Published in final edited form as:
J Res Pers. 2013 October 1; 47(5): 505–513. doi:10.1016/j.jrp.2013.05.003.

Personality Stability from Childhood to Midlife: Relating Teachers' Assessments in Elementary School to Observer- and Self-Ratings 40 Years Later

Grant W. Edmonds, **Lewis R. Goldberg**, **Sarah E. Hampson**, and **Maureen Barckley**
Oregon Research Institute, Eugene, Oregon

Abstract

We report on the longitudinal stability of personality traits across an average 40 years in the Hawaii Personality and Health Cohort relating childhood teacher assessments of personality to adult self- and observer- reports. Stabilities based on self-ratings in adulthood were compared to those measured by the Structured Interview for the Five-Factor Model (SIFFM; Trull & Widiger, 1997), and trait ratings completed by interviewers. Although convergence between self-reports and observer-ratings was modest, childhood traits demonstrated similar levels of stability across methods in adulthood. Extraversion and Conscientiousness generally showed higher stabilities, whereas Neuroticism showed none. For Agreeableness and Intellect/Openness, stability was highest when assessed with observer-ratings. These findings are discussed in terms of differences in trait evaluativeness and observability across measurement methods.

Keywords

personality stability; Big Five; person perception; judgment accuracy; Self-Other

> No man ever steps in the same river twice, for it is not the same river and he is not the same man.
>
> -Heraclitus of Ephesus

The famous quotation from Heraclitus draws attention to the complexities of studying personality stability from childhood to adulthood. Just as Heraclitus questioned the possibility of traversing the same river twice, it is impossible to repeat teacher assessments of children's personality traits in adulthood. However it is possible to assess adult personality using a variety of other informant-based methods in addition to self-reports, and to evaluate the degree to which such methodological differences affect long-term stability coefficients. In the current study, we evaluated the degree of convergence between teacher assessments in childhood and observer-ratings in adulthood for participants in the Hawaii Longitudinal Study of Personality and Health. We compared these findings to stability coefficients based on self-reports in adulthood, and we related self- and observer-reports in adulthood. Observer-ratings were derived from a clinical interview that was conducted an

Correspondence concerning this article should be addressed to Grant W. Edmonds, Oregon Research Institute, 1776 Millrace Drive, Eugene, OR 97403-2536. gedmonds@ori.org.

average of 40 years after the childhood teacher-assessments. Recent work on rank-order trait stability and accuracy of observer-ratings of personality informed our theoretical perspective on how these different sources of personality information may affect long-term stability coefficients.

Rank-Order Stability

Rank-order stability is the test-retest correlation of a trait within the same sample over time, and represents the degree to which individuals retain their relative position with respect to each other. Rank-order stability operates independently from mean-level changes. Previously, Hampson and Goldberg (2006) obtained rank-order stability coefficients for the Big Five traits for participants in the Hawaii study based on teachers' assessments in childhood and self-reports in adulthood. Stabilities were highest for Extraversion (.30), followed by Conscientiousness (.25), Intellect/Openness (.17), and Agreeableness (.09), and were zero for Neuroticism. In this article, we evaluate the degree to which differences in measurement method in adulthood influence estimates of personality stability from childhood. To the degree that each trait may have been affected equally by a difference in measurement method across time, all of our stability estimates may have been underestimated. Alternatively, we may find that a change in method has little effect on some traits, while attenuating others more. Correcting a measurement confound could result in a reduction in the degree to which stabilities vary across traits.

How much improvement in stability estimates can be expected? Meta-analytic results suggest that the average 40-year rank-order stability correlation for personality traits should be approximately .25 (Roberts & DelVecchio, 2000). Roberts and DelVecchio arrived at this estimate by regressing stability coefficients taken from 152 longitudinal studies onto the time interval for each study while controlling for age. The results of this regression were then used to estimate the predicted stability over a variety of time intervals starting at age 20. While their 40 estimate does not correspond to the same age range in our data, it represents the best empirically derived estimate of the average stability resulting from accumulated trait change over 40 years. The level of stability we observed previously for Extraversion and Conscientiousness already exceeds this figure, suggesting it is unlikely that a methodological change will produce an increase in stability for these domains. However, there is considerable room for improvement for the other Big Five domains, particularly for Neuroticism. According to Fraley and Roberts' (2005) model of asymptotic decline, trait stability declines in a non-linear fashion approaching a non-zero asymptote, which results from a persistent constant that maintains stability in the face of forces that would otherwise lead to the accumulation of changes over time. Asymptotic decline of this nature is consistent with a model where trait stability is intermediate between a model based on a fixed set-point (i.e., high stability), and one where trait levels are entirely experience dependent (i.e., zero stability), (Ormel, Riese, & Rosmalen, 2012). By midlife, trait stability will have reached this asymptote suggesting that the zero-level of stability observed previously for Neuroticism was an underestimate and indicating that greater stability estimates could be achieved for this domain with methodological improvements.

Accuracy: The Self versus Knowledgeable Informants

We take as a premise that anything that is likely to affect the accuracy of personality assessments cross-sectionally must also be considered in longitudinal research. In the Hawaii cohort, where the child assessments are based on a method that cannot be perfectly reproduced in adulthood, these considerations are especially important and are likely to affect any study that spans child and adult phases of life.

The Brunswik lens model represents one of the first formal models applied to understanding accuracy in personality judgments (Brunswik, 1956). In its most basic formulation as a model for accurate assessments of personality, the lens model specifies that judgment accuracy depends on the presence of valid cues (availability) and also on judges' use of these cues (cue utilization). The model allows for researchers to characterize conditions where accuracy is less than optimal. This may result from a lack of available cues, from raters failing to use valid cues, or from the presence and utilization of non-valid cues. Building on the lens model, Funder's (1995, 1999) Realistic Accuracy Model (RAM) gives a more detailed account of the conditions that must be met for accurate personality judgments to occur: (1) relevant trait information must exist; (2) it must be available to raters; (3) raters must notice available cues; and (4) cues must be correctly interpreted and used to form judgments.

