantially by gender or among the Big Five traits. Finally, given that previous studies had demonstrated that the childhood data conformed to the Big Five trait structure (Digman & Inouye, 1986; Goldberg, 2001), we predicted that we would demonstrate construct continuity between the child and adult versions of these latent traits.

Method

Child Participants and Procedures

The original child cohort included 2,404 elementary school children comprising six samples obtained by John M. Digman between 1959 and 1967. These samples, which are described in detail in Goldberg (2001), can be divided into two sets on the basis of their relative sizes. Each of the first three samples was reasonably large, and the personality characteristics assessed in these samples were all unipolar in their format (e.g., Gregarious). In contrast, each of the remaining three samples, all drawn from the Laboratory School of the University of Hawaii, was relatively small in size, and the variables were in a bipolar format (e.g., Gregarious vs. Solitary); some of the Lab School children were assessed more than once by different teachers in different elementary school years. The following descriptions of each of the six samples have been adapted from Goldberg (2001).

Oahu: Grades 1 and 2 (N = 885). In 1965, each of 885 children from eight schools on the island of Oahu were assessed by one of 29 teachers on 49 personality attributes. Each attribute consisted of a single word or short phrase, followed by a more extensive definition (e.g., *Energetic*: Active; full of pep; vigorous; movements are quick, darting).

Oahu: Grades 5 and 6 (N = 834). During the same year, each of 834 older children from the same eight schools were assessed by one of 28 teachers by using the identical set of 49 variables.

Kauai: Grade 6 (N = 502). In 1967, each of 502 sixth-grade children from eight schools on the island of Kauai were assessed by one of 17 teachers on 43 personality attributes. Again, each attribute consisted of a single word or short phrase, followed by a more extensive definition; 39 of these variables had been included in the two Oahu samples.

Lab '59: Grades 1 and 2 (N = 102). The remaining three partially overlapping samples were all drawn from the University of Hawaii Laboratory School. Four teachers assessed each of the children in their classrooms using the bipolar variables previously used by Cattell and Coan (1957).

Lab '60: Grades 1, 2, and 3 (N = 149). Of the 102 children assessed in 1959, 93 were reassessed a year later, along with 56 new students, mostly first graders. Two 1st-grade, two 2nd-grade, and two 3rd-grade teachers provided these assessments. The set of 50 bipolar variables used with this sample includes variants of most of those used in the 1959 sample plus 20 new ones devised by Digman.

Lab '63: Grades 5 and 6 (N = 100). This sample of fifth and sixth graders included 69 of those assessed in 1959 and 73 of those assessed in 1960; of the 100 children, 67 were assessed on all three occasions. Four teachers provided the 1963 child descriptions. Again, Digman modified many of the variables that he had used previously and added a number of new ones, resulting in an enlarged set of 63 bipolar variables.

For present purposes, the six child samples were aggregated into two sets: (a) the 2,221 children, originally assessed on a single occasion in Oahu or Kauai schools and (b) those Laboratory School children who were assessed on more than one occasion. The former sample is here used to

assess childhood-to-adulthood personality trait stability, and the latter is used solely to estimate the stability of teacher assessments during childhood.

General Procedures Common to All of the Child Samples

The teachers were provided with the names of the students in their classes and with sheets of paper or cardboard, on each of which a particular variable label was printed at the top; the teachers were instructed to rank their students from the highest to the lowest on that attribute. Rectangular boxes were included on each sheet to be used by the teacher for recording the students' names. All data were collected one attribute at a time, with a fixed nine-step quasi-normal distribution used for each classroom. Teachers were instructed to alternate between the two poles of the distribution in their rankings. The procedure was identical to that of a typical Q-sort, except that individuals, rather than attributes, were ranked (sorted). The resulting data distributions are therefore symmetric and quasi-normal, with very comparable means and standard deviations across both personality variables and classrooms. For the Kauai and the two Oahu samples, each personality attribute was accompanied by a definition of the characteristic—definitions that had been developed beforehand from focus groups of teachers who had been asked to provide typical examples of classroom behaviors relating to that concept.

Adult Participants and Procedures

Since July of 1998, our research team has been attempting to locate each of the roughly 2,000 now-adult living members of the childhood cohort and to recruit as many of them as possible into a study of adult health behaviors and outcomes. The details of our location and recruitment procedures are described in Hampson et al. (2001). As the adult participants are recruited for the project, they are periodically mailed a series of questionnaires, and they are invited to attend a half-day session at a medical setting where they complete an extensive battery of physical, medical, personality, and cognitive measures. The present analyses are based on the roughly 400 men and 400 women who completed the first two of our adult questionnaires before the end of 2004. For most of these participants, the interval of time between the teacher assessments during childhood and the completion of the adult questionnaires was approximately 40 years.

Measures of Adult Personality Traits

Our first questionnaire was a 16-page survey of demographic variables and health-related behaviors (e.g., smoking, drinking, diet, exercise); included in this questionnaire were the 44 items in the Big Five Inventory (BFI; John & Srivastava, 1999), which is perhaps the best of the various brief sets of five factor markers. John developed each of the five BFI scales to fall roughly halfway between the lexical Big Five factors (Goldberg, 1992) and the five domain scores from the NEO-Personality Inventory—Revised (NEO-PR-R; Costa & McCrae, 1992), and consequently he named his inventory with the lexical expression and used the NEO labels for the scales: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience. The BFI items were administered with a 5-point response scale, ranging from 1 (very inaccurate) to 5 (very accurate) as self descriptors; the BFI items can be found in Appendix A. For the sake of consistency, we use the BFI labels for the Big Five traits throughout the remainder of this article.

Our second questionnaire was administered between 2 and 4 years after the first; the mean length of time between completion of the surveys was 2.8 years. This two-sided single-page questionnaire focused entirely on personality characteristics. One side comprised 46 items: a readministration of the 44 BFI items plus two items in the same format that assessed self-perceived physical attractiveness. The other side comprised 84 items: the 40 personality trait adjectives from Saucier's (1994) Mini-Markers

(SMM), which are a subset of terms from Goldberg's (1992) 100 unipolar markers of the Big Five factor structure that have been found to be the most univocally associated with the same factor across diverse subject samples; two adjectives tapping self-perceived physical attractiveness; and adult versions of 42 of the 49 childhood variables (6 childhood variables overlapped with 6 SMM variables, and the childhood variable "self-minimizing" was omitted because of its unfamiliarity and replaced with "self-centered," reverse scored). The childhood variables included the 39 variables that were common to the teacher assessments in Oahu and Kauai. Omitted from the adult questionnaire were all of the defining child behavioral indicators that had been listed under each personality trait term in the teacher assessments. The combined set of 84 trait-descriptive adjectives, listed in alphabetical order, was administered with a 5-point response scale, ranging from 1 (very false) to 5 (very true) as self-descriptors.

Representativeness of the Adult Sample to the Child Cohort

To evaluate the extent to which the adult participants might constitute a biased or truncated sample of the original child cohort, we examined the means and standard deviations of their childhood Big Five factor scores, which are provided in Table 1. The values in this table are based on the child factor scores from the total Oahu and Kauai samples (i.e., the Laboratory School samples are excluded), computed separately within the subsamples of boys and girls; thus, if the adult men and women subsamples were completely unbiased, they would have means of zero; and if the adult subsamples were completely unrestricted in range, they would have standard deviations of 1.00. Clearly, there is virtually no range restriction for any of the adult samples on any of the child factors, with the standard deviations ranging from .98 to 1.03. Nor is bias likely to be much of a problem, with the possible exception of Conscientiousness, on which, predictably, the adult samples are about one-tenth of a standard deviation higher than the cohort mean. In evaluating this small effect, one must keep in mind that it is probably at least partially artifactual because we know from previous studies (e.g., Friedman et al., 1995) and from prior analyses of our own cohort (Hampson et al., 2001) that early mortality is associated with low childhood Conscientiousness; thus, at least some of the children who were seen by their teachers as low in Conscientiousness are unlikely to be available for adult recruitment.

Analyses

We begin our analyses with the three Laboratory School samples, where we have access to test–retest teacher assessments for 1-, 3-, and 4-year intervals. We then turn to the adult participants, for whom we also have short-term (2- to 4-year) test–retest administrations of the BFI. Next we present Big Five factor structures in childhood and adulthood, prior to examining the extent of association between the two sets of personality factors. In addition to rank-order correlations, we evaluate the long-term stability and construct validity of personality traits by using canonical analysis and structural equation modeling. In this article, we focus on findings for the Big Five factors. We have also conducted stability analyses with the middle-level trait clusters and individual items. These, and analyses conducted separately by gender, are available from Sarah E. Hampson and Lewis R. Goldberg.

