Journal of Personality and Social Psychology
2008, Vol. 94, No. 4, 718–737

0022-3514/08/$12.00 DOI: 10.1037/0022-3514.94.4.718

The Developmental Psychometrics of Big Five Self-Reports: Acquiescence, Factor Structure, Coherence, and Differentiation From Ages 10 to 20

Christopher J. Soto and Oliver P. John
University of California, Berkeley

Samuel D. Gosling
University of Texas at Austin

Jeff Potter
Cambridge, Massachusetts

How do youths' personality reports differ from those of adults? To identify the year-by-year timing of developmental trends from late childhood (age 10) to early adulthood (age 20), the authors examined Big Five self-report data from a large and diverse Internet sample. At younger ages within this range, there were large individual differences in acquiescent responding, and acquiescence variability had pronounced effects on psychometric characteristics. Beyond the effects of acquiescence, self-reports generally became more coherent within domains, and better differentiated across domains, at older ages. Importantly, however, different Big Five domains showed different developmental trends. Extraversion showed especially pronounced age gains in coherence but no gains in differentiation. In contrast, Agreeableness and Conscientiousness showed large age gains in differentiation but only trivial gains in coherence. Neuroticism and Openness showed moderate gains in both coherence and differentiation. Comparisons of items that were relatively easy versus difficult to comprehend indicated that these patterns were not simply due to verbal comprehension. These findings have important implications for the study of personality characteristics and other psychological attributes in childhood and adolescence.

Keywords: five-factor model, personality structure, reliability, validity

The years spanning late childhood and adolescence constitute a time of considerable biological, social, and psychological change. These rapid developments raise questions about the capacity of children and adolescents to report about their own personality traits, that is, their characteristic patterns of thoughts, feelings, and behavior. At what age can youths provide meaningful personality self-reports? In what ways do the reports of youths differ from those of adults? Is it easier for young respondents to report about some personality characteristics than about others? These basic questions are crucial for theorists, researchers, and practitioners interested in personality, emotional, and social development prior to adulthood. Definitive answers, however, are not yet available. In this article, we derive specific hypotheses about the characteristics of youths' personality self-reports. We then test these hypotheses by using data from a large and diverse sample of children, adolescents, and young adults who provided self-reports for the Big Five personality trait domains. The large sample size allowed separate analyses—with high statistical precision—at each year of age from 10 to 20.

Important Developmental Changes in Late Childhood and Adolescence

Over the course of late childhood and adolescence, youths experience several developmental changes with implications for the capacity to provide personality self-reports.

Changes in Cognitive Capacities and Self-Concept

Late childhood (often defined approximately as ages 8–11) and adolescence (often defined approximately as ages 12–17) have long been recognized as periods of change in fundamental reasoning abilities (Inhelder & Piaget, 1958). For example, compared with younger children, adolescents have greater capacities to evaluate the logical consistency of statements (Osherson & Markman, 1975) as well as to think abstractly and consider the hypothetical, rather than being tied to the concrete and immediate (Flavell, Miller, & Miller, 1993). Increases in capacities for abstract reasoning are accompanied, not coincidentally, by capacities and motivations to think about the self and consider questions of identity: Who am I? Am I a worthwhile person? What makes me different from others? Indeed, the search for a coherent self-

Christopher J. Soto and Oliver P. John, Department of Psychology and Institute of Personality and Social Research, University of California, Berkeley; Samuel D. Gosling, Department of Psychology, University of Texas at Austin; Jeff Potter, Cambridge, Massachusetts.

Christopher J. Soto's work on this project was supported by a Chancellor's Opportunity Predoctoral Fellowship from the University of California, Berkeley, and by a Graduate Research Fellowship from the National Science Foundation; Christopher J. Soto and Oliver P. John also received support from National Institute of Mental Health Grant MH-43948 and a University of California Faculty Research Grant. We thank Ravenna Helson, Sheldon Zedeck, and Anne E. Cunningham for their helpful comments on a draft of this article.

Correspondence concerning this article should be addressed to Christopher J. Soto, Institute of Personality and Social Research, University of California, Berkeley, 4143 Tolman Hall, MC 5050, Berkeley, CA 94720-5050. E-mail: cjsoto@berkeley.edu

concept is often considered the central psychological task of adolescence (Erikson, 1968).

As adolescents work on the task of identity development, their self-concept changes; these changes have received considerable and ongoing attention from developmental and educational psychologists. For example, one recent study found that correlations among different self-aspects (e.g., skill and interest in sports vs. quality of relationship with parents) decreased with age in a sample of French students in Grades 2 through 6 (Marsh & Ayotte, 2003). More generally, compared with younger children, adolescents are more likely to think about and describe themselves with abstract and psychological terms (e.g., "I am popular and athletic" vs. "I have a friend named Betty and I like baseball"), to differentiate among multiple aspects of the self, to recognize consistencies and inconsistencies among these self-aspects, and to organize them in a clear way (Byrne & Shavelson, 1996; Donahue, 1994; Harter, 1999, 2006; Harter & Monsour, 1992; Marsh, 1989; Marsh & Ayotte, 2003; Montemayor & Eisen, 1977). By contrast, in the study of personality traits (which we review below), age trends in coherence and differentiation have received much less attention.