The Brunswik lens and Funder's RAM models lay out conditions for accurate assessments of personality based on observer-reports. Evaluating accuracy formally requires the specification of an accuracy criterion and, in the case of personality judgments, this is often based on self-reports. Vazire's (2010) Self-Other Knowledge Asymmetry (SOKA) model specifically addresses sources of asymmetry in self- and observer-reports. This model posits two premises based on trait observability and evaluativeness. Principle 1 states that others will know more about highly observable traits while the self has more knowledge of less visible traits. Thus, trait observability leads to asymmetries across self- and observer-ratings. Principle 2 states that because of self-serving perceptual biases (i.e., self-enhancement) others will know more than the self when viewing highly evaluative traits, which also leads to asymmetries in self- and observer-ratings. These principles of self-other asymmetry follow the general consensus that observability and evaluativeness are essential elements that affect accurate personality judgments (John & Robins, 1993; Tetlock, 1984).

Much of the work on judgeable traits has focused on observability in the context of social interactions, where Extraversion is highly judgeable in comparison to other Big Five factors. This remains true even in studies where judges have very limited exposure to the target (e.g., Borkenau, Brecke, Möttig, & Paelecke, 2009; Norman & Goldberg, 1966). Because the Big Five factors vary in the degree to which they are observable and subject to biases resulting from evaluativeness, these properties may explain the degree to which traits vary in rater agreement. A typical pattern of the rank-order of inter-rater agreement has emerged across the literature. Extraversion tends to have the highest agreement, while Agreeableness and Neuroticism tend to show the lowest levels, with Conscientiousness and Intellect/Openness falling in the middle (Norman & Goldberg, 1966). When comparing self-other agreement to observer-observer agreement, the same general pattern emerges (Albright, Kenny, & Malloy, 1988; Funder & Colvin, 1988; Watson, 1989). However, while the rank-order of convergence across traits remains the same, observer-observer agreement tends to be monotonically higher than self-observer agreement (John & Robins, 1993). While differences in observability and evaluativeness are useful for characterizing some of these differences, chiefly high agreement for Extraversion and low agreement for Neuroticism, Agreeableness stands out as a trait that shows low levels of agreement across a variety of methods. This is puzzling given that Agreeableness displays similar levels of observability and evaluativeness when compared to Conscientiousness, a trait that tends to consistently show moderate levels of convergence across many different measurement scenarios (Vazire & Gosling, 2004).

The Current Study

Testing the degree to which different methods of assessment in adulthood affect levels of trait stability in the Hawaii cohort should provide valuable information about patterns of

personality stability and change across the life course. The Hawaii cohort data are unique in that they include a remarkably well-collected set of elementary-school teacher assessments in childhood, measures of the same traits in adulthood by the self and observers, a long interval between child and adult measures, and a large ethnically diverse sample. In this report, we estimated child-adult stability coefficients using interviewer/observer-ratings based on the Structured Interview for the Five Factor Model (SIFFM; Trull & Widiger, 1997), plus interviewer/observer-ratings following the interview using three different personality measures. By comparing stability obtained from self- versus observer-reports at midlife, we can examine the influence of trait observability and evaluativeness on these estimates. To the extent that differences in rater perspective and measurement instruments result in a methodological confound, we expect to find the highest stability estimates using measures and measurement methods that vary the least across measurement occasions. Following this reasoning, we treat the childhood assessments as a kind of accuracy criteria, and assume that higher stability coefficients result from greater accuracy in measurement, while recognizing that stability estimates are also limited by the amount of true change occurring over time. Extraversion and Conscientiousness already displayed stabilities that exceeded the best meta-analytic estimate of the average stability of personality traits over a 40- year interval, so substantial increases in estimates of stability for these traits seemed unlikely. The traits most likely to show increases in stability coefficients due to improvements in measurement are Agreeableness, Intellect/Openness, and Neuroticism.

Methods

Child Participants and Procedures

The original child cohort consisted of 2,404 elementary-school students from six samples collected by John M. Digman between 1959 and 1967. Detailed descriptions of these samples are given in Goldberg (2001). The three largest samples were collected in schools on the islands of Oahu and Kauai. Three smaller samples were assessed at the University of Hawaii Laboratory School. Wherever possible we have maximized the size of our adult sample by including participants from all six of the child samples. For some analyses we have narrowed the focus to participants drawn from the three largest of these samples, consisting of 2,221 children in elementary school on the islands of Oahu and Kauai. The vast majority of the children were in the 1st, 2nd, 5th, or 6th grades when they were assessed. The same core set of 39 traits was assessed across all schools, with some variation among schools in terms of additional traits. The teachers were provided with sheets of paper on each of which a particular trait label was printed at the top; the teachers were instructed to rank all the students in their classroom from the highest to the lowest on that attribute. Rectangular boxes set out in the form of a fixed nine-step quasi-normal distribution were included on each sheet for recording the students' names. Each attribute consisted of a single word or short phrase (e.g., gregarious), followed by a more extensive definition. Definitions for each personality trait had been developed beforehand from focus groups of teachers who had been asked to provide typical examples of classroom behaviors relating to that concept. 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 a listing of the 39 core items organized by Big Five trait, refer to appendix A.

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 elsewhere (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, and cognitive measures, and the personality interview (SIFFM). 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 and the clinic visit before the end of 2011. For most of these participants, the interval of time between the teacher assessments during childhood and the completion of the adult questionnaires and personality interview at the clinic visit was approximately 40 years. The adult sample represents the ethnic diversity of Hawaii and participants identified themselves as Japanese Americans (36%), Hawaiian or part Hawaiian (19%), Caucasian (16%), Filipino (10%), or Chinese (7%), Okinawan (4%), Latino (2%), Korean (1%), or other Pacific Islander (1%). A small number of participants (5%) identified as other (1%) or declined to respond (5%).

Representativeness of the Adult Sample to the Child Cohort

To address questions of whether 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. We found virtually no range restriction for the adult sample on any of the child factors (Hampson & Goldberg, 2006). Similarly we found very little evidence of bias, with the possible exception of Conscientiousness where the adult sample is about one-tenth of a standard deviation higher than the cohort mean. In evaluating this small effect, one must keep in mind that this is probably at least partially an artifact of mortality. 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.