Results

Short-Term Retest Stability of the Childhood Teacher Assessments

Table 2 provides the stability correlations from the three Laboratory School samples, for each of the Big Five factors. Stability correlations over 1 year ranged from .52 for Extraversion to .28 for

Table 1
Representativeness of the Original Child Cohort in the Adult Samples: Mean Childhood Factor Scores in Each of the Adult Samples

| | Childhood factor scores (*M*s) | | | | *SD*s | | | |
| | Male | | Female | | Male | | Female | |
Factor	*n* = 396	*n* = 375	*n* = 403	*n* = 387	*n* = 396	*n* = 375	*n* = 403	*n* = 387
Extraversion	.02	.03	−.04	−.04	1.03	1.03	1.03	1.03
Agreeableness	.05	.06	.06	.07	0.97	0.97	0.98	0.98
Conscientiousness	.09	.10	.13	.12	1.04	1.03	0.99	0.98
Neuroticism	.07	.06	.00	.00	0.98	0.97	1.02	1.02
Openness	.05	.05	.08	.09	0.98	0.98	1.00	1.01

Note. The mean values are deviations (in standard deviation units) for each of the adult samples from the childhood cohort means (0.00), which were computed separately within the total childhood male and female samples; the sample standard deviations can be compared with the total childhood male and female cohort standard deviation of 1.00.

Neuroticism; over 3 years they ranged from .43 for Agreeableness to .22 for Neuroticism; and over 4 years they ranged from .55 for Openness to .36 for Extraversion. One provocative finding from these analyses suggests that the five factors may have different patterns of short-term stability: Agreeableness and Extraversion conformed to the predicted pattern of a linear decrease in stability over the length of the time interval, whereas, paradoxically, for the other three factors stability was higher over 4 years than over 3. The short-term stability of Neuroticism was lower than that of the other four factors.

Short-Term Retest Stability of One of the Adult Self-Report Measures

Table 3 provides evidence concerning the stability of the BFI scales over an approximately 3-year period of middle adulthood. The stability correlations were somewhat higher for Extraversion and Openness than for the other three factors. The magnitude of these test–retest correlations was considerably higher than the short-term stability of the childhood assessments.

The Childhood Variables and Big Five Factors

Appendix A provides a listing of the 39 variables that were common to the teacher assessments in the Oahu and Kauai sam-

ples, ordered by their factor loadings on five varimax rotated factors, when these factors were derived in the sample (*N* = 799) for whom adult personality measures are available. Although the content of the Big Five factors is clearly apparent, the varimax solution is far less simply structured than one finds in analyses of adult personality variables. As evidence of the factorial complexity of these childhood variables, almost three-quarters (29 of the 39) of them have factor loadings of .30 or higher on two or more factors, and some of the secondary loadings are quite substantial in size. Indeed, a number of variables that help to define the Neuroticism factor (e.g., nervous, touchy, complains about others) have somewhat larger loadings on other factors.

Two Strategies for Assessing Big Five Personality Factors in Adulthood

Appendix B provides the varimax rotated six-factor solution from an analysis including all 130 variables in the second of the two adult questionnaires. The content of the Big Five factors plus a Physical Attractiveness factor is strikingly clear, with only about one-quarter (36 of 130) of the variables having secondary loadings of .30 or higher (as compared with three-quarters for the childhood

Table 2
Stability (r) Across Years of the Teacher Assessments for the Big Five Factor Scores in the University of Hawaii Laboratory School Samples

| | Laboratory School samples | | | |
Factor	1959 vs. 1960[a]	1960 vs. 1963[b]	1959 vs. 1963[c]	*M*
Extraversion	.52	.38	.36	.42
Agreeableness	.51	.43	.45	.46
Conscientiousness	.53	.31	.41	.42
Neuroticism	.28	.22	.38	.29
Openness	.49	.33	.55	.46
M	.47	.33	.43	

[a] *N* = 93.　[b] *N* = 73.　[c] *N* = 69.

Table 3
Internal Consistency (Coefficient Alpha) and Test–Retest Stability of Each of the Five BFI Scales

BFI scale	No. of items	α_1	α_2	Retest *r*	Corrected *r*
Extraversion	8	.84	.84	.79	.94
Agreeableness	9	.78	.80	.70	.87
Conscientiousness	9	.80	.81	.70	.87
Neuroticism	8	.82	.85	.71	.85
Openness	10	.79	.83	.79	.98
M		.81	.83	.74	.90

Note. *N* = 799. The two values of coefficient alpha are based on the first (α_1) and the second (α_2) administrations of the Big Five Inventory (BFI) items. The test–retest correlations are corrected for the unreliabilities at each of the two administrations. The average length of the test–retest interval was approximately 1,000 days (2.8 years).

variables). The Big Five factor scores from this six-factor structure were used as one of our two sets of adult personality criteria.

One advantage of the Big Five factor scores is that they were derived from a sizeable set of variables; a disadvantage is that all of these variables had been administered on the same occasion. To provide another set of criteria at the adult level, we used factor scores from an analysis of the $44 \times 2 = 88$ BFI items that had been administered on two occasions, roughly 3 years apart. These latter factors have the advantage of correcting for short-term fluctuations in adulthood for each of the Big Five traits. The correlations between the two sets of criterion factors (i.e., Big Five factor scores and BFI factor scores) are as follows: Extraversion = .91, Agreeableness = .86, Conscientiousness = .88, Neuroticism = .84, and Openness = .91.

Personality Stability Across the 40-Year Period

Table 4 provides the correlations between the corresponding childhood and adult Big Five personality factors for each of the two sets of adult criterion factors. In separate analyses for children assessed in Grades 1 and 2 compared with those assessed in Grades 5 and 6, we found no differences in the patterns of correlations. The 40-year stability correlations for each adult criterion were highest for Extraversion and Conscientiousness, were around zero for Neuroticism, and were intermediate for Openness. In this table, the Big Five domains are ordered by their relative stability from Extraversion (most stable) to Neuroticism (least stable).

For both procedures for calculating adult factors, we examined the statistical significance of the differences in stability correlations among all pairs of factors. When ordered by size of the stability correlation (as in Table 4) all differences in correlations between adjacent factors (E vs. C; C vs. O; O vs. A; and A vs. N) are not statistically significant, whereas all other pairs of such differences are significant. Indeed, all pairs of factors that are two steps removed from each other (E vs. O; C vs. A; and O vs. N) differ at least at the .01 level of significance, whereas all other pairs of factors (E vs. A; E vs. N; C vs. N) differ well beyond the

Table 4

Correlations Between the Corresponding Childhood and Adult Personality Factors (r) and Significant Multiple Correlations

Factor	r	K	R	F	df	p
Extraversion	.27**	3	.38	25.22	5, 756	.00
	.29**	3	.39	36.03	4, 794	.00
Conscientiousness	.25**					
	.23**					
Openness	.17**					
	.16**					
Agreeableness	.09*	2	.28	15.82	4, 757	.00
	.08*					
Neuroticism	.00					
	.00					

Note. The values appearing in the top of each cell are based on the adult factors derived from 130 variables (Big Five factor scores), and the values appearing in the bottom of each cell are based on the adult factors derived from the two administrations of the Big Five Inventory factor scores. K = Number of childhood factors included in the prediction.
* p < .05. ** p < .01.

.0001 level of significance. Clearly, then, these differences in factor stability correlations are likely to be highly reliable within this sample.

Table 4 also presents the findings from the stepwise multiple-regression analyses in which the five childhood factors were used to predict each of the adult criterion. In these analyses, the Gender $\times$ Childhood factor interactions were also tested for the significant predictors. The multiple correlations are reported only when more than one childhood personality factor predicted the adult criterion. Only two of the adult factors were predicted by a combination of childhood factors, and only two gender interactions were significant. The three-variable combination of childhood Extraversion (βs = .29 and .31), Disagreeableness (βs = .13 and .11), and lack of Conscientiousness (βs = .24 and .14) predicted adult Big Five Extraversion and BFI Extraversion, respectively. Lack of Conscientiousness was significant for men but not for women for the Big Five Extraversion criterion. The two-variable combination of childhood Agreeableness (β = .08) and, for men but not for women, Extraversion (β = .10) predicted the adult Big Five Agreeableness factor. Childhood Agreeableness predicted BFI Agreeableness only for men (β = .15).

Canonical Correlations Between the Two Sets of 39 Childhood and Adult Variables

What is the maximum possible correlation between the childhood personality variables and their adult counterparts? Perhaps the most straightforward procedure for answering this question is based on optimizing the correlation between sets of variables by using canonical correlation analysis. In our context, each of the 39 childhood variables and each of the 39 adult variables were separately weighted so as to optimize the correlation between those two sets. Each such canonical variate provides an optimal weighting of the variables within each set, after the variance in the preceding canonical variate has been residualized; thus, the intercorrelations among all canonical variates are zero. In our case, only the first two canonical correlations were statistically significant: The first canonical variate had a correlation of .47, which, when corrected for capitalization on chance, corrects to a value of .35; for the second variate, the correlation was .43, which corrects to a value of .30.