Changes in Verbal Comprehension

Late childhood and adolescence are times of ongoing gains in verbal capacities. The continued growth of vocabulary throughout childhood and adolescence (Smith, 1941) includes the addition of new personality-relevant words, as well as increasingly frequent and complex use of such words (Donahue, 1994; Yuill, 1993). There are also changes in key reading comprehension skills (Siegler, 1998). Adolescents are better able to monitor their own comprehension and take corrective action when a passage is not fully understood (Garner, 1990). Such skills can help adolescents make sense of many kinds of texts, including personality questionnaires.

Changes in Basic Personality Trait Structure

Over the course of childhood and adolescence, the basic organization of cognition, affect, and behavior itself may change. Recent reviews of the available literature (Caspi & Shiner, 2006; Shiner, 2006) have offered two conclusions. On the one hand, some researchers have moved toward accepting the Big Five as a representation of basic personality trait dimensions in childhood and adolescence (e.g., Digman, 1994; Goldberg, 2001; Halverson et al., 2003; Kohnstamm, Halverson, Mervielde, & Havill, 1998). On the other hand, several findings are not easily reconciled with the canonical Big Five structure (e.g., Hampson & Goldberg, 2006; John, Caspi, Robins, Moffitt, & Stouthamer-Loeber, 1994; Lamb, Chuang, Wessels, Broberg, & Hwang, 2002; van Lieshout & Haselager, 1994), and further research is needed before the issue can be considered settled.

A pair of studies examined the factor structure of ratings of children's and adolescents' personality traits by adult informants with an instrument specifically designed for the description of child personality, namely the California Child Q-set (Block & Block, 1980). Their findings, emerging in samples from the United States (John et al., 1994) and the Netherlands (van Lieshout & Haselager, 1994), converge in suggesting age differences in trait structure. In both studies, Extraversion did not cohere as it typically does in ratings of adults; instead, it was represented by two distinct components.

A third study employing California Child Q-set parent reports followed 102 Swedish children from early childhood into adolescence (Lamb et al., 2002). Although the sample size of this longitudinal study was not large enough to provide robust factor-analytic results, the internal consistency of the Extraversion scale developed by John et al. (1994) increased substantially from early childhood to adolescence, as did its correlation with a California Child Q-set scale assessing activity level. Consistent with all of these results, Eaton (1994) argued that activity level is a distinct, basic dimension of personality during infancy and childhood that then retreats to lesser status as a facet of Extraversion by adulthood. Thus, the Extraversion domain may be fundamentally less cohesive in childhood and early adolescence than in adulthood, although the exact timing of this change is not yet well documented.

Testing Implications for Personality Self-Reports

The developmental changes reviewed above should affect how well children and adolescents can report about their own personalities. Here we discuss the implications of these changes for youths' personality self-reports, review previous literature that has examined such reports, identify open questions, and articulate specific hypotheses tested in the present research.

General Changes in Within-Domain Coherence

Within-domain coherence, or internal consistency, is the extent to which items intended to measure the same trait are highly correlated with one another; stronger correlations among items within a domain indicate greater coherence. In general, the within-domain coherence of personality self-reports should increase with age, for at least two reasons. First, most older adolescents and adults have available clear self-concepts that should provide useful reference points for making coherent judgments about their personalities. Second, older adolescents are better at evaluating issues of logical consistency, which should help them avoid contradictions in their personality reports (e.g., claiming that they are both generally talkative and generally quiet) and thereby strengthen relations among items within the same domain.

One previous study compared the coherence of self-reports on a Big Five measure across adolescent age groups. In a sample of Estonian secondary school students, the mean alpha reliability for the five scales of the 60-item NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992) was lower for 12-year-olds ($M = .72$) than for 16- and 18-year-olds ($M = .81$); the mean alphas for these older student samples were about as high as that observed in an adult sample ($M = .82$; Allik, Laidra, Realo, & Pullman, 2004). Another study found that NEO-FFI responses in a small sample of young Mexican Americans and European Americans ($N = 130$, age range = 10.00 to 12.66 years) yielded alpha coefficients substantially lower ($M = .55$) than those typically observed in adult samples (Markey, Markey, Tinsley, & Ericksen, 2002).

These findings suggest that important developmental changes in within-domain coherence occur at some point prior to late adolescence. However, pinpointing the precise timing and extent of age differences requires more detailed data than have been available

previously. Here, we examine the within-domain coherence of youths' personality self-reports at each year of age from 10 to 20. We hypothesize that, when averaged across the Big Five domains, coherence is better at older ages, with rapid gains in late childhood and early adolescence followed by smaller gains in late adolescence.

General Changes in Between-Domain Differentiation

Between-domain differentiation is the extent to which scales that assess conceptually distinct traits have low correlations with each other; lower between-domain correlations indicate better differentiation. In general, the differentiation of youths' personality self-reports should increase with age. As research on self-representations and identity has shown (see above), adolescents can more readily distinguish among different social roles than can younger children. This increased differentiation of self-aspects should be accompanied by a similar increase in the capacity to rate distinct personality traits independently of each other.

In the Estonian sample of secondary school students mentioned above, the mean intercorrelation among the five NEO-FFI scales was .24 among 12-year-olds, compared with .12 among 18-year-olds (Allik et al., 2004). However, the extent and timing of age differences in between-domain differentiation remain to be investigated more precisely. Here, we examine between-domain differentiation at each year of age from 10 to 20, and we test the hypothesis that—averaged across the Big Five domains—differentiation is substantially better at older ages, with particularly large gains during late childhood and early adolescence.