Measures of Child Personality Traits

Since the item content varied across the six childhood samples, factor scores were derived separately within each subsample using all of the variables available for each sample. Goldberg (2001) described two sets of Big Five factor scores using the childhood items; one derived using maker sets that clearly establish the five factor structure in the childhood subsamples, and a second set of factor scores using all of the available items in each sample. The extended factor scores showed high convergence with those based on markers, and additionally provided more robust associations with adult outcomes. We have used the factor scores based on the extended items for all subsequent analyses. An advantage of using the extended factors scores is that this maximizes the number of items for each trait, improving the measurement of traits that would otherwise be measured with few items (see Appendix A). For example, using the extended factor scores the number of items for Extraversion ranged from 6 to 14 items, while the number of items assessing Neuroticism ranged from 3 to 16 items across samples. Alpha reliabilities for the child factors scores were estimated following methods described by Ten Berg and Hofstee (1999). Across the six childhood subsamples, the mean alphas for each trait were; .75 (Extraversion), .62 (Agreeableness), .77 (Conscientiousness), .68 (Neuroticism), and .60 (Intellect/Openness). These measures of the childhood Big Five have consistently shown validity for predicting adult outcomes, such as health status, health behaviors, educational attainment, and occupational environments (Hampson, Goldberg, Vogt, & Dubanoski, 2006; Hampson et al. 2007; Woods & Hampson, 2010).

Measures of Adult Personality Traits

Adult self-reports—The first adult questionnaire (Q1) included demographic variables and the 44 items of the Big Five Inventory (BFI; John & Srivastava, 1999). The BFI items were administered with a 5-point response scale, ranging from 1 (*very inaccurate*) to 5 (*very accurate*) as self-descriptors. We administered a second questionnaire (Q2) between 2 and 4 years after the first, with a mean interval between assessments of 2.8 years. Q2 was a two-sided single-page survey. On one side we administered the 44 BFI items; the other side included the 40 trait adjectives from Saucier's Mini-Markers (SMM; Saucier 1994), and an adult version of the teacher-assessed childhood variables that included the 39 common traits. Behavioral definitions of the child items, which had been given to the teachers, were not provided. The combined set of items was administered with a 5-point scale, ranging from 1 (*very false*) to 5 (*very true*). All three inventories provide measures of each of the Big Five factors. From these we also generated two sets of composite measures: (a) Factor scores derived from an analysis of the BFI items across both measurement occasions; and (b) Big Five factor scores using all 126 variables from the second adult questionnaire.

Alpha reliabilities for the first administration of the BFI were .84 (Extraversion), .78 (Agreeableness), .80 (Conscientiousness), .82 (Neuroticism), and .79 (Intellect/Openness). Alpha reliabilities for second administration of the BFI were .84 (Extraversion), .80 (Agreeableness), .81 (Conscientiousness), .85 (Neuroticism), and .83 (Intellect/Openness).

The alpha reliabilities for the SMM administered as part of Q2 were .79 (Extraversion), .81 (Agreeableness), .84 (Conscientiousness), .78 (Neuroticism), and .80 (Intellect/Openness). Alpha reliabilities for Big Five scores derived from 39 child items were .65 (Extraversion), .71 (Agreeableness), .74 (Conscientiousness), .79 (Neuroticism), and .65 (Intellect/Openness).

The Structured Interview for the Five-Factor Model (SIFFM)—The SIFFM was originally designed to provide a dimensional measure of normal personality traits that could be used as a bridge between the five-factor model of personality and clinical semi-structured interviews designed to tap specific Axis-II psychiatric disorders (Trull & Widiger, 1997, Trull et al., 1998). Semi-structured clinical interviews are designed to elicit specific information relevant to clinical judgments, while avoiding evaluative biases inherent in self-reports. Rather than focusing on a single specific disorder, the SIFFM applies this rationale to assessing personality traits. Its 120 items were directly modeled after those in the NEO-PI-R (Costa & McCrae, 1992). Each of the questions includes an associated follow-up probe (e.g., "Do you worry a lot about the future or things that might go wrong? If yes, what kinds of things do you worry about? What proportion of the time?"). Each item is scored on a 3-point scale where 0 indicates the participant said *no* to the question, and 1 and 2 indicate two increasing levels of strength of a *yes* answer, depending on the response to the follow-up probe. The SIFFM assesses the five domains of the NEO-PI-R using 24 items for each domain.

Participants who took part in the clinic assessment were administered this semi-structured personality interview by a trained interviewer. The interview typically took about an hour to complete. A subset of the interviews (n =183) were scored independently both by the interviewer and an observer who sat in during the interview. Inter-rater agreement was high, with correlations between raters ranging between .79 and 1.00; 59% of these rater pairs demonstrated perfect congruence in their ratings. Alpha reliabilities for the SIFFM domain scores were: .80 (Extraversion), .62 (Agreeableness), .72 (Conscientiousness), .79 (Neuroticism), and .73 (Intellect/Openness).

Additional interviewer personality ratings—In addition to providing a structured interview assessment of personality traits, the SIFFM provided a standardized context in which to elicit personality-relevant data. After the participant completed the SIFFM and three tests from the Woodcock-Johnson cognitive test battery (Woodcock-Johnson, McGrew, & Mather, 2001; not included in the present report), the interviewer described the participant using the 44-item BFI, Saucier's 40-item mini-markers, and the same 39 personality traits common to three largest subsamples in the childhood Hawaii cohort. For these interviewer assessments we used the same rating scales as the self-report versions. Each of these measures provides Big Five scores. Alpha reliabilities for the BFI interviewer ratings were high: .91 (Extraversion), .92 (Agreeableness), .89 (Conscientiousness), .90 (Neuroticism), and .91 (Intellect/Openness). Big Five scores on the mini-markers demonstrated similarly high alpha reliabilities: .90 (Extraversion), .92 (Agreeableness), .90 (Conscientiousness), .77 (Neuroticism), and .87 (Intellect/Openness). Big Five scores derived from the 39 child traits were very slightly lower: .79 (Extraversion), .84 (Agreeableness), .83 (Conscientiousness), .74 (Neuroticism), and .76 (Intellect/Openness).

Analyses

We first examined the convergence between the SIFFM scores and both adult self-reports and interviewer ratings on the BFI, SMM, the 39 traits, and aggregates of these. Next, we evaluated the convergence between the interviewer ratings and self-reported Big Five factor scores that aggregated across the available measures. Having tested the degree of convergence in our adult measures, we then used the SIFFM and the additional interviewer ratings to estimate 40-year stability coefficients using the child personality assessments. We also used structural equation modeling (SEM) to model the stability of each of the Big Five factors using interviewer ratings on the 39 childhood items common to the two largest childhood subsamples. SEM has the added advantage of modeling and correcting for measurement error. Next we compared the average stabilities at the Big Five level across self- and interviewer-ratings, and across different measures. This allowed us to contrast long-term stabilities using self-versus interviewer-ratings in adulthood and across measures that used the same versus different trait content at both time points. To the degree that the previous findings reported by Hampson and Goldberg (2006) were subject to confounding resulting from different measurement methods, we predicted that using the same method and same measure would give the highest stability estimates. In contrast, using a different method and different measure should give lower stability estimates. Other combinations of measures and methods should produce intermediate levels of stability.