Those variables with the highest correlations with these two orthogonal factors are presented in Table 5. The largest of the two canonical variates combines variables from the Conscientiousness, Agreeableness, and Neuroticism domains, linking childhood classroom disruptiveness (e.g., restless, careless with others' property, fidgety, impulsive, and irresponsible vs. mannerly and careful of personal belongings) with adult undependability (e.g., irresponsible and impulsive versus conscientious and neat in appearance). The second canonical variate links childhood Extraversion traits (e.g., outspoken, socially confident, verbally fluent, and assertive versus submissive) with some adult Extravert counterparts (e.g., verbally fluent, gregarious, outspoken, assertive versus fearful).

Latent Modeling of Construct Continuity

Up to this point, all of our findings have been based on rather traditional analytic procedures, typically involving only correlations and multiple correlations. We now turn to some analyses

Table 5

The Variables With the Highest Correlations With Each of the First Two Canonical Variates From the Canonical Analyses (39 Variables)

Variable	Child	Adult
First canonical variate		
Restless	.63	.33
Careless of others' property	.57	.36
Fidgety	.56	.25
Impulsive	.48	.47
Irresponsible	.45	.54
Fickle	.46	.17
Nervous	.44	−.03
Spiteful	.34	.35
Rude	.31	.38
Mannerly	−.55	−.36
Careful of personal belongings	−.52	−.27
Persevering	−.47	−.26
Planful	−.43	−.36
Neat in appearance	−.41	−.41
Conscientious	−.39	−.41
Second canonical variate		
Outspoken	.49	.36
Socially confident	.46	.14
Verbally fluent	.43	.46
Assertive	.33	.28
Gregarious	.12	.37
Submissive	−.48	−.22
Fearful	−.27	−.36

Note. For the first canonical variate, the correlation between the childhood and adult variable sets is .47 (.35 when corrected for capitalization on chance). For the second variate, the correlation is .43 (.30 when corrected).

involving structural equation models to provide additional insight about the long-term stability and continuity of personality-trait constructs. For our analyses of latent construct continuity, we began with the 39 variables assessed by teachers in childhood and the equivalent set of variables assessed by self-reports at midlife (with "self-centered," reverse scored, substituted for "self-minimizing" in the adult assessment). In exploratory factor analyses of the adult self-ratings, these 39 variables were factored with and without the inclusion of the BFI items. Inspection of the orthogonal varimax five factor solutions in both analyses suggests that we should omit 6 of the 39 variables (mannerly, eccentric, impulsive, happy, seclusive, and lethargic), which were insufficiently strongly or univocally associated with any dimension. Thus, latent construct continuity was examined by using 33 variables as indicators of the adult constructs of Extraversion (assertive, energetic, gregarious, outspoken, socially confident, and verbally fluent vs. submissive), Agreeableness (considerate, adaptable, and self-minimizing/self-centered vs. rude, spiteful, and rigid), Conscientiousness (conscientious, careful, planful, neat in appearance, and persevering vs. careless and irresponsible), Neuroticism (nervous, fidgety, restless, jealous, fickle, fearful, touchy, complaining, and concerned about acceptance), and Openness (imaginative, original, curious, and aesthetically sensitive).

Construct continuity was examined separately for each of the Big Five dimensions by using structural equation modeling conducted with Mplus (Version 3), with standard maximum-

likelihood estimation (Muthén & Muthén, 2004). As indices of model fit, we examined the comparative fit index (CFI) and the root-mean-square error of approximation (RMSEA). The same trait indicators were used for both the child and adult latent constructs. To improve the fit of the models, correlated errors were allowed among the indicators within the child and within the adult constructs but not between them. Gender differences were evaluated by conducting multiple-group analyses. However, for none of the Big Five factors did the multiple-group analysis indicate that there were significant gender differences on any parameters, and therefore we report the findings based on the total sample.

For each of the Big Five factors, one model was tested in which the parameter estimates for the indicators of each construct were constrained to be equal in childhood and middle age, and another in which the parameters were unconstrained. If freeing the relative weights of the child indicators instead of forcing them to be identical to those for adults produced a better fit, this would suggest some differences in the way the Big Five dimensions were measured for children versus adults. Figures 1–5 show the unconstrained best-fitting structural equation models and provide the standardized path coefficients so that the relative strengths of the variables might be compared. Fit indices are provided in the figure captions. The standardized path coefficients between the child and the adult latent constructs, representing the correlations between them, demonstrated the same pattern of stability observed in our previous analyses, with Extraversion being the most stable of the Big Five factors and Neuroticism the least stable.

These analyses also evaluated construct continuity. The indicators (all $p < .05$) are ordered in terms of the absolute size of their path estimates for the adult latent construct. Variables with the largest discrepancy between these coefficients for the child versus the adult latent constructs appear to contribute differently to child and adult versions. The variables that showed the largest differences in parameter estimates for the child versus the adult constructs were identified from the best-fitting unconstrained models. The paths for these variables as indicators of the child latent construct were then sequentially freed, and the model fit was compared with the best-fitting model in which all the parameters for the child and adult constructs were constrained to be equal. These analyses suggested that the fit for Extraversion was significantly improved by freeing the variables socially confident and submissive, $\chi^2_{\text{Difference}}(2, N = 2221) = 19.27, p < .001$ (see Figure 1); for Conscientiousness, no significant improvement in fit was achieved (see Figure 2); for Openness the fit was significantly improved by freeing aesthetically sensitive, $\chi^2_{\text{Difference}}(1, N = 2221) = 4.42, p < .05$ (see Figure 3); for Agreeableness the fit was improved by freeing rigid and adaptable, $\chi^2_{\text{Difference}}(2, N = 2221) = 132.95, p < .001$ (see Figure 4); and for Neuroticism the fit was improved by freeing concerned about acceptance and fearful, $\chi^2_{\text{Difference}}(2, N = 2221) = 18.92 p < .02$ (see Figure 5).

The traits that significantly altered the fit of the models suggest some discontinuities between the child and adult versions of the constructs. For Extraversion, being socially confident was more important for adults (i.e., had a larger path estimate), and not being submissive was more important for children. For Agreeableness, being adaptable and not being rigid were more important for adults than for children. Being fearful was a stronger indicator of adult than of child Neuroticism, whereas being concerned about acceptance was a better indicator of childhood than was adult Neuroti-

Figure 1. The structural equation model for child and adult Extraversion; $\chi^2(68, N = 2221) = 126.86$, $p = .000$. Comparative fit index = .991, root-mean-square error of approximation = .020, 90% confidence interval = .014–.025.

cism. Aesthetic sensitivity was the weakest indicator for both childhood and adult Openness, but it was more strongly related to the childhood construct.

Discussion

The combination of childhood and midlife personality assessments in the Hawaii Personality and Health Cohort offered a unique opportunity to test the principles of rank-order personality stability across a hitherto unexamined span of 40 years from middle childhood to middle age. As discussed in the following paragraphs, our findings support the principles of personality sta-

bility identified by Caspi et al. (2005), with one exception. These findings contribute to the enduring issue of stability versus change in personality over the life course and have implications for life-course models of the effects of childhood personality on adult outcomes.

Trait Stability Within Childhood and Adulthood

Consistent with Principle 2, that rank-order stability increases with age, the 3-year stability correlations within childhood ranged from .36 to .55 (see Table 2). We can compare these correlations with those obtained within adulthood over the roughly comparable