Unique Developmental Trends in Coherence and Differentiation for Specific Big Five Domains

While age gains in within-domain coherence and between-domain differentiation should hold in the aggregate, there are also reasons to expect a more complex pattern at the level of the individual domains. That is, some of the Big Five should show developmental trends in coherence and differentiation that are quite different from the overall trends. One reason for this expectation concerns previous findings suggesting that the basic organization of personality traits changes with age. If the broad domain of Extraversion is fundamentally less cohesive in childhood and early adolescence than in adulthood, or even consists of distinct components, then measures of Extraversion should show especially dramatic age gains in coherence.

A second reason concerns the focus of early socialization processes. Parents and teachers convey to young children that there are appropriate (good) and inappropriate (bad) ways to behave. As articulated by Digman (1997), in Big Five terms, being a good child involves acting both agreeably (e.g., "don't hit," "share with your brother/sister") and conscientiously (e.g., "do as you're told," "pick up after yourself"). By contrast, being a bad child involves making trouble for parents and teachers, as captured by the externalizing construct in child psychopathology (Achenbach, 1966, 1978, 1979), which has been linked empirically to low levels of both Agreeableness and Conscientiousness (John et al., 1994; Robins, John, & Caspi, 1994). Given these elaborated notions of what constitutes good and bad behavior, children should have relatively coherent concepts of Agreeableness and Conscientiousness but may have trouble distinguishing between these two ways of being good. Put another way, Agreeableness and Conscientiousness should show high within-domain coherence from a young age but should lack clear differentiation until later on.

A third reason to expect domain differences in age trends concerns item comprehension. Even for adults, questionnaire items differ in how easy or difficult they are to comprehend, and less comprehensible items tend to show less desirable psychometric characteristics (e.g., Angleitner, John, & Löhr, 1986). Items written to assess some personality domains may be particularly difficult for young respondents. For example, the Openness scale of the NEO-FFI includes several items that seem difficult for children to comprehend and relate to their own lives, such as an item asking them whether they are tolerant of other lifestyles and another asking whether students should be allowed to listen to controversial speakers. Perhaps because of this, even in a sample of academically gifted early adolescents (M = 12.0 years) who completed the NEO-FFI, fully one-third of the 12 Openness items did not load substantially on the principal component representing Openness (Parker & Stumpf, 1998). Even for personality questionnaires that are generally easier to understand than is the NEO-FFI, measures of Openness may show particularly large age gains in coherence.

The issue of whether self-reports for different personality domains show different age trends in within-domain coherence and between-domain differentiation remains to be examined systematically. Are young respondents better able to report about some personality characteristics than about others? Here we test three novel hypotheses: (a) that Extraversion will show particularly large age gains in coherence; (b) that Agreeableness and Conscientiousness will show large gains in differentiation—especially in differentiation from each other; and (c) that Openness will show especially large gains in coherence.

Changes in Factor Structure

General age trends in coherence and differentiation, as well as specific trends for specific personality domains, should combine to influence the factor structure of multidimensional personality measures. Because all of these trends affect overall factor structure simultaneously, it is difficult to anticipate how any specific trend will be manifested in the overall structure. In general, however, the intended factor structures of personality measures should be recovered less clearly in the responses of children and adolescents than in those of adults.

Some previous studies of adolescents' self-reports have recovered versions of the Big Five structure in samples as young as age 12 (Allik et al., 2004; McCrae et al., 2002; Parker & Stumpf, 1998). However, these studies all used a single instrument, the NEO-FFI, and the adolescent structures were not identical or nearly identical to this measure's adult structure, even when target rotation was used to maximize congruence (Allik et al., 2004). For example, in a sample of academically gifted early adolescents, 9 of the 60 NEO-FFI items (15%), did not have substantial loadings on their intended primary component (Parker & Stumpf, 1998).

We do not yet know whether the findings of these studies using the NEO-FFI can be generalized to other measures. Moreover, the structure of responses from individuals younger than age 12 remains to be examined separately from older age groups, raising the

question of whether the Big Five structure can be recovered at ages younger than this. Here we examine data from a diverse sample that was large enough to conduct factor analyses separately for each specific year of age from 10 to 20. We use a Big Five measure written approximately at the reading level of a fifth grader or 11-year-old (Benet-Martínez & John, 1998), namely the Big Five Inventory (BFI; John, Donahue, & Kentle, 1991; John, Naumann, & Soto, in press; John & Srivastava, 1999). We therefore hypothesize that the intended five-factor structure of the BFI is recognizable by age 11.

Changes in Variability of Acquiescent Responding

Acquiescent response style refers to the tendency of an individual to consistently agree (yea-saying) or consistently disagree (nay-saying) with questionnaire items, regardless of their content (Jackson & Messick, 1958). When individuals within a sample differ substantially in their acquiescence tendencies (i.e., some tend to agree more and others tend to agree less), this variation introduces a positive covariance for every item pair, regardless of item content and scoring direction, and thus tends to positively bias interitem correlations. Correlations that would be positive in the absence of acquiescence variation are made even more positive, and correlations that would be negative are made less negative.

This positive biasing of interitem correlations has undesirable effects on personality self-reports. At the item level, if acquiescence variation is not removed prior to factor analysis, individual differences in acquiescent responding often result in the emergence of a factor representing the within-person mean item response (McCrae, Herbst, & Costa, 2001; Ten Berge, 1999). At the scale level, if a scale does not have balanced keying (i.e., an equal number of true-keyed and false-keyed items), its internal consistency will be artificially inflated by the positively biased interitem correlations, but its validity will be decreased because measurement of the desired construct will be confounded with acquiescence.