Results

SIFFM Convergence with Other Adult Measures

We tested the convergence of the SIFFM scores with self-reports on the BFI, the SMM, the 39 childhood traits, and an aggregate of the three adult inventories. These correlations are reported in Table 1. Conscientiousness and Extraversion showed the highest convergent correlations, ranging from .50 to .61. Agreeableness demonstrated the lowest convergences, ranging from .26 to .47. The remaining correlations fell in the range of .40 to .54. Overall, the SIFFM showed substantial convergent correlations with self-reports, with Agreeableness showing the lowest levels of convergence.

We also correlated the SIFFM scores with the interviewer ratings on the BFI, SMM, the set of 39 traits used in the childhood assessments, and an aggregate using all of the items from these three inventories. These correlations are also reported in Table 1. Once again Agreeableness showed the lowest level of convergence, ranging from .29 to .41 across measures. For the remaining factors convergent correlations ranged from .48 to .59.

Convergence between the Interviewer-Ratings and the Self-Reports

We next evaluated the associations between self-ratings on the BFI and interviewer-ratings on the BFI, the SMM, the 39 child traits, and an aggregate of these. These correlations are reported in Table 2. The self-versus other correlations ranged from .27 for Neuroticism to . 59 for Extraversion. Neuroticism and Agreeableness show the lowest levels of convergence. The degree to which the self- and interviewer-ratings converged was lower than that between the SIFFM and either of the other two adult sources.

Forty-Year Stability

We correlated the childhood Big Five personality factors derived from the teacher assessments with four observer measures of the Big Five in adulthood: the SIFFM scores, and the interviewer ratings on the BFI, SMM, the 39 traits used in childhood, and an aggregate of these three measures. We report these child-adult correlations in Table 3. For comparison purposes, the stability coefficients between the childhood Big Five and adult self-reported Big Five are shown on the last two columns of Table 3.

Across all of the adult interviewer ratings, Intellect/Openness showed the highest long-term stability correlations (.22 to .25). Extraversion and Conscientiousness demonstrated very slightly lower levels of stability (Conscientiousness ranged from .20 to .22; Extraversion varied between .21 and .22). At the other extreme, Agreeableness demonstrated a low level of stability and Neuroticism demonstrated zero long-term stability.

Child-adult stabilities based on the SIFFM demonstrated a somewhat different pattern. Agreeableness demonstrated a level of stability similar to that of Extraversion (r = .19 for Agreeableness vs. r = 20 for Extraversion). Conscientiousness and Intellect/Openness demonstrated a moderate level of stability (r = .17). Neuroticism once again showed zero long-term stability.

SEM Estimates of Forty-Year Stability

We constructed separate models for each Big Five trait using teacher assessments in childhood and interviewer ratings in adulthood on the same 39 traits to model stability coefficients in SEM. We allowed indicators to have correlated errors within but not across measurement occasions. Standardized path estimates of the longitudinal stability of each of the Big Five based on informant ratings, along with estimates from our previous analysis (Hampson & Goldberg, 2006) and fit indices of the new models, are presented in Table 4. All of the five models achieved acceptable fit (CFIs range from .97 to .99; RMSEAs range from .03 to .05). Extraversion showed the highest level of stability (β = .38) followed by Intellect/Openness (β = .29) and Conscientiousness (β = .26). Agreeableness demonstrated a lower level of stability (β = .12), and Neuroticism showed zero stability.

Comparison of Forty-Year Stability Estimated for Self- and Interviewer-Ratings

In summary, across all of the methods we used, the long-term stability correlations based on interviewer-ratings generally were as high as those based on self-reports, and in some cases yielded higher coefficients. Using SIFFM scores to estimate stabilities resulted in a dramatic increase in the stability estimate for Agreeableness compared to either self-ratings or interviewer assessments. The stability of Conscientiousness was slightly lower based on the SIFFM scores than with other methods, but it was generally consistent across the other measures (r = .20 to r = .25). The zero stability of Neuroticism was consistent across all methods and measures. When 40-year stability was estimated in an SEM framework using interviewer ratings, Agreeableness, Conscientiousness and Neuroticism showed similar levels of stability to those derived from self-reports. Extraversion and Intellect/Openness show higher SEM based stabilities using interviewer ratings. The increase in estimated

stability was greatest for Intellect/Openness ($\beta = .29$ versus $\beta = .12$), a statistically significant increase in the stability coefficient ($z = 3.17$, $p < .05$).

Crossing Method and Measure

We compared the average stability correlations for each of the Big Five factors across four adult measurement conditions differing on whether the assessment in adulthood was by self- or observer-report and whether the measure used had the same or different trait content as the childhood measure (self-report, different measure; self-report, same measure; observer-report, different measure; observer-report same measure). Across those four conditions the average 40-year stability correlations for the Big Five factors ranged from .14 to .16. Clearly this difference is far too small to indicate a gross methodological confounding. Consequently, the results ruled out any methods confound large enough to account for the zero stability obtained for Neuroticism.

Summary of Main Findings

The childhood versus adulthood stability of measures derived from interviewer/observer-ratings were generally as high as or higher than those based on self-reports for all of the Big Five factors. Across all measures and methods, Extraversion and Conscientiousness displayed substantial levels of child-adult stability. Equally consistent was the absence of stability for Neuroticism. The average stabilities for each of the Big Five factors across the different assessment modalities did not differ to any large degree, which was inconsistent with any systematic effects of measurement-method confounding across all the Big Five. Nevertheless, close comparison of stabilities achieved by the different measurement methods revealed effects for two of the Big Five factors. Most strikingly, the factor of Agreeableness showed a consistently low level of stability for both interviewer ratings and self-reports, but a markedly higher level of stability when assessed by the SIFFM. Intellect/Openness showed higher levels of stability when assessed by the interviewer ratings than by any other method. This difference was most pronounced in the SEM analysis, where observer-ratings on the 39 childhood traits resulted in a large increase in the stability of Intellect/Openness.