Figure 2. The structural equation model for child and adult Conscientiousness; $\chi^2(68, N = 2221) = 97.79$, $p = .000$. Comparative fit index = .997, root-mean-square error of approximation = .014, 90% confidence interval = .007–.020.
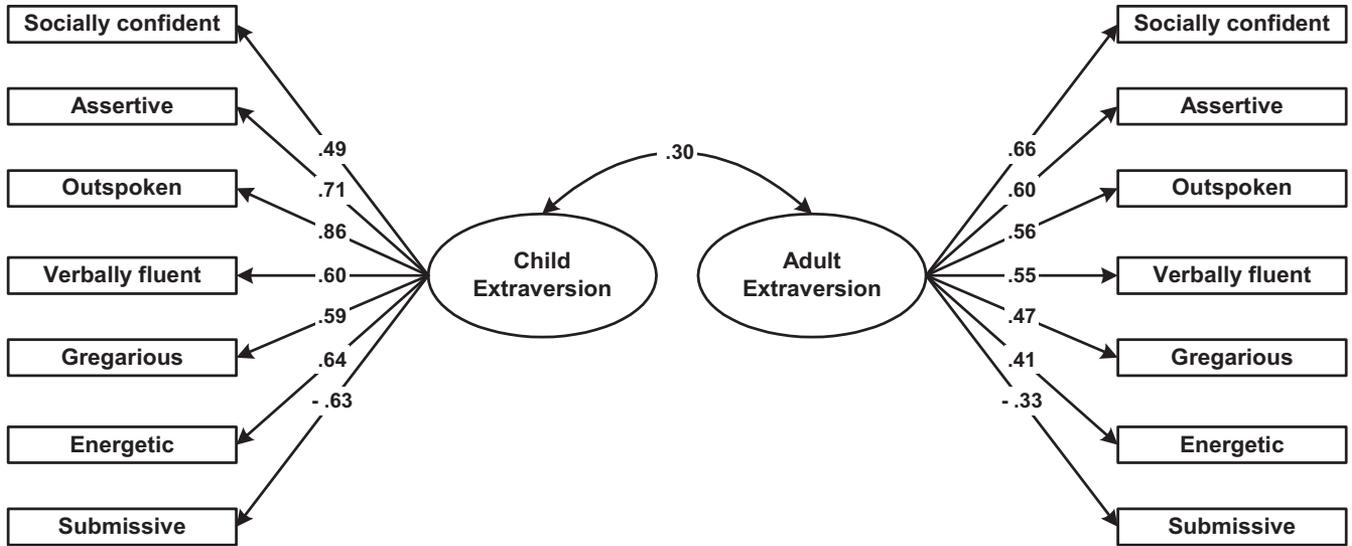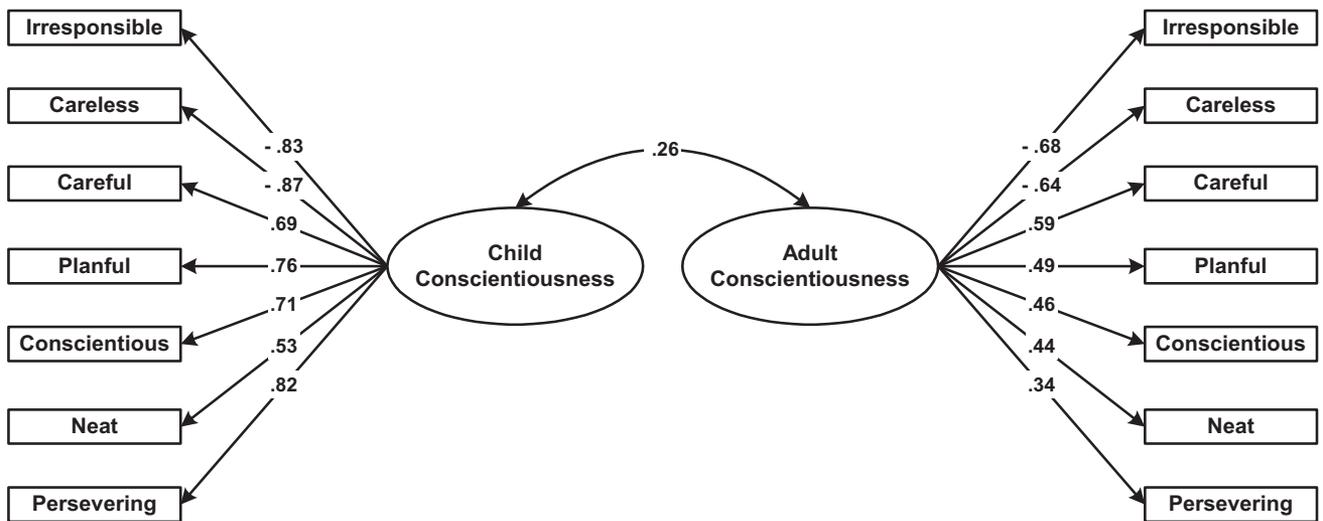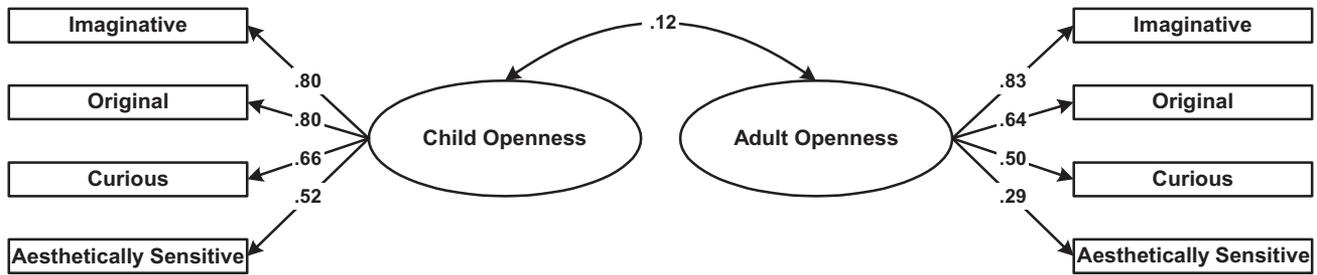
Figure 3. The structural equation model for child and adult Openness; $\chi^2(19, N = 2221) = 22.41$, $p = .264$. Comparative fit index = .991, root-mean-square error of approximation = .009, 90% confidence interval = .000–.021.

test–retest interval of 2.8 years, which ranged from .70 to .79 (see Table 3). Consistent with Principle 3, that personality stability peaks in the fourth and fifth decades (Roberts & DelVecchio, 2000), the test–retest correlations found here indicate considerably higher test–retest stability at midlife than in childhood. The test–retest correlations in childhood were relatively low for at least two reasons: Different teachers rated the children on each occasion, and this period of middle childhood is when personality is developing and hence is not expected to be highly stable. In contrast, the test–retest correlations in adulthood were based on self-reports during the period of maximum personality stability. However, direct comparisons of these short-term stability correlations between childhood and adulthood should be viewed cautiously, given that the childhood subsamples used for these analyses were from the University of Hawaii Laboratory School, which were smaller and less typical of the elementary school population at the time than our other childhood subsamples, and consequently were not included in our long-term analyses.

Trait Stability Between Childhood and Adulthood

According to Principles 1, 2, and 3 of rank-order stability offered by Caspi et al. (2005), stability coefficients over relatively short periods within childhood and adulthood should be consider-

ably higher than stability coefficients for the 40-year interval between childhood and adulthood. And, this conclusion was indeed confirmed. The stability correlations within both childhood and adulthood were higher than the stability correlations spanning the interval from childhood to adulthood: These long-term stability correlations ranged from a low of around zero for Neuroticism to a high of .29 for Extraversion. Given that the childhood assessments were conducted when the personality was developing, and the adult assessments occurred at a developmental plateau when adult personality had stabilized and prior to any changes in later life, to obtain any significant correlations between these two assessments could be seen as remarkable. We anticipated lower stability correlations than those reported in previous studies. For example, reports of correlations within adulthood across decades have ranged from .14 to .37 across 50 years (Haan et al., 1986), and from .07 to .38 across 30 years (Soldz & Vaillant, 1999). However, despite crossing the child–adult boundary in our 40 year follow-up, we found correlations displaying a similar range, from around .00 to .30.

Differences Among Traits

The meta-analysis by Roberts and DelVecchio (2000) showed the highest rank-order stability (population estimates controlling
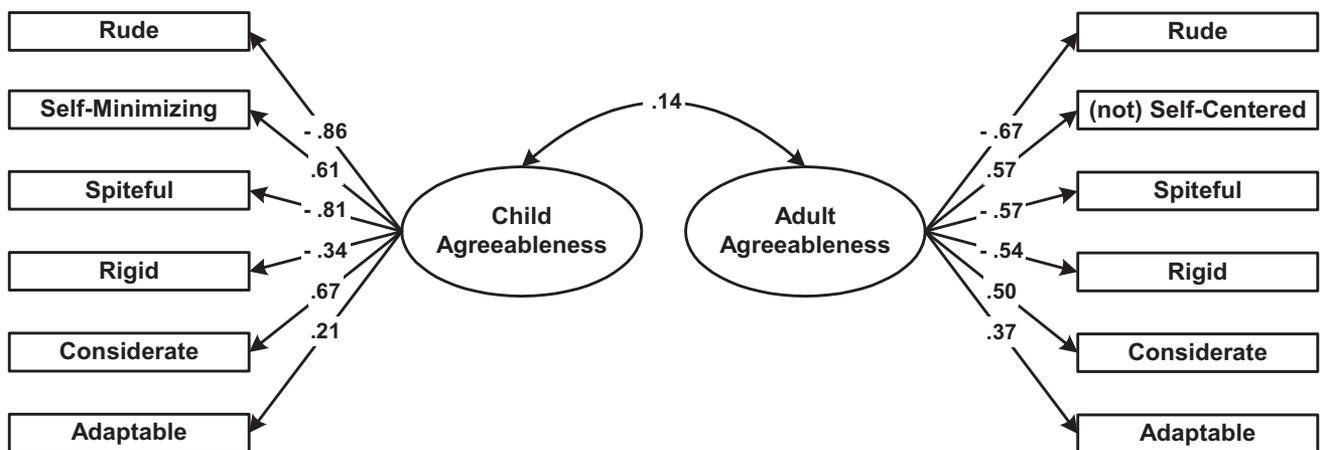