Moreover, individual differences in acquiescent responding influence the mean scores on a scale with an unequal number of true- and false-keyed items. A respondent who tends to agree with items, regardless of their content, will score too high on a scale with more true-keyed items than false-keyed items and too low on a scale with more false-keyed items. Because of this, acquiescence variability will also bias the correlation between two scales that do not have balanced item keying. This correlation will be positively biased if both scales have mostly true-keyed or mostly false-keyed items, or negatively biased if one scale is mostly true-keyed and the other is mostly false-keyed.

Variability in acquiescent responding should be greater at younger ages, for several reasons. As reviewed above, older adolescents have clearer self-concepts, as well as larger vocabularies and better verbal comprehension skills, than do younger adolescents and children. Older adolescents are also likely to have more experience with both informal questionnaires (e.g., personality and attitude quizzes in teen magazines) and formal questionnaires (e.g., teaching evaluations), and thus a better understanding of implicit norms about how to respond to questionnaire items (e.g., that responses should be centered around the midpoint of a multiple-step rating scale). For all of these reasons, younger respondents should be less certain about how to interpret and respond to any particular questionnaire item, leaving greater room for the influence of systematic response styles (Goldberg, 1963), including individual differences in acquiescence. Changes in critical reasoning capacities during early adolescence, and thus the presence of substantial individual differences in these capacities, may also contribute to especially great variability in acquiescent response style during this period (cf. Knowles & Condon, 1999).

Despite these reasons to expect greater individual differences in acquiescent responding at younger ages, the influence of acquiescence on youths' personality self-reports remains to be examined. Here, we test the novel hypotheses that (a) younger respondents show greater individual differences in acquiescent responding; and (b) individual differences in acquiescence have greater effects on the coherence, differentiation, and overall structure of personality self-reports at younger ages.[1]

The Role of Verbal Comprehension

Because understanding a personality questionnaire item is a necessary precondition to providing a valid response, age differences in comprehension might explain why youths' personality self-reports differ from those of adults. One relevant finding is that Estonian adolescents with higher nonverbal intelligence scores provided more internally consistent responses to the NEO-FFI than did adolescents with lower scores (Allik et al., 2004). Moreover, as noted above, in adult samples questionnaire items that are difficult to comprehend tend to show less desirable measurement characteristics than do easier items (Angleitner et al., 1986). However, the link between verbal comprehension and the coherence of youths' personality self-reports remains to be tested. Are age and domain differences in the coherence of youths' personality self-reports simply due to the ongoing development of verbal comprehension skills?

The approach taken here focuses on comparing BFI scales and items that are particularly difficult for youths to comprehend with those that are particularly easy to comprehend. If developmental trends in the coherence of youths' personality self-reports are due to verbal comprehension, then particularly difficult scales and items should show much less coherence at young ages than do easier scales and items, and the relation between item difficulty and coherence should be stronger in childhood than in adulthood. Moreover, difficult scales and items should show especially large age gains in coherence, whereas easy scales and items should show smaller age differences.[2]

---

[1] We did not have clear expectations regarding whether younger respondents would show more or less yea-saying, on average, than would older respondents. Exploratory analyses did not show an effect of age on the mean level of acquiescence; analyzed with a 2 (gender) $\times$ 11 (age) analysis of variance, the effect of age on mean acquiescence was not statistically significant, $F(10, 230025) = 3.59$, *ns.* However, even if there had been mean-level age differences in acquiescence, they would not have affected the patterns of interitem and interscale correlations examined here. They would have affected only cross-age comparisons of scale means.

[2] It is less clear how item difficulty should relate to between-domain differentiation. On the one hand, younger respondents may have less clear understandings of the personality characteristics they are asked to rate, hurting differentiation. On the other hand, they may respond to a difficult item less reliably, thus "helping" differentiation (by reducing the correlation of that item with all other scales and items). Exploratory analyses indicated that item-comprehension difficulty was not significantly related to differentiation in either direction.

Overview of the Present Research

The present research examines responses from a large and diverse sample of children (ages 10–11), adolescents (ages 12–17), and young adults (ages 18–20) who completed the BFI, a well-validated and widely used measure of the Big Five personality trait domains, on the Internet. With these data, we first examine age trends in the variability of acquiescent responding as well as the effects of this variability on the characteristics of youths' personality self-reports. We then trace the within-domain coherence, between-domain differentiation, and overall factor structure of personality self-reports from ages 10 to 20—and test whether age trends in coherence and differentiation differ meaningfully by personality domain. Finally, we investigate the role of verbal comprehension in explaining these patterns.