Discussion

In the present investigation, we estimated child-adult stability coefficients using a variety of measures in adulthood: self-reports, scores from the SIFFM, and interviewer ratings based on the BFI, the SMM, and the 39 child traits. The method that was closest to the childhood teacher assessments was the interviewer assessments using the 39 childhood traits. However, matching methods in this way did not yield a systematic increase in childhood-adult stability coefficients. Indeed, methodological differences in the adult assessments made remarkably little difference overall. The failure to improve upon the stability coefficients for Extraversion and Conscientiousness was consistent with our expectation that these factors were already at or near the likely maximum level of measurable stability as indicated by other research (Roberts & DelVecchio, 2000). More significant was the confirmation of zero long-term stability for Neuroticism. Most puzzling were the differences in stability among measurement methods observed for Agreeableness and Intellect/Openness. Below, we discuss possible explanations for each of these findings, drawing on the concepts of evaluativeness and observability that have proved useful for research on person perception and judgment accuracy (Funder, 1995, 1999; Vazire, 2010).

Reliability of the Childhood traits

The childhood factors scores tended to have lower reliability than the adult trait scores, with Agreeableness, Intellect/Openness, and Neuroticism, showing the lowest reliability. When

deriving orthogonal factors scores, later derived factors necessarily account for less variance, and as a consequence show lower alpha reliability. Performing an orthogonal rotation tends to redistribute variance across factors, but the total variance accounted for, and the total sum of alpha reliability across factors remains conserved (Ten Berg & Hofstee, 1999). This may have accounted for the low alpha reliability reported for the childhood factor scores of these three traits. Since they also tended to show the lowest stability over time, this raise the question of the degree to which our stability estimates are attenuated by low reliability in the childhood factor scores. In some cases, the extracted factors had few items. For example, Neuroticism is measure buy just 3 items in the Kauai sample factor score. This may also account for lower reliability in this trait.

In contrast, in the SEM analyses each trait was modeled separately. As a result the variance for each trait was maximized without regard to overall orthogonally. Furthermore, SEM explicitly estimates error and produces path estimates that correct for this. Comparing the SEM results with those using the factor scores gives an indication of the degree to which measurement reliability may have attenuated stability estimates. In our SEM models, the estimated stabilities tend to be higher for all of the traits with the exception of Neuroticism. Intellect/Openness and Extraversion showed the largest increases in estimated stability estimates using SEM models versus factor scores, suggesting that these two traits were the most affected by measurement error. It is important to note that in contrasting stability estimates based on different methods in adulthood all of our estimates using the child factor scores are equally affected by any attenuation resulting from the unreliability in the child measures.

Agreeableness and Evaluativeness

Near-zero levels of stability for Agreeableness were obtained for all observer-reports except those based on the SIFFM. One possible explanation for the SIFFM finding is related to the high evaluativeness of Agreeableness, and hence to the potential for bias by both the self and observers. When rating highly evaluative traits, self-reports may elicit self-enhancement biases, whereas observers may fall prey to leniency biases. These both represent evaluative biases. In both cases, neither the self nor the interviewer cares to cast aspersions, with the result that assessments of this factor tend to be truncated, and consequent correlations with any other variables tend to be attenuated. The SIFFM was designed to be a clinical interview for assessing the NEO domains that bridges the assessment of normal personality traits and personality disorders. As such, the questions and follow-up probes permit the interviewee to provide examples of undesirable characteristics. For example, a question assessing the Altruism facet of Agreeableness asks “Are you generally very reluctant to get involved in the problems of other people?” and the probes distinguish between whether or not the interviewee thinks other people see this tendency as selfish or not selfish. The clinically sensitive nature of the questions in the SIFFM may have made it easier for interviewees to describe themselves in more undesirable ways. The clearly defined coding scheme made it difficult for interviewers to code leniently. In support of this explanation, the convergence among adult measures for this factor was always as low or lower than for any of the other Big Five factors. Especially low were the convergence coefficients between the SIFFM domain score for Agreeableness and the self-report and observer trait ratings of this factor on the BFI, SMM, and the 39 traits.

Of all the measures in adulthood, SIFFM Agreeableness also had the lowest reliability (alpha = .62). It may appear puzzling that a relatively low reliability measure achieved the highest stability with respect to this trait. One possibility is that the alpha reliability is an underestimate of the true reliability of the measure, which tends be the case when measures violate the assumption of tau-equivalence across items (Miller, 1995). This may be especially true of the SIFFM items, since the format of follow-up questions and hence the

meaning of the rating scale used, varies across items. Another alternative is that the low alpha reliability accurately reflects low interrelatedness across SIFFM Agreeableness items in our sample, and a broader bandwidth of trait assessment. This interpretation is consistent with the low convergence of SIFFM Agreeableness with our other adult measures. In this case, both the low alpha reliability and the higher stability observed with this measure may have resulted from specific items carrying unique variance. While the convergent correlations could equally be attenuated by low reliability, the high observed stability using this measure argues in favor of the high bandwidth interpretation (Hogan & Roberts 1996).

Intellect/Openness and Observability

In contrast to Agreeableness, SIFFM scores did not increase the long-term stability of Intellect/Openness. However, observer-ratings of this domain for the other adult measures demonstrated higher levels of stability than did self-reports on these same measures. Why? Some clues are available from studies that have given observers access to unusual personality information. In studies where observers made judgments based on cues from targets' bedrooms, offices, and websites, Intellect/Openness emerged as the most judgeable trait (Gosling, Ko, Mannarelli, & Morris, 2002; Vazire & Gosling, 2004). These findings highlight the functional value of valid cues. After the SIFFM, the interviewers administered three subtests from a cognitive-ability battery (Woodcock-Johnson, McGrew, & Mather, 2001). This experience provided the interviewers with trait-relevant information that was particularly informative for the assessment of Intellect. The information provided by these cognitive tests is also similar to the kinds of information on intellectual abilities available to the teachers who performed the original child personality assessments. Thus, the increase in stability estimates for Intellect/Openness on all the adult observer-ratings, but not the SIFFM, may have resulted from the enhanced observability of this trait inadvertently provided to the interviewers by administering and scoring the cognitive assessments.