Figure 4. The structural equation model for child and adult Agreeableness; $\chi^2(45, N = 2221) = 67.78$, $p = .016$. Comparative fit index = .996, root-mean-square error of approximation = .015, 90% confidence interval = .007–.022.
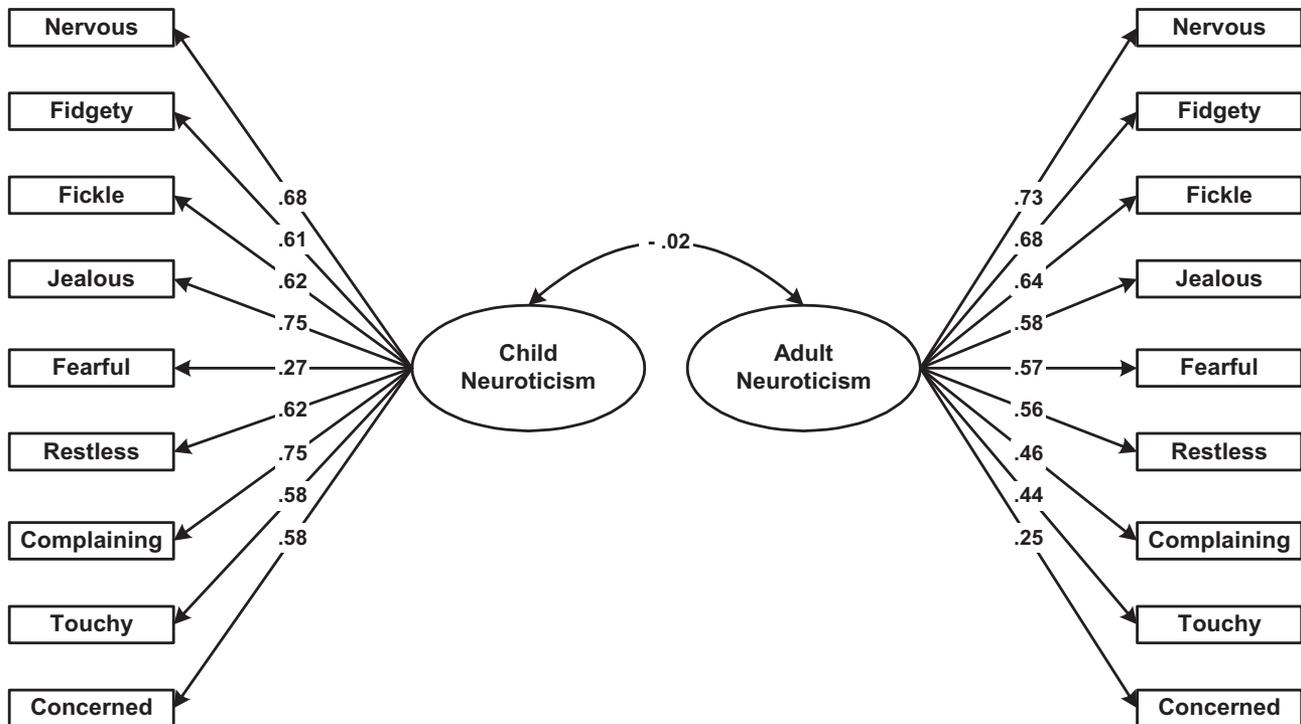
Figure 5. The structural equation model for child and adult Neuroticism; χ^2(112, N = 2221) = 183.43, p = .000. Comparative fit index = .992, root-mean-square error of approximation = .017, 90% confidence interval = .012–.021.

for time intervals and age of the samples) for Extraversion and Agreeableness (each .54) and the lowest for Conscientiousness (.51), Neuroticism (.50), and Openness (.51). These values formed the basis for Caspi et al.'s (2005) Principle 6 that stability does not vary by trait. Contrary to that conclusion, however, our study suggests that in the Hawaii cohort, all traits are not equally consistent from childhood to adulthood. Rather, we found several differences among the Big Five factors in their 40-year stabilities. At one extreme is Extraversion, followed closely by Conscientiousness, both of which demonstrate statistically significant links between childhood and adulthood, no matter the type of methodological procedure that is used to analyze those traits. At the other extreme, is Neuroticism, followed closely by Agreeableness, neither of which displays significant longitudinal stabilities over this long time period. In the middle, is the somewhat motley construct of Openness, which tends to provide statistically significant stability correlations, albeit at a much reduced level than that of Extraversion. Although this is but one study to compare with the many used in the meta-analysis, the Hawaii cohort is unique, bringing hitherto unavailable new data to this discussion. None of the studies included in the meta-analysis contributed rank-order stability coefficients spanning childhood and adulthood approaching anywhere near 40 years.

We have provided evidence of both the short-term and the long-term relative instability for the construct of Neuroticism. This finding makes good sense, given that the broad dimension of Neuroticism is the most "state-like" of the Big Five traits (Chaplin, John, & Goldberg, 1988) and the domain commonly targeted for

change by therapeutic endeavors. Neuroticism is also the factor that is most private and least visible to outside observers, perhaps even including classroom teachers. Indeed, it is the factor that has been found to be least concordant between descriptions provided by oneself and those of knowledgeable others (e.g., Funder, 1995, 1999). Therefore, the relative instability of Neuroticism found here may have been enhanced by the use of observer reports in childhood and self-reports in adulthood and may be less marked when the same assessment method is used at each time point.

At the other extreme is Extraversion versus Introversion, the least evaluatively charged of the Big Five, the most easily observable of them all, and the factor that tends to be most concordant between self-descriptions and those of others, even relative strangers (e.g., Norman & Goldberg, 1966). Both Extraversion and Neuroticism are normally considered the two dimensions of *temperament*, easily measurable in infancy and toddlerhood. Yet, our data suggest that they differ in both their short-term and their long-term stabilities, and this finding poses a challenge for those developmental theories linking infant temperament with adult personality traits.

In between the two extremes are three factors that appear to differ somewhat among themselves in their long-term (but not their short-term) stabilities. Conscientiousness appears relatively stable, whereas Agreeableness (often the largest of the Big Five factors in lexical studies, and a relatively observable trait) does not. To the extent to which Openness entails aspects of cognitive and other school-related abilities, one would expect significant childhood–adulthood stability. On the other hand, to the extent to

which this construct involves aspects of aesthetic appreciation and related artistic or creative experiences not readily available in childhood, one might expect some discontinuities between these two age periods. Perhaps, then, it is no wonder that the broad construct that combines these various personality characteristics displays some modest childhood-to-adulthood stability but not as strong as that of Extraversion and Conscientiousness.

Limitations to the Interpretation of the Stability Coefficients

Although the differences among stability correlations for the Big Five appear to be reliable for the Hawaii cohort, given the lack of comparable studies spanning such a long time interval and bridging the child–adult divide, the generalizability of these findings is yet to be determined. Therefore, it remains a possibility that the differences among the Big Five observed here are unique to this cohort and the relation between these two particular assessment points. Moreover, as noted at the outset of this report, a limitation of this study was that different assessment methods were used in childhood (teacher assessments) and adulthood (self-reports) on different instruments. This may have led to an underestimation of trait stability for some traits (e.g., Neuroticism) and certainly makes comparisons with findings from previous studies in which the same methods were used on both occasions more problematic. Whereas zero stability for Neuroticism is consistent with this methodological explanation, the close-to-zero stability for Agreeableness is less amenable to this account and remains a conundrum. A weakness of any study of trait stability with only two time points is that it does not address the development process, so we cannot make inferences from these single coefficients about differences between the trajectories of trait stability or change over time among the Big Five (Fraley & Roberts, 2005).

One striking difference between the childhood and adulthood assessments was the complexity of the Big Five factor structure in childhood compared with in adulthood. Teachers may not have known their students well enough to make fine-grained distinctions among 39 traits, but this seems unlikely given the precautions taken to ensure that teachers were familiar with their students and understood the meaning of the traits to be assessed. Because we know from previous research involving peer descriptions of adults that the structure of peer ratings and self-ratings is virtually identical (e.g., Goldberg, 1990), the complex structure of these teacher assessments could reflect the complexities in their students' personality traits at this early age, which has implications for construct continuity.

Construct Continuity

Whenever one finds instability in a trait between childhood and adulthood, it is worrisome unless the finding reflects nothing more than construct drift over time. It is plausible that teachers' concepts of children's personality traits could differ from trait concepts accessed 40 years later via self-reports. We examined construct continuity in the following three ways: (a) by multiple regression analyses, (b) by canonical correlations, and (c) by structural equation modeling. If the childhood version of a trait predicting the corresponding adult version contributed to the prediction of another (i.e., "off-diagonal") adult trait, this would imply construct discontinuity. In our analyses, most of the stepwise regression analyses that used the five childhood factors to predict each of the adult factors stopped at or before the first step. These analyses provide more support for continuity than for discontinuity at the level of the Big Five factors.