Method

Sample

Participants were 230,047 volunteers who provided personality and demographics data over the World Wide Web. All respondents were between the ages of 10 and 20 (M = 16.6 years) and were residents of either the United States (90% of the sample) or Canada (10%). In preliminary data screening, a small proportion (2.5%) of cases was eliminated from the final sample because multiple questions were left unanswered or because the same response was repeated over and over (cf. Costa & McCrae, 1992; Gosling, Vazire, Srivastava, & John, 2004).[3] This exclusion rate is similar to what we have observed with undergraduate samples (ages 18 to 23 years) recruited from a research participant pool; in such samples, exclusion rates have typically varied between 2% and 5%. Even after the elimination of potentially suspect cases, sample sizes for each specific year of age were large: 1,754 for age 10; 3,590 for age 11; 6,508 for age 12; 14,310 for age 13; 21,152 for age 14; 27,138 for age 15; 31,272 for age 16; 32,483 for age 17; 36,397 for age 18; 29,317 for age 19; and 26,126 for age 20.[4]

Compared with non-Internet samples recruited for personality and social psychology studies published in the *Journal of Personality and Social Psychology* (Gosling et al., 2004), the present sample was split more evenly between male and female participants and was at least as diverse in terms of ethnicity. Overall, 60% of respondents were female and 40% were male. Respondents reported their ethnicity as one of seven categories: 12,540 (5.5%) were Asian; 7,684 (3.3%) were Black; 9,895 (4.3%) were Latino; 972 (0.4%) were Middle Eastern; 1,468 (0.6%) were Native American; 178,761 (77.7%) were White; 11,865 (5.2%) reported their ethnicity as Other; and 6,862 (3.0%) did not report their ethnicity.

The sample was also diverse in terms of socioeconomic status. A question about social class was added during the survey period, and answers to this question were available for most of the sample. Of these respondents, 11,866 (9.2%) described their parents' social class as working class; 13,586 (10.5%) as lower-middle class; 55,758 (43.2%) as middle class; 43,198 (33.4%) as upper-middle class; and 4,729 (3.7%) as upper class.

Because of the large sample size, all significance tests of individual-level (as opposed to group-level) variables were evaluated at the conservative $\alpha = .001$ level. Tests of group-level variables were evaluated at the conventional $\alpha = .05$ level.

Procedure and Data Collection

The data analyzed here were collected by Jeff Potter via his Web site (www.outofservice.com) and made available for scientific analysis in exchange for authorship. The outofservice.com Web site offers its visitors free feedback on several surveys and personality measures. Potential respondents could reach this Web site in a number of ways, including search engines, unsolicited links from other Web sites, and informal channels, such as e-mail and online discussion forums.

Many children and adolescents may have difficulty appreciating the risks and costs of participating in a research procedure, so special care must be taken to protect the welfare of minor participants in scientific research. For most research involving minors, this includes obtaining the informed consent of parents for their child's participation. Initially, Jeff Potter—a Web entrepreneur who is not a psychologist or associated with an academic institution—obtained a portion of the present data without seeking review from an institutional review board (IRB). When the potential significance of this emerging data set became apparent, Oliver P. John and Samuel D. Gosling obtained separate IRB approvals to (a) analyze the then-existing data archive and (b) collect new data. The present research, including a waiver of parental consent, was approved by two IRBs (a) because the research procedure—anonymously answering questions about global personality characteristics and receiving general and innocuous feedback—involves no more than minimal risk to the participants, (b) because the waiver of consent would not adversely affect the rights or welfare of the participants, and (c) because the research could not practicably be carried out without the waiver (United States Department of Health and Human Services, 2005). More generally, we encourage researchers interested in collecting questionnaire data from minors, whether over the Internet or by other means, to carefully consider the welfare of their participants when selecting the questions to be asked and any feedback to be given.

Measurement

The BFI. The BFI (John et al., 1991; John et al., in press; John & Srivastava, 1999) was used to obtain personality self-reports. The 44 BFI items are short, easy-to-understand phrases that assess personality traits central to each of the Big Five domains. An

[3] Exclusion rates varied somewhat by age. For ages 10 to 13, 2.7% of all cases were excluded because of missing responses, compared with 1.3% of cases for ages of 14 to 16 and 0.8% of cases for ages of 17 to 20. Similarly, the proportion of cases excluded for repeatedly selecting the same response option was 3.1% for ages 10 to 13, 1.5% for ages 14 to 16, and 0.9% for ages 17 to 20. Thus, aberrant responding was somewhat more common among younger respondents, but still quite unusual.

[4] Data from some respondents younger than age 10 were also available. However, the fifth-grade reading level of the BFI suggests that this inventory may be too difficult for respondents much younger than age 11. Indeed, sample sizes decreased rapidly below age 10: Sample size dropped by more than half from age 10 to age 9, and again from age 9 to age 8. Younger samples were thus too small to yield reliable PCA results at the item level. Moreover, for participants as young as age 10, self-selection concerns are ameliorated by comparisons with findings from other studies, but previous studies examining age trends in youths' responses to a Big Five questionnaire have not included any participants younger than age 10.

example item from the Extraversion scale is "Is outgoing, sociable"; a reverse-keyed example item from the Neuroticism scale is "Is emotionally stable, not easily upset." Each item was rated on a scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The brevity of the BFI, and its fifth-grade reading level (Benet-Martínez & John, 1998), make it well suited to a survey of young respondents that expects each participant to devote only a limited amount of time. In adult samples, the BFI scales have shown high internal consistency, retest reliability, and clear factor structure, as well as strong convergence with longer Big Five measures (e.g., Benet-Martínez & John, 1998; John et al., in press; John & Srivastava, 1999) and substantial agreement between self- and peer-reports (e.g., DeYoung, 2006; Rammstedt & John, 2007).