Neuroticism

The complete lack of long-term stability for Neuroticism contradicts the model of asymptotic decline in personality traits over time proposed by Fraley and Roberts (2005). Their model proposes that the stability of personality declines over increasing intervals of time in a non-linear fashion, stabilizing at a level greater than zero. It has also been suggested that whether or not Neuroticism stabilizes at a non-zero asymptote is a key feature of models of long-term stability that focus on changes resulting from life experiences (Ormel et al., 2012). In light of these considerations, the replication of zero stability for Neuroticism across both self- and observer-reports in adulthood merits careful consideration.

Although widely viewed as the least observable of the Big Five factors (Funder & Dobroth, 1987), the traits selected for the original teacher ratings in childhood were specifically chosen to include behaviors that could be observed by teachers (e.g., Fidgets, Fearful, Touchy), and the teachers were given behavioral definitions of the traits. However, it is conceivable that none of the adult measures of Neuroticism tapped the same aspects of the domain that the teachers judged. Another explanation for the zero stability for Neuroticism is that teachers are unable to make reliable and valid judgments of this trait. However, a Neuroticism factor was recovered in analyses of all six subsamples comprising the Hawaii childhood sample, and previous studies of teachers' ratings of elementary school children's personality traits have identified reliable Neuroticism factors in other samples (e.g., Digman & Shmelyov, 1996; Hagekull & Bohlin, 2003; Shiner & De Young, in press). Project Competence (Shiner & Masten, 2012) provides strong evidence for the predictive validity of childhood Neuroticism measured at age 10 by a combination of teacher, parent, and observer ratings. Neuroticism predicted outcomes at age 20 (lower academic attainment, less rule abiding conduct, less social competence) and age 30 (lower work competence, lower

romantic competence). In the Hawaii sample, higher levels of childhood Neuroticism predicted greater alcohol use 40 years later for both men and women, and men who were less neurotic as children had higher BMIs as adults (Hampson, Goldberg, Vogt & Dubanoski, 2006). We conclude that Neuroticism is a valid personality construct for children in middle childhood.

Many mechanisms have been proposed as drivers of both personality change and personality stability over time (see Caspi & Roberts 2001; Roberts & Jackson 2008 for reviews). These include biological factors, life events and environmental factors, goals and motives, and interactions among these. Of all of the mechanisms that affect personality development across these stages of life, we must ask if Neuroticism is distinct in its formation as a trait, or the factors that affect it. One feature that distinguishes Neuroticism from the remaining Big Five is that one pole of the trait, namely being high on Neuroticism, is inherently aversive for the person experiencing it. As a result, we might expect individuals high on Neuroticism to be strongly motivated to change. Additionally, our childhood assessments predate by at least a decade the average age of onset of the most common forms of depression and anxiety (de Girolamo et al. 2012), disorders which are integral to lifespan pathways of Neuroticism at the individual level. While we can only speculate, these features together may contribute a high higher rate of change for Neuroticism in the span of time represented in our data.

A lack of stability across this 40 year interval has implications for how Neuroticism is likely to operate in lifespan models of development and influence consequential outcomes. If the effects of childhood Neuroticism are not evoked through a persistent stable component, they must be encoded in biological, cognitive, or behavioral structures outside the domain of Neuroticism that persist over long periods of time. Such encoding may occur during a critical period in childhood where levels of Neuroticism are especially impactful. Furthermore, instances where childhood Neuroticism impacts adult outcomes would necessarily depend on a mediator external to Neuroticism to carry the observed effect.

Limitations, Conclusions, and Future Directions

The findings reported here confirmed that personality stability coefficients differed considerably among the Big Five, even when equating the assessment method across childhood and adulthood. In addition, the findings suggest that measurement differences related to evaluativeness and observability may impact stability coefficients in ways similar to their effect on self- versus observer-reports. Although unlikely, it remains a possibility that these findings are unique to the Hawaii cohort. The lack of comparable studies spanning childhood to midlife makes generalizability difficult to determine.

We are not aware of prior work that has integrated the rich literature on person perception and judgment accuracy with models of long-term personality stability. Optimal personality assessment is vital for estimating levels of stability, and is also necessary for accurately characterizing personality change. Any estimate of the stability of personality is likely to be attenuated by error in measurement at each time point. As we have argued, varying rater perspectives and enhancing the availability of personality-relevant cues may serve to reduce some of this error. Here, we treated the child ratings as criteria and interpreted any increase in stability as an indication that we had reduced a specific source of rater error. However, from a wider perspective, low convergence across different rating perspectives and instruments is not necessarily best characterized as error. If the goal is maximal prediction of a non-trait outcome such as health status, we may find, for example, that different rating modalities provide increments of independent variance, each having unique predictive validity. Furthermore, in the context of a lifespan model of personality, it is possible that for any given trait and outcome, different modalities of personality measurement may have unique predictive validity for specific points in the lifespan. Future longitudinal research

incorporating repeated assessments of personality from multiple perspectives using multiple methods and relating these to consequential outcomes such as health should address these important questions.

Acknowledgments

This research was supported by grant AG020048 from the National Institute on Aging. The authors thank Melody Fo and CrisYamabe, and all members of the Lifestyle, Culture, and Health team at the Center for Health Research-Hawaii, for conducting the adult personality interviews. The authors are indebted to John (Jack) M. Digman (1923–1998) for obtaining the original teacher assessments of childhood personality traits. Appreciation is expressed to Brent W. Roberts, Patrick L. Hill, and Erika N. Carlson for their comments on an earlier draft of this manuscript.

References

Borkenau P, Brecke S, Möttig C, Paelecke M. Extraversion is accurately perceived after a 50-ms exposure to a face. Journal of Research in Personality. 2009; 43:703–706.10.1016/j.jrp.2009.03.00

Albright L, Kenny DA, Malloy TE. Consensus in personality judgments at zero acquaintance. Journal of Personality and Social Psychology. 1988; 55:387–395.10.1037/0022-3514.55.3.38 [PubMed: 3171912]

Brunswik, E. Perception and the representative design of psychological experiments. Berkeley: University of California Press; 1956.

Caspi A, Roberts BW. Personality development across the life span: The argument for change and continuity. Psychological Inquiry. 2001; 12:49–66.

Caspi A, Roberts BW, Shiner RL. Personality development: Stability and change. Annual Review of Psychology. 2005; 56:453–484.10.1146/annurev.psych.55.090902.14191

Costa, PT., Jr; McCrae, RR. Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment Resources; 1992.