When we used the common set of 39 variables in a canonical correlation analysis, we found two substantial orthogonal canonical variates that optimally linked the childhood and adult variables. One of these combined many of the traits in the Conscientiousness, Agreeableness, and Neuroticism domains to form a broad prosocial dimension (reflecting a concern with getting along with others) and the other formed a clear Extraversion dimension (reflecting a concern with getting ahead of others). These canonical variates mirror virtually identically the two higher order dimensions of the Big Five factor structure (see Digman, 1997; Saucier & Goldberg, 2003) proposed by the same investigator, John M. Digman, who much earlier had collected this unique set of teacher assessments. This analytic approach, conducted at the level of the 39 common variables that were allowed to be differentially weighted, implies considerably less construct continuity than the regression approach. Construct continuity was optimized at the level of the two higher order factors rather than at the more specific level of the Big Five.

Finally, the structural equation modeling approach to the question of construct continuity explored differences in the measurement of each of the latent constructs for the Big Five childhood versus adult factors. These analyses were conducted separately for each of the Big Five factors because the complexity of the childhood data makes it difficult to get models to converge when cross-loadings between factors are permitted. Some measurement discontinuities were identified in these analyses, demonstrating that improved model fit could be achieved by freeing indicators of the child latent construct to differ from those of the adult latent construct. These differences reflect what may be thought of as variations in more versus less prototypical indicators for the child and the adult versions of the same Big Five trait.

Taken together, the various approaches illustrate that the degree of construct continuity observed depends on how it is studied. If indicators of the latent constructs are allowed to freely "cross-load," as in the canonical correlations, then the optimal prediction between child and adult is achieved through latent constructs at a broader level than the Big Five traits. If the Big Five factors are created at each of the two age points and then related through the use of regression analyses, they suggest moderate construct continuity. If the Big Five are assumed, yet their indicators are allowed to vary, as in the structural modeling approach, then subtle differences between the child and adult latent constructs can be observed.

Before concluding, we turn briefly to Caspi et al.'s (2005) Principle 5 that rank-order stability does not differ by gender. Our findings confirm this conclusion. In the regression analyses, gender only moderated the prediction of adult personality factors by childhood personality factors in two cases (Extraversion and Agreeableness), and no gender effects were obtained in the structural equation models.

Conclusions

This study has contributed to the continuing scientific debate about the lifetime stability of personality traits by presenting findings across a 40-year interval spanning childhood to middle age. Our findings suggest that for members of the Hawaii cohort there may be differences in stability among the Big Five traits. There is no inconsistency between this finding and the evidence that the Big Five are about equally heritable. Just as a trait could be extremely stable but not heritable (e.g., a trait completely determined by one strong kind of environmental influence), so a trait could be strongly heritable but not at all stable (e.g., the number of hair follicles per square centimeter of scalp in males). It is the genetic developmental program that gets inherited, not the phenotypical expression of a trait at any particular point in time. A further reason for studying trait stability is in service of the larger goal of identifying life span pathways from childhood predictors to adult morbidity and mortality. For example, childhood Conscientiousness has been shown to predict health behaviors that may mediate its effects on longevity (e.g., Friedman et al., 1995; Hampson, Goldberg, Vogt, & Dubanoski, in press). Presumably, if more conscientious children grow up into more conscientious adolescents and adults, then they are more likely to consistently benefit from engaging in health-enhancing behaviors and so increase their chances for a long and healthy life. The moderate stability of Conscientiousness observed here supports such a life course pathway involving this trait. Moreover, the even greater stability of Extraversion spotlights this trait for future studies of its influence on life course pathways to health.

References

Asendorpf, J. B. (1992a). Beyond stability: Predicting inter-individual differences in intra-individual change. *European Journal of Personality, 6,* 103–117.

Asendorpf, J. B. (1992b). A Brunswikean approach to trait continuity: Application to shyness. *Journal of Personality, 60,* 53–77.

Backteman, G., & Magnusson, D. (1981). Longitudinal stability of personality characteristics. *Journal of Personality, 49,* 148–160.

Caspi, A., & Roberts, B. W. (1999). Personality continuity and change across the life course. In L. Pervin & O. P. John (Eds.), *Handbook of personality psychology: Theory and research* (2nd ed., pp. 300–326). New York: Guilford Press.

Caspi, A., & Roberts, B. W. (2001). Personality development across the life course: The argument for change and continuity. *Psychological Inquiry, 12,* 49–66.

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology, 56,* 453–484.

Caspi, A., & Silva, P. A. (1995). Temperamental qualities at age three predict personality traits in young adulthood: Longitudinal evidence from a birth cohort. *Child Development, 66,* 486–498.

Cattell, R. B., & Coan, R. A. (1957). Child personality structure as revealed in teachers' ratings. *Journal of Clinical Psychology, 13,* 315–327.

Chaplin, W. F., John, O. P., & Goldberg, L. R. (1988). Conceptions of states and traits: Dimensional attributes with ideals as prototypes. *Journal of Personality and Social Psychology, 54,* 541–557.

Conley, J. J. (1984). Longitudinal consistency of adult personality: Self-reported characteristics across 45 years. *Journal of Personality and Social Psychology, 47,* 1325–1333.

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

Costa, P. T., Jr., & McCrae, R. R. (1988). Personality in adulthood: A 6-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. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Odessa, FL: Psychological Assessment Resources.

Digman, J. M. (1989). Five robust trait dimensions: Development, stability, and utility. *Journal of Personality, 57,* 195–214.

Digman, J. M. (1997). Higher order factors of the Big Five. *Journal of Personality and Social Psychology, 73,* 1246–1256.

Digman, J. M., & Inouye, J. (1986). Further specification of the five robust factors of personality. *Journal of Personality and Social Psychology, 50,* 116–123.

Digman, J. M., & Shmelyov, A. G. (1996). The structure of temperament and personality in Russian children. *Journal of Personality and Social Psychology, 71,* 341–351.

Digman, J. M., & Takemoto-Chock, N. K. (1981). Factors in the natural language of personality: Reanalysis, comparison, and interpretation of six major studies. *Multivariate Behavioral Research, 16,* 149–170.

Fraley, R. C., & Roberts, B. W. (2005). Patterns of continuity: A dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course. *Psychological Review, 112,* 60–74.

Friedman, H. S., Tucker, J. S., Schwartz, J. E., Tomlinson-Keasey, C., Martin, L. R., Wingard, D. L., & Criqui, M. H. (1995). Psychosocial and behavioral predictors of longevity: The aging and death of the "Termites." *American Psychologist, 50,* 69–78.

Friedman, H. S., Tucker, J. S., Tomlinson-Keasey, C., Schwartz, J. E., Wingard, D. L., & Criqui, M. H. (1993). Does childhood personality predict longevity? *Journal of Personality and Social Psychology, 65,* 176–185.

Funder, D. C. (1995). On the accuracy of personality judgment: A realist approach. *Psychological Review, 102,* 652–670.

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception.* San Diego, CA: Academic Press.

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

Goldberg, L. R. (1992). The development of markers for the Big Five factor structure. *Psychological Assessment, 4,* 26–42.

Goldberg, L. R. (2001). Analyses of Digman's child-personality traits: Derivation of Big Five factor scores from each of six samples. *Journal of Research in Personality, 69,* 709–743.

Haan, N., Millsap, R., & Hartka, E. (1986). As time goes by: Change and stability in personality over 50 years. *Psychology and Aging, 1,* 220–232.

Hampson, S. E., Dubanoski, J. P., Hamada, W., Marsella, A. J., Matsukawa, J., Suarez, E., & Goldberg, L. R. (2001). Where are they now? Locating former elementary school students after nearly 40 years for a longitudinal study of personality and health. *Journal of Research in Personality, 35,* 375–387.

Hampson, S. E., Goldberg, L. R., Vogt, T. M., & Dubanoski, J. P. (2006). Forty years on: Teachers' assessments of children's personality traits predict self-reported health behaviors and outcomes at midlife. *Health Psychology, 25,* 57–64.

Hampson, S. E., Goldberg, L. R., Vogt, T. M., & Dubanoski, J. P. (in press). Mechanisms by which childhood personality traits influence adult health status: Educational attainment and healthy behaviors. *Health Psychology.*

John, O. P., Caspi, A., Robins, R. W., Moffitt, T. E., & Stouthamer-Loeber, M. (1994). The "Little Five": Exploring the nomological network of the five-factor model of personality in adolescent boys. *Child Development, 65,* 160–178.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality, 61,* 521–552.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York: Guilford.

Kagan, J. (1980). Perspectives on continuity. In O. G. Brim, Jr., & J. Kagan (Eds.), *Constancy and change in human development* (pp. 26–74). Cambridge, MA: Harvard University Press.

Kenny, D. A. (1994). *Interpersonal relations: A social relations analysis.* New York: Guilford Press.

Lewis, M. (2001). Issues in the study of personality development. *Psychological Inquiry, 12,* 67–83.

McCrae, R. R., Costa, P. T., Jr., Ostendorf, F., Angleitner, A., Hrebickova, M., Avia, M. D., et al. (2000). Nature over nurture: Temperament, personality, and life span development. *Journal of Personality and Social Psychology, 78,* 173–186.