Measuring individual differences in acquiescent response style. Individual differences in acquiescent responding were indexed with acquiescence scores (within-person response means) computed from a set of 16 pairs of BFI items with opposite implications for personality (e.g., "Is talkative" vs. "Tends to be quiet"; see the Appendix for a complete list of these pairs as well as scoring procedures). These pairs were selected by Christopher J. Soto and Oliver P. John on the basis of item content and interitem correlations. An acquiescence score was computed for each participant across the 32 items from these pairs (rather than from the full set of 44 BFI items) because, in the full set, direction of item keying is confounded with personality content. Specifically, all of the BFI scales have more true-keyed items than false-keyed items, although the ratio of true- to false-keyed items varies across scales. This confound is not present in the set of 16 opposite item pairs, which equates, for each Big Five domain, the number of true-keyed and false-keyed items.

Ipsatized data: controlling for response style. Ipsatization (or within-person standardization) is the procedure of transforming data so that each participant's set of item responses has the same mean and standard deviation (e.g., McCrae et al., 2001; Ten Berge, 1999). Many researchers routinely ipsatize their data to control for individual differences in acquiescent and extreme responding (e.g., Ashton et al., 2004; Goldberg, 1990, 1992). Typically, ipsatization is accomplished by subtracting each participant's mean response (across all items) from each of their individual item responses, then dividing these differences by the standard deviation of that person's responses. The resulting set of transformed responses has a mean of 0 and a standard deviation of 1 for each participant.[5] To avoid confounding response style with personality content (as described in the previous paragraph), within-person response means and standard deviations here were computed across the 32 items from the set of 16 opposite-item pairs; these values were then used to ipsatize the full set of 44 items, as detailed in the Appendix.

Assessing the comprehension difficulty of the BFI items: item understandability and trait-word familiarity. Two approaches were used to assess comprehension difficulties posed by the BFI items. First, following Angleitner et al. (1986), 13 psychology graduate students reviewed the BFI items and then judged how difficult each item would be for a typical 10-year-old to understand. As in Angleitner et al. (1986), psychology students were used as judges because of their familiarity with personality measures and their general psychological mindedness. These judges were asked to consider both the semantic and syntactic features of the items; an easy-to-understand item was defined as one that is clearly written and uses only basic vocabulary (i.e., words that almost all fourth graders/10-year-olds would know), whereas a difficult item was defined as one that uses more advanced vocabulary or syntax. Each judge rated each BFI item on a scale ranging from 1 (*very easy to understand*) to 7 (*very difficult to understand*). Across the set of 44 items, the alpha reliability for the mean of the 13 judges was .93.

A second approach to assessing comprehension difficulty focused exclusively on a semantic aspect of the BFI items, namely the unfamiliarity of the trait concepts being assessed: Items that include unfamiliar trait terms are presumably more difficult for respondents to comprehend. Demonstrating that the familiarity of trait terms can be assessed reliably, Saucier (1997) collected familiarity ratings for 1,248 person-descriptive adjectives from a sample of 112 adult community residents and undergraduate psychology students, using a scale ranging from 0 (*I don't know the meaning of the word*) to 9 (*I know the meaning of the word and think it is used extremely often to describe people*). The alpha reliability of these judgments exceeded .90.

The BFI item phrases were derived from trait adjectives that experts had rated as uniquely prototypical of each Big Five domain (see John, 1990), and so we were able to index the unfamiliarity of each item by matching its personality descriptors to Saucier's (1997) list of ratings. We reverse-keyed Saucier's original ratings so that high scores would indicate unfamiliarity—that is, items that were more unfamiliar and thus more difficult to comprehend. For items that included more than one descriptor (e.g., "Is outgoing, sociable"), we indexed overall unfamiliarity as the mean of all adjectives for which ratings from Saucier (1997) were available. For four BFI items (e.g., "Tends to find fault with others"), the corresponding trait adjectives (e.g., faultfinding) were not included on the Saucier (1997) list, and so these four items could not be used in the present analyses of unfamiliarity.

Examples of BFI items that were identified by both understandability and unfamiliarity ratings as particularly difficult are "Is ingenious, a deep thinker" (difficulty of understanding = 5.15, unfamiliarity = 3.72); "Can be cold and aloof" (difficulty of understanding = 5.15, unfamiliarity = 2.58); and "Likes to reflect, play with ideas (difficulty of understanding = 5.08, unfamiliarity = 3.76). These items clearly include more difficult language than do three items identified as particularly easy: "Is talkative" (difficulty of understanding = 1.23, unfamiliarity = 1.48); "Is sometimes rude to others" (difficulty of understanding = 1.62, unfamiliarity = 0.76); and "Tends to be quiet" (difficulty of understanding = 1.77, unfamiliarity = 0.92).

How should these two approaches to defining comprehension difficulty be related? Note that the difficulty-of-understanding ratings took into account both the syntax and semantics of the full item text, whereas the unfamiliarity scores were based only on the individual trait terms. Moreover, the difficulty-of-understanding judges focused on the language skills of typical 10-year-olds,

[5] Two other methods of controlling for individual differences in acquiescent responding are (a) within-person centering and (b) partialling the mean response. In the present research, the three methods produced virtually identical results; results from analyses of the ipsatized data are presented here because this is the approach most commonly used in personality research.

whereas the adults and undergraduates who judged familiarity in Saucier's (1997) research were asked to consider only their own knowledge of the personality terms. Despite these differences, difficulty of understanding and unfamiliarity were strongly intercorrelated across the 40 BFI items for which the unfamiliarity index was available ($r = .48$, $p < .05$).