Chang L, Connelly BS, Geeza AA. Separating method factors and higher order traits of the Big Five: A meta-analytic multitrait–multimethod approach. Journal of Personality and Social Psychology. 2012; 102:408–426.10.1037/a0025559 [PubMed: 21967007]

De Girolamo G, Dagani J, Purcell R, Cocchi A, McGorry PD. Age of onset of mental disorders and use of mental health services: needs, opportunities and obstacles. Epidemiology and Psychiatric Sciences. 1:1–11.

DeYoung CG. Higher-order factors of the Big Five in a multi-informant sample. Journal of Personality and Social Psychology. 2006; 91:1138–1151.10.1037/0022-3514.91.6.113 [PubMed: 17144770]

Digman JM. Higher order factors of the Big Five. Journal of Personality and Social Psychology. 1997; 73:1246–1256.10.1037/0022-3514.73.6.124 [PubMed: 9418278]

Digman JM, Shmelyov AG. The structure of temperament and personality in Russian children. Journal of Personality and Social Psychology. 1996; 71:341–351.10.1037/0022-3514.71.2.34 [PubMed: 8765485]

Fraley RC, Roberts BW. Patterns of continuity: A dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course. Psychological Review. 2005; 112:60–74.10.1037/0033-295X.112.1.6 [PubMed: 15631588]

Friedman HS, Tucker JS, Schwartz JE, Tomlinson-Keasey C, Martin LR, Wingard DL, Criqui MH. Psychosocial and behavioral predictors of longevity: The aging and death of the “Termites”. American Psychologist. 1995; 50:69–78.10.1037/0003-066X.50.2.6 [PubMed: 7879989]

Friedman HS, Tucker JS, Tomlinson-Keasey C, Schwartz JE, Wingard DL, Criqui MH. Does childhood personality predict longevity? Journal of Personality and Social Psychology. 1993; 65:176–185.10.1037/0022-3514.65.1.17 [PubMed: 8355139]

Funder DC. On the accuracy of personality judgment: A realist approach. Psychological Review. 1995; 102:652–670.10.1037/0033-295X.102.4.65 [PubMed: 7480467]

Funder, DC. Personality judgment: A realistic approach to person perception. San Diego, CA: Academic Press; 1999.

Funder DC, Colvin CR. Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. Journal of Personality and Social Psychology. 1988; 55:149–158.10.1037/0022-3514.55.1.14 [PubMed: 3418488]

Funder D, Dobroth K. Differences between traits: Properties associated with interjudge agreement. Journal of Personality and Social Psychology. 1987; 52:409–418.10.1037/0022-3514.52.2.40 [PubMed: 3559898]

Goldberg LR. Analyses of Digman's child-personality traits: Derivation of Big-Five factor scores from each of six samples. Journal of Personality. 2001; 69:709–743.10.1111/1467-6494.69516 [PubMed: 11575511]

Gosling SD, Ko SJ, Mannarelli T, Morris ME. A room with a cue: Personality judgments based on offices and bedrooms. Journal of Personality and Social Psychology. 2002; 82:379–398.10.1037/0022-3514.82.3.37 [PubMed: 11902623]

Hagekull B, Bohlin G. Early temperament and attachment as predictors of the Five Factor Model of personality. Attachment & Human Development. 2003; 5:2–18.10.1080/146167303100007864 [PubMed: 12745826]

Hampson SE, Dubanoski JP, Hamada W, Marsella AJ, Matsukawa J, Suarez E, Goldberg LR. Where are they now? Locating former elementary school students after nearly 4 years for a longitudinal study of personality and health. Journal of Research in Personality. 2001; 35:375–387.10.1006/jrpe.2001.231

Hampson SE, Goldberg LR. A first large-cohort study of personality-trait stability over the 4 years between elementary school and midlife. Journal of Personality and Social Psychology. 2006; 91:763–779.10.1037/0022-3514.91.4.76 [PubMed: 17014298]

Hogan J, Roberts B. Issues and non-issues in the fidelity/bandwidth tradeoff. Journal of Organizational Behavior. 1996; 17:627–637.

John, OP.; Srivastava, S. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In: Pervin, LA.; John, OP., editors. Handbook of personality: Theory and research. 2. New York: Guilford; 1999. p. 102-138.

Miller MB. Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. Structural Equation Modeling. 1995; 2:255–273.

Nave CS, Sherman RA, Funder DC, Hampson SE, Goldberg LR. On the contextual independence of personality: Teachers' assessments predict directly observed behavior after four decades. Social Psychological and Personality Science. 2010; 1:327–334.10.1177/194855061037071

Norman WT, Goldberg LR. Raters, ratees, and randomness in personality structure. Journal of Personality and Social Psychology. 1966; 4:681–691.10.1037/h002400

Ormel J, Riese H, Rosmalen JGM. Interpreting neuroticism scores across the adult life course: immutable or experience-dependent set points of negative affect? Clinical Psychological Review. 2012; 32:71–79.10.1016/j.cpr.2011.10.00

Roberts BW, DelVecchio WF. The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. sychological Bulletin. 2000; 126:3–25.10.1037/0033-2909.126.1.

Roberts BW, Jackson JJ. Sociogenomic personality psychology. Journal of Personality. 2008; 76:1523–1544. [PubMed: 19012657]

Saucier G. Mini-markers: A brief version of Goldberg's unipolar Big Five markers. Journal of Personality Assessment. 1994; 63:506–516.10.1207/s15327752jpa6303_8 [PubMed: 7844738]

Shiner, RL.; DeYoung, CG. The structure of temperament and personality traits: A developmental perspective. In: Zelazo, P., editor. Oxford handbook of developmental psychology. New York: Oxford University Press; in press

Shiner RL, Masten AS. Childhood personality as a harbinger of competence and resilience in adulthood. Development and Psychopathology. 2012; 24:507–528.10.1017/S095457941200012 [PubMed: 22559127]

Ten Berge JMF, Hofstee WK. Coefficient alpha and reliabilities of unrotated and rotated components. Psychometrika. 1999; 69:83–90.

Tetlock, PE. Toward an intuitive politician model of attribution processes. In: Schlenker, BR., editor. The self in social life. New York, NY: McGraw-Hill; 1984. p. 203-234.

Trull, TJ.; Widiger, TA. Structured interview for the Five-Factor Model of Personality. Odessa, FL: Psychological Assessment Resources; 1997.