Measelle, J. R., John, O. P., Ablow, J. C., Cowan, P. A., & Cowan, C. P. (2005). Can children provide coherent, stable, and valid self-reports on the Big Five dimensions? A longitudinal study from ages 5 to 7. *Journal of Personality and Social Psychology, 89,* 90–106.

Mervielde, I., Buyst, V., & De Fruyt, F. (1995). The validity of the Big Five as a model for teachers' ratings of individual differences among children aged 4–12 years. *Personality and Individual Differences, 18,* 525–534.

Muthén, L. K., & Muthén, B. O. (2004). Mplus User's Guide (3rd ed.). Los Angeles, CA: Authors.

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

Roberts, B. W., & Chapman, C. (2000). Change in dispositional well-being and its relation to role quality: A 30-year longitudinal study. *Journal of Research in Personality, 34,* 26–41.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126,* 3–25.

Rubin, K. H., Hymel, S., & Mills, R. S. L. (1989). Sociability and social withdrawal in childhood: Stability and outcomes. *Journal of Personality, 57,* 237–255.

Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big Five markers. *Journal of Personality Assessment, 63,* 506–516.

Saucier, G., & Goldberg, L. R. (2003). The structure of personality attributes. In M. R. Barrick & A. M. Ryan (Eds.), *Personality and work: Reconsidering the role of personality in organizations* (pp. 1–29). San Francisco, CA: Jossey-Bass.

Shiner, R. L. (1998). How shall we speak of children's personalities in middle childhood? A preliminary taxonomy. *Psychological Bulletin, 124,* 308–332.

Shiner, R. L., & Caspi, A. (2003). Personality differences in childhood: Measurement, development, and consequences. *Journal of Child Psychology and Psychiatry, 44,* 2–32.

Soldz, S., & Vaillant, G. E. (1999). The Big Five personality traits and the life course: A 45-year longitudinal study. *Journal of Research in Personality, 33,* 208–232.

(*Appendixes follow*)

Appendix A

Five Factors Derived From 39 Childhood Variables From the Sample for Whom Adult
Personality Measures Were Available

Childhood variables	C	A	O	E	N
Conscientiousness (C)					
Careful of personal belongings	**−.78**[a]	−.11	.02	.02	−.07
Planful	**−.71**[a]	−.06	**.34**	.04	−.23
Persevering	**−.71**[a]	−.16	**.32**	.02	−.17
Neat in appearance	**−.67**[a]	−.02	−.03	.13	.07
Conscientious	**−.59**[a]	**−.40**	.23	−.04	−.19
Mannerly	**−.54**[a]	**−.38**	**.31**	.07	.00
Irresponsible	**.74**[a]	.18	−.22	−.03	.11
Careless of others' property	**.72**[a]	**.33**	−.07	.06	.08
Fidgets	**.65**[a]	**.35**	.00	**.30**	.20
Fickle	**.62**[a]	.28	−.13	.23	**.32**
Restless	**.60**[a]	**.40**	.06	**.39**	.18
Eccentric	**.51**[a]	.11	.15	**−.37**	.03
Nervous habits	**.48**[a]	.23	−.03	.09	**.46**
Agreeableness (A)					
Self-minimizing	−.24	**−.63**[a]	−.27	−.24	−.02
Considerate	**−.45**	**−.61**[a]	.19	.10	−.01
Submissive	−.03	**−.59**[a]	**−.40**	−.26	.19
Spiteful	**.36**	**.70**[a]	.02	.12	.16
Rude	**.43**	**.70**[a]	.02	.08	.15
Jealous	.22	**.70**[a]	−.01	.08	**.35**
Assertive	.10	**.68**[a]	**.30**	**.32**	.01
Touchy	.17	**.57**[a]	−.10	−.14	**.47**
Complains about others	.22	**.55**[a]	.09	.18	**.40**
Impulsive	**.50**	**.53**[a]	.12	**.33**	.18
Outspoken	.19	**.53**[a]	**.45**	**.38**	−.07
Openness (O)					
Imaginative	−.05	.02	**.80**[a]	.11	.05
Original	−.05	.06	**.78**[a]	.08	−.12
Curious	−.06	.18	**.70**[a]	.25	−.07
Socially Confident	−.20	.22	**.60**[a]	.28	−.23
Esthetically sensitive	**−.34**	−.18	**.57**[a]	.13	.18
Verbally fluent	−.02	**.32**	**.56**[a]	.27	−.07
Adaptable	**−.30**	−.03	**.55**[a]	.16	**−.44**
Rigid	.28	.25	**−.45**[a]	−.14	**.37**
Extraversion (E)					
Gregarious	.12	**.30**	.20	**.72**[a]	.01
Energetic	.12	.27	**.30**	**.64**[a]	−.05
Happy	−.18	−.24	**.38**	**.56**[a]	−.28
Seclusive	.01	−.14	−.28	**−.71**[a]	.06
Lethargic	.21	−.16	**−.40**	**−.49**[a]	.06
Neuroticism (N)					
Fearful	.09	−.05	−.24	−.16	**.72**[a]
Concerned about acceptance	.15	**.31**	.01	−.03	**.67**[a]

Note. $N = 799$. These 39 variables were common to the Kauai and the two Oahu samples. Only short-hand labels are used in this table; for the complete variable descriptions, see Goldberg (2001). Correlations of .30 or more are listed in boldface type.
[a] Highest loading for each variable.

Appendix B

Six Varimax Factors Derived From 130 Personality Items in the Adult Sample

Item	N	O	C	E	A	PA
Neuroticism (N)						
Tense [C]	**.76**[a]	.00	−.07	−.01	−.11	−.07
Nervous [C]	**.70**[a]	−.04	−.14	−.12	.08	−.05
Fretful [M]	**.68**[a]	−.08	−.15	−.04	.02	−.06
Moody [M]	**.66**[a]	.02	−.08	−.09	−.21	.09
Jealous [C, M]	**.65**[a]	−.05	−.16	.02	−.09	.02
Can be tense [B]	**.64**[a]	.03	−.03	−.06	−.12	−.08
Temperamental [M]	**.64**[a]	.09	−.02	.09	−.15	.06
Worries a lot [B]	**.63**[a]	−.08	−.14	−.10	.08	−.11
Gets nervous easily [B]	**.63**[a]	−.17	−.17	−.17	.07	−.12
Envious [M]	**.62**[a]	−.12	−.10	.02	−.03	.03
Can be moody [B]	**.62**[a]	.05	−.10	−.14	−.21	.03
Restless [C]	**.59**[a]	.12	−.13	.14	−.04	−.11
Fidgety [C]	**.58**[a]	.07	−.18	.03	−.02	−.01
Complaining [C]	**.58**[a]	−.11	−.03	.09	−.20	−.04
Spiteful [C]	**.58**[a]	−.05	−.02	−.04	−.29	.03
Is depressed, blue [B]	**.57**[a]	−.01	−.20	−.22	−.14	−.02
Suspicious [C]	**.57**[a]	.06	.04	−.06	−.09	−.10
Harsh [M]	**.56**[a]	.04	.02	.04	**−.38**	−.00
Fearful [C]	**.55**[a]	−.12	−.18	−.11	.04	−.07
Touchy [C, M]	**.53**[a]	.06	−.05	.04	−.04	.04
Tends to find fault with others [B]	**.52**[a]	.03	.03	.08	**−.31**	−.08
Rigid [C]	**.51**[a]	.03	.06	−.08	−.28	−.12
Fickle [C]	**.50**[a]	−.02	−.16	−.05	−.11	.08
Starts quarrels with others [B]	**.45**[a]	.02	−.15	.11	**−.35**	.03
Self-centered [C]	**.41**[a]	.16	−.14	.02	**−.36**	−.04
Lethargic [C]	**.39**[a]	−.12	−.29	−.22	−.10	.04
Concerned about acceptance [C]	**.37**[a]	−.06	−.03	−.03	.18	−.10
Submissive [C]	**.28**[a]	−.12	−.11	−.24	.23	−.09
Impulsive [C]	**.28**[a]	.27	−.24	.15	.03	.09
Is relaxed, handles stress well [B]	**−.59**[a]	**.30**	.16	.04	.10	.03
Is emotionally stable, not easily upset [B]	**−.57**[a]	.23	.17	.02	.25	.00
Relaxed [M]	**−.47**[a]	.15	.09	−.10	.26	.07
Emotionally stable [C]	**−.45**[a]	.11	**.31**	.08	.21	−.01
Remains calm in tense situations [B]	**−.45**[a]	**.33**	.27	−.06	.18	.05
Unenvious [M]	**−.33**[a]	.06	.02	−.10	.01	−.16
Adaptable [C]	**−.30**[a]	**.30**	.16	.07	.29	−.06
Openness (O)						
Imaginative [C, M]	−.07	**.71**[a]	.05	.09	.15	.15
Is original, comes up with new ideas [B]	−.05	**.70**[a]	.05	.17	.06	.17
Is inventive [B]	−.10	**.68**[a]	.05	.08	−.02	.16
Has an active imagination [B]	.07	**.68**[a]	−.01	.12	.03	.08
Creative [M]	.00	**.67**[a]	.04	.08	.11	.13
Likes to reflect, play with ideas [B]	−.11	**.65**[a]	.16	.07	.14	.00
Is curious about many different things [B]	.01	**.64**[a]	.09	.12	.17	−.04
Original [C]	−.03	**.62**[a]	.12	.10	.11	.23
Philosophical [M]	.03	**.62**[a]	.11	.03	.06	−.05
Is ingenious, a deep thinker [B]	.12	**.61**[a]	.22	−.03	−.04	−.01
Values artistic, aesthetic experiences [B]	−.06	**.59**[a]	.06	.08	.20	−.02
Curious [C]	.07	**.58**[a]	.05	.13	.13	−.11
Intellectual [M]	−.10	**.56**[a]	.20	.11	.04	.07
Is sophisticated in art, music, or lit. [B]	−.06	**.55**[a]	.01	.08	.02	.07
Deep [M]	.14	**.54**[a]	.16	.02	.14	.00
Knowledgeable [C]	−.12	**.48**[a]	.23	.17	.14	.14
Perceptive [C]	−.06	**.48**[a]	.29	.13	.13	−.01
Persevering [C]	−.07	**.46**[a]	**.41**	.10	.15	−.15
Complex [M]	.27	**.45**[a]	.09	−.05	−.10	−.12
Verbally fluent [C]	−.03	**.38**[a]	.19	.34	.01	−.01
Is full of energy [B]	−.17	**.34**[a]	.28	.33	.12	.17