We also tested whether these two sets of ratings were sufficiently sensitive to children's verbal abilities. Three elementary school teachers particularly familiar with the vocabulary and reading skills of 10- and 11-year-old children rated how difficult it would be for a typical 10-year-old (fourth grader) to understand each of the BFI items. Across the 44 items, the means of the teacher ratings correlated very strongly with the means of the graduate student ratings ($r = .82$, $p < .05$). These teacher ratings also correlated substantially with the unfamiliarity scores derived from Saucier (1997; $r = .39$, $p < .05$).[6]

Gender Differences

In preliminary analyses, we examined differences between the characteristics of responses made by male and female participants. At each age, within-domain coherence and between-domain differentiation for males and females differed by no more than two hundredths of a correlation point. More importantly, age trends in coherence and differentiation were not substantively different for male and female participants. Therefore, specific tests of gender differences are not reported here. Instead, the confounding of age effects with gender effects was avoided by analyzing male and female participants separately. In the interest of clarity, all figures plot age-group values as the unweighted average of the male and female values.

Results and Discussion

Individual Differences in Acquiescent Responding

We hypothesized that younger respondents would show greater individual differences in their acquiescent response tendencies than would older respondents. There was considerable variation in respondents' acquiescence scores overall, and the extent of this variation differed systematically by age. Figure 1 shows the variance of the acquiescence scores at each age; higher values indicate greater individual differences in acquiescent responding. As expected, variance in acquiescence scores was substantially larger in younger age groups; for Levene's test for equality of variance, $F(21, 230025) = 118.15$, $p < .001$. In fact, as Figure 1 shows, there was fully twice as much variability in acquiescence at age 10 (variance = 0.14) as at age 20 (variance = 0.07).

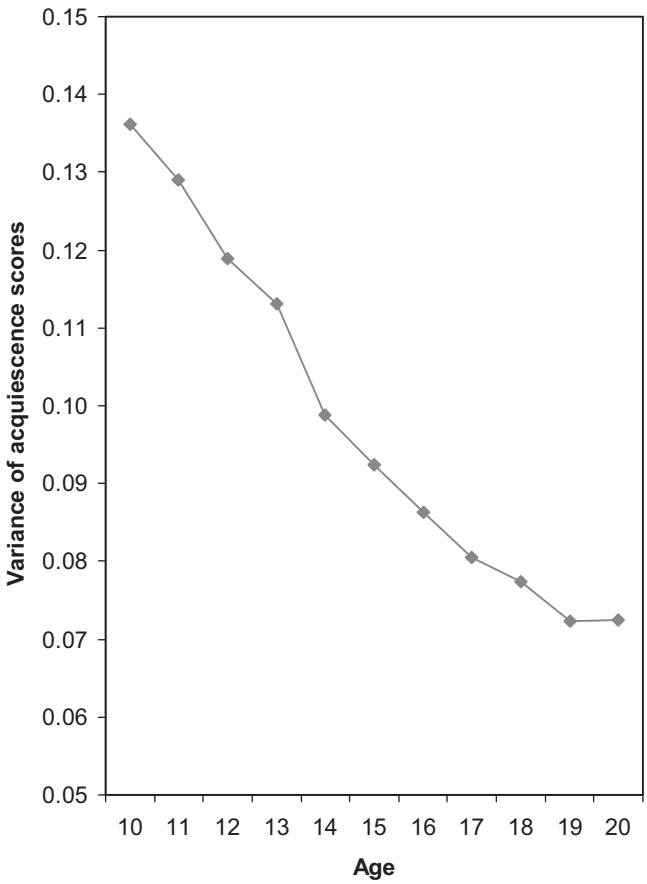


Figure 1. Variance of the acquiescence scores (i.e., extent of individual differences in acquiescent responding) by age group.

Principal Component Structure of the BFI Across Ages

The influence of acquiescence variation on eigenvalues. As a first test for the effects of individual differences in acquiescent responding on the characteristics of youths' personality self-reports, we examined eigenvalues from principal components analyses (PCAs) of the raw and ipsatized data conducted separately for each age group from 10 to 20. Figure 2a presents scree plots for analyses of the raw data (i.e., acquiescence uncontrolled) at ages 20, 18, 16, 14, 12, and 10; plots for the odd age groups are omitted because they were very similar to those of adjacent, even ages.

The age-20 plot (i.e., the top panel of Figure 2a) provides an adult comparison standard.[7] This plot indicates a major break after six components, with five large components presumably representing the Big Five, plus a smaller one presumably representing individual differences in acquiescence. The first six eigenvalues are therefore plotted as filled circles, whereas the remaining eig-

[6] Another approach would be to use the structural characteristics of the BFI items (e.g., average number of words per sentence, average number of syllables per word) to compute an estimate of comprehension difficulty (e.g., Flesch–Kincaid grade level; see Benet-Martínez & John, 1998). However, such indices assume reliable structural estimates (i.e., averaged across many sentences) and so are not well suited to estimating the difficulty of individual sentences or, worse, the short phrases that constitute the BFI items. Thus, not surprisingly, when we computed Flesch–Kincaid reading grade levels for the individual BFI items, these scores were related only modestly to difficulty of understanding as rated by psychology graduate students ($r = .24$) and elementary school teachers ($r = .14$) and were not related at all to adjective familiarity ($r = .00$).

[7] To be sure that there were no substantial changes in the structure of the BFI after age 20, we analyzed data from a comparable sample of 30-year-olds (Srivastava et al., 2003). The age-20 and age-30 samples were virtually identical in terms of all measurement characteristics examined in the present research.

envalues, representing the "scree" (rubbish variance), are plotted as open circles.

If the sixth non-scree component indeed represents acquiescent responding, then this component should have two characteristics. First, most items should have a small, positive loading on this component, regardless of their content (e.g., both "Is talkative" and "Tends to be quiet" should load positively). Indeed, when six components were extracted and varimax-rotated at age 20, 38 of the 44 BFI items (86%) had positive loadings on the sixth component.

Second, controlling for individual differences in acquiescence should effectively remove the additional component. Scree plots from PCAs of the ipsatized BFI responses (i.e., acquiescence controlled) are shown in Figure 2b. At age 20 (shown in the top panel of Figure 2b), there was a substantial break after the first five eigenvalues (plotted as filled circles); unlike in analyses of the raw data (Figure 2a), there was no clear break after the sixth component. That is, ipsatizing the data to control for individual differences in acquiescent responding effectively removed the extra component found in PCAs of the raw data.

As Figure 2 shows, this pattern of six non-scree components in the raw data and five non-scree components in the ipsatized data was generally consistent across age groups but was qualified by two important trends. One trend is that, as shown in Figure 2a, the separation of the five expected components from the sixth, extra component in the raw data was smallest—indeed, practically non-existent—at age 10, then increased systematically to age 20. This effect can be quantified by using the difference between the fifth and sixth eigenvalues, which was only 0.12 at age 10 but increased to 1.30 by age 20; across the full age range, this difference correlated .98 with age group. This trend is consistent with the hypothesis that individual differences in acquiescent responding have larger structural effects at younger ages.

A second developmental trend is that the five expected components accounted for a much larger proportion of the total variance in the older age groups. At age 10, the first five components in the ipsatized data (see Figure 2b) accounted for 34% of the total variance, which increased to 45% by age 20—a gain of approximately one-third. Across the full age range, the correlation between the variance accounted for by the first five components and age group was .96. This trend suggests that, even after controlling for individual differences in acquiescence, the BFI's intended five-factor structure may be recovered more clearly at older ages.

Five-component solutions. As noted above, at age 10 the scree plot for the raw data suggested six non-scree components (see Figure 2a). By contrast, once acquiescence was controlled through ipsatization, the plot indicated only five non-scree components (see Figure 2b). The presence of an extra acquiescence component in the raw data raises the question of whether the familiar Big Five structure could be recovered at age 10 if individual differences in acquiescent responding were not first controlled. To address this question, and to further examine the effects of variation in acquiescent responding on the overall structure of the BFI, we varimax-rotated five components in the raw and in the ipsatized data for each specific age group from 10 to 20.

At age 20, the intended Big Five structure was very clearly recovered in varimax-rotated five-component solutions, as in previous research using the BFI with college-aged samples (e.g., Benet-Martínez & John, 1998; John & Srivastava, 1999). In both the raw and ipsatized data, each of the 44 BFI items had its largest loading on its intended component, and these loadings were substantial in size, with absolute values averaging .63.

This five-dimensional structure was also clearly recovered throughout middle and late adolescence. However, it was recovered much less clearly in early adolescence and was not recognizable at age 10 when individual differences in acquiescence were not controlled. Consider first the congruence coefficients presented in Table 1, which compare the five components at each age with the corresponding components from our age-20 standard. In general, high congruence coefficients indicate that the two components being compared have similar patterns of loadings across the 44 BFI items, with congruence of 1 indicating perfect similarity of relative loadings. The left half of Table 1 presents congruence coefficients for the raw data. Going backwards from age 19, congruence with the age-20 solution declined gradually from early adulthood through middle adolescence. However, even as far back as age 13, all five coefficients were at least .96, and inspection of the loading matrix indicated that at this age 43 of the 44 BFI items (98%) still had their strongest loading on the intended component.

Age differences in congruence for the raw data were much more pronounced below age 13. As shown in Table 1, three of the five congruence coefficients (for Extraversion, Conscientiousness, and Neuroticism) declined rapidly from age 13 to age 10. Similarly, the number of items that had their strongest loading on the intended component decreased to 42 items (95%) at age 12, to 38 (86%) at age 11, and to only 34 (77%) at age 10. Indeed, at age 10, the intended Conscientiousness and Neuroticism dimensions were not recovered. Instead, one of the five rotated components was defined by the five high-Conscientiousness items and three low-Neuroticism items, along with a low-Openness item and a low-Extraversion item; another component was defined by the four low-Conscientiousness items, the five high-Neuroticism items, and a low-Extraversion item.

As the congruence coefficients for the ipsatized data (in the right half of Table 1) show, controlling for individual differences in acquiescence greatly aided recovery of the intended Big Five structure. Going backwards in age, congruence declined very gradually from early adulthood into early adolescence; even at age 13, congruence was at least .97 for each component. As in the raw data, the five factors were recovered less clearly in late childhood than in adolescence and early adulthood, and year-to-year differences in congruence were largest during late childhood. However, unlike in the raw data, the Big Five structure was recognizable in the ipsatized data even at age 10; at this age, congruence coefficients were at least .93 for each of the five components, and 43 of the 44 items (98%) already had their strongest loading on the intended component.

The high congruence coefficients for the ipsatized data indicate that, after controlling for individual differences in acquiescence, the patterns of relative loadings in varimax-rotated PCA solutions were quite similar across late childhood, adolescence, and early adulthood. Nonetheless, the item loading matrices indicated considerable room for improvement with age in terms of absolute loading sizes. Even in the ipsatized data, the magnitude of the 44 intended primary loadings