Trull TJ, Widiger TA, Useda JD, Holcomb J, Doan BT, Axelrod SR, Stern BL, Gershuny BS. A structured interview for the assessment of the five-factor model of personality. Psychological Assessment. 1998; 10:229–240.10.1037/1040-3590.10.3.22

Vazire S, Gosling SD. e-Perceptions: Personality impressions based on personal websites. Journal of Personality and Social Psychology. 2004; 87:123–132.10.1037/0022-3514.87.1.12 [PubMed: 15250797]

Vazire S. Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. Journal of Personality and Social Psychology. 2010; 98:281–300.10.1037/a001790 [PubMed: 20085401]

Watson D. Strangers' ratings of the five robust personality factors: Evidence of a surprising convergence with self-report. Journal of Personality and Social Psychology. 1989; 57:120–128.10.1037/0022-3514.57.1.12

Woodcock, RW.; McGrew, KS.; Mather, N. Woodcock-Johnson III Tests of Cognitive Abilities. Itasca, IL: Riverside Publishing; 2001.

Appendix A

The 39 Childhood Variables Common to the Three Largest Childhood Samples Listed by Big Five Trait Domain

Childhood items		
Extraversion	Agreeableness	Conscientiousness
Energetic	Assertive	Careful of personal belongings
Gregarious	Complains about others	Careless of others' property
Happy	Considerate	Conscientious
Lethargic	Impulsive	Eccentric
Seclusive	Jealous	Fickle
	Outspoken	Fidgets
	Rude	Irresponsible
	Self-minimizing	Mannerly
	Spiteful	Neat in appearance
	Submissive	Nervous habits
	Touchy	Persevering
		Planful
		Restless
Neuroticism	Openness	
Concerned about acceptance	Adaptable	
Fearful	Curious	
	Esthetically sensitive	
	Imaginative	
	Original	
	Rigid	
	Socially confident	
	Verbally fluent	

Note: Factor analytic results of the 39 childhood common variables can be found in Hampson & Goldberg 2006

Highlights

We tested Big-Five stability over 40 years using observer ratings in adulthood.

Overall results are similar to those using self-ratings in adulthood.

Intellect/Openness showed highest stability when using observer ratings.

Agreeableness showed highest stability using a structured interview in adulthood.

Neuroticism showed zero stability across all methods.

Table 1

Convergence between Scores from the Structured Interview for the Five Factor Model (SIFFM) Scores and Self- and Interviewer-Ratings

	Adult Self-Reports				Interviewer-Ratings			
SIFFM	**BFI[a] *n* = 579**	**SMM *n* = 554**	**39 Child traits *n* = 554**	**Aggregated Set[b] *n* =554**	**BFI *n* = 779**	**SMM *n* = 779**	**39 Child Traits *n* = 779**	**Aggregated Set[b] *n* = 779**
Extraversion	.59	.50	.50	.56	.57	.54	.52	.56
Agreeableness	.47	.40	.26	.47	.36	.32	.29	.41
Conscientiousness	.59	.60	.56	.61	.49	.53	.48	.53
Neuroticism	.53	.40	.44	.48	.57	.57	.50	.55
Intellect/Openness	.54	.47	.47	.51	.59	.49	.48	.45

Note:

[a]BFI self-reports are aggregated factor scores from two administrations of this measure approximately three years apart.

[b,c]Aggregated scores derived from BFI, SMM, and 39 child items. SIFFM = Structured Interview for the Five Factor Model; BFI = Big Five Inventory; SMM = Saucier's Mini-Markers; All correlations are significant at the $p < .05$ level.

Table 2

Convergence between the Self-Reports on the BFI and the Interviewer Ratings on the 39 Child Traits, BFI, SMM, and an Aggregate of These Inventories

Self-Reports	Interviewer-Ratings			
BFI[a]	BFI	SMM	39 Child Traits	Aggregate[a]
Extraversion	.53	.54	.50	.59
Agreeableness	.25	.24	.22	.29
Conscientiousness	.38	.46	.35	.45
Neuroticism	.38	.35	.27	.35
Intellect/Openness	.46	.45	.41	.44

Note: n = 579. BFI = Big Five Inventory; SMM = Saucier's Mini-Markers;

[a]Aggregate ratings include items from the BFI, SMM, and 39 child trait. All correlations are significant at the $p < .05$ level.

Table 3

Forty-Five Year Stability Correlations from Childhood to Adulthood for Each of the Big Five Factors: Comparing Interviewer Ratings and Self-Reports in Adulthood with Teachers' Assessments in Childhood.

Teacher Assessments In Childhood	Interviewer-Ratings					Self-Reports[a]	
	39 Child Traits *n* = 771	BFI *n* = 771	SMM *n* = 771	Aggregate[b] *n* = 771	SIFFM *n* = 780	BFI[c] *n* = 916	Aggregate[d] *n* = 916
Extraversion	**.21**	**.22**	**.21**	**.22**	**.20**	**.29**	**.27**
Agreeableness	**.07**	**.08**	.06	**.09**	**.19**	**.08**	**.09**
Conscientiousness	**.20**	**.20**	**.20**	**.22**	**.17**	**.23**	**.25**
Neuroticism	−.02	−.04	−.06	.04	−.03	.00	.00
Intellect/Openness	**.22**	**.24**	**.25**	**.24**	**.17**	**.16**	**.17**

Note: BFI= Big Five Inventory; SMM= Saucier's Mini-Markers;

[a]Previously reported in Hampson & Goldberg 2006.

[b,d]Aggregate ratings include items from the BFI, SMM, and 39 child traits.

[c]BFI self-reports are aggregated factor scores from two administrations of this measure approximately three years apart. Correlations in bold differ significantly from zero at the $p < .05$ level.

Table 4

Standardized Path Coefficients from SEM Models Relating Childhood Personality Factors to Adult Self- and Informant-Ratings on the Same 39 Traits

	Interviewer *n*= 690			Self[a] *n* = 762
Big Five	**CFI**	**RMSEA**	***β***	***β***
Extraversion	.97	.05	**.38**	**.30**
Agreeableness	.98	.05	**.12**	**.14**
Conscientiousness	.98	.04	**.26**	**.26**
Neuroticism	.98	.03	.01	−.02
Intellect/Openness	.99	.03	**.29**	**.12**

Note:

[a]Results are based on self-reports, along with fit indices for each model previously reported in Hampson & Goldberg (2006). Standardized path coefficients in bold differ significantly from zero at the $p < .05$ level.