(Appendixes continue)

Appendix B (*continued*)

Item	N	O	C	E	A	PA
Self-reliant [C]	−.11	.34[a]	**.31**	.02	.06	−.11
Eccentric [C]	**.32**	.32[a]	−.14	−.07	−.16	.06
Aesthetically sensitive [C]	.20	.32[a]	.05	.02	.21	−.02
Uncreative [M]	.16	−.55[a]	−.08	−.10	−.14	**−.32**
Prefers work that is routine [B]	.22	−.43[a]	.09	−.15	.16	.11
Unintellectual [M]	.17	**−.43[a]**	−.20	−.12	−.12	−.18
Has few artistic interests [B]	.15	**−.42[a]**	.02	.00	−.05	−.03
Conscientiousness (C)						
Organized [M]	−.07	.12	**.75[a]**	.06	.05	.14
Efficient [M]	−.08	.10	**.71[a]**	.09	.12	.03
Does things efficiently [B]	−.08	.18	**.67[a]**	.01	.19	.11
Makes plans and follows through with them [B]	−.16	.19	**.59[a]**	.13	.18	−.07
Planful [C]	.00	.24	**.58[a]**	.03	.10	−.07
Does a thorough job [B]	−.04	.17	**.56[a]**	.09	.05	.05
Perseveres until the task is finished [B]	−.10	.21	**.55[a]**	.01	.11	−.09
Careful of my belongings [C]	.03	.01	**.55[a]**	.05	.17	.08
Systematic [M]	−.02	**.32**	**.52[a]**	−.01	−.03	−.09
Practical [M]	−.10	.12	**.51[a]**	−.09	.22	−.16
Conscientious [C]	.01	.16	**.48[a]**	.05	.18	−.22
Neat in appearance [C]	−.01	.07	**.42[a]**	.21	**.33**	.27
Is a reliable worker [B]	−.06	.17	**.39[a]**	.09	.16	.03
Sensible [C]	−.21	.23	**.38[a]**	.02	**.33**	−.03
Energetic [C, M]	−.14	**.30**	**.30[a]**	.30	.15	.13
Tends to be disorganized [B]	.20	.00	**−.69[a]**	−.07	.00	−.21
Disorganized [M]	.21	.01	**−.68[a]**	−.02	.04	−.14
Inefficient [M]	.24	−.04	**−.63[a]**	−.10	−.11	−.07
Sloppy [M]	.22	.01	**−.58[a]**	−.12	−.07	−.27
Irresponsible [C]	**.31**	−.02	**−.56[a]**	.03	−.21	−.02
Careless [M]	**.30**	.00	**−.56[a]**	.01	−.08	−.11
Tends to be lazy [B]	.26	−.11	**−.49[a]**	−.17	−.06	−.14
Can be somewhat careless [B]	**.30**	.05	**−.49[a]**	.06	−.07	−.19
Is easily distracted [B]	**.35**	.04	**−.40[a]**	−.02	.05	−.10
Extraversion (E)						
Talkative [C, M]	.15	.20	.01	**.74[a]**	.13	−.02
Is talkative [B]	.15	.20	.01	**.72[a]**	.09	−.02
Is outgoing, sociable [B]	−.11	.22	.09	**.69[a]**	**.30**	.08
Outspoken [C]	.18	**.32**	.10	**.61[a]**	−.17	−.05
Extraverted [M]	.07	.23	.07	**.58[a]**	.11	−.00
Has an assertive personality [B]	.12	**.36**	.13	**.57[a]**	−.05	.09
Socially confident [C]	−.23	.29	.22	**.53[a]**	.15	.17
Assertive [C]	.06	**.35**	.17	**.52[a]**	−.09	.03
Generates a lot of enthusiasm [B]	−.08	**.44**	.09	**.49[a]**	.26	.09
Bold [M]	.02	**.37**	.14	**.48[a]**	−.13	.08
Gregarious [C]	.03	.22	−.01	**.44[a]**	.19	.20
Tends to be quiet [B]	.00	.01	.03	**−.81[a]**	−.03	.00
Quiet [M]	−.05	.02	.08	**−.78[a]**	.04	.04
Shy [M]	.13	.03	−.10	**−.72[a]**	.10	−.08
Is sometimes shy, inhibited [B]	.23	.01	−.07	**−.70[a]**	.07	−.09
Is reserved [B]	.10	.05	.08	**−.65[a]**	−.07	−.09
Bashful [M]	.18	−.02	−.06	**−.64[a]**	.08	−.09
Withdrawn [M]	**.43**	−.01	−.17	**−.47[a]**	−.22	−.15
Seclusive [C]	**.42**	.10	−.09	**−.46[a]**	−.23	−.05
Agreeableness (A)						
Is considerate and kind to almost everyone [B]	−.18	.09	.15	−.04	**.72[a]**	.06
Warm [M]	−.11	.18	.11	.18	**.69[a]**	.10
Is helpful and unselfish with others [B]	−.13	.13	.17	.01	**.68[a]**	.06
Kind [M]	−.10	.16	.16	.04	**.67[a]**	.17
Sympathetic [M]	.05	.12	.11	.09	**.62[a]**	−.08
Considerate [C]	−.11	.18	.26	.04	**.61[a]**	.04
Cooperative [M]	−.16	.16	.21	.00	**.60[a]**	.01
Thoughtful [C]	−.03	.25	.22	.08	**.59[a]**	.17
Likes to cooperate with others [B]	−.19	.13	.18	.02	**.58[a]**	−.08
Is generally trusting [B]	−.18	.13	.03	−.01	**.50[a]**	.05
Has a forgiving nature [B]	−.28	.14	−.06	.04	**.47[a]**	.06
Obedient [C]	−.02	−.02	.27	−.18	**.40[a]**	.03
Mannerly [C]	.03	.21	.33	.08	**.39[a]**	.11
Happy [C]	**−.34**	.17	.18	.21	**.35[a]**	.09

Appendix B (*continued*)

Item	N	O	C	E	A	PA
Rude [C, M]	**.44**	−.02	−.09	.06	−.51[a]	−.05
Is sometimes rude to others [B]	**.38**	.05	−.05	.04	−.50[a]	−.06
Unsympathetic [M]	.21	−.17	−.15	−.11	**−.50[a]**	−.05
Can be cold and aloof [B]	.39	.15	.04	−.25	**−.46[a]**	−.06
Cold [M]	**.41**	.08	−.04	−.14	**−.46[a]**	−.07
Physical Attractiveness (PA)						
Is physically attractive	−.04	.24	.17	.17	.13	**.71[a]**
Good-looking	−.02	.28	.15	.16	.17	**.71[a]**
Is not good looking	.14	−.18	−.17	−.19	−.15	**−.72[a]**
Unattractive	.15	−.17	−.15	−.15	−.16	**−.71[a]**

Note. *N* = 762. Loadings of .30 or more are listed in boldface type. [B] = Big Five Inventory; [M] = Mini-Markers; [C] = childhood variables.
[a] Highest factor loading for each item.

<div style="border:1px solid black; padding:1em;">

Instructions to Authors

For Instructions to Authors, please visit www.apa.org/journals/psp and click on the "Instructions to Authors" link in the Journal Info box on the right.

</div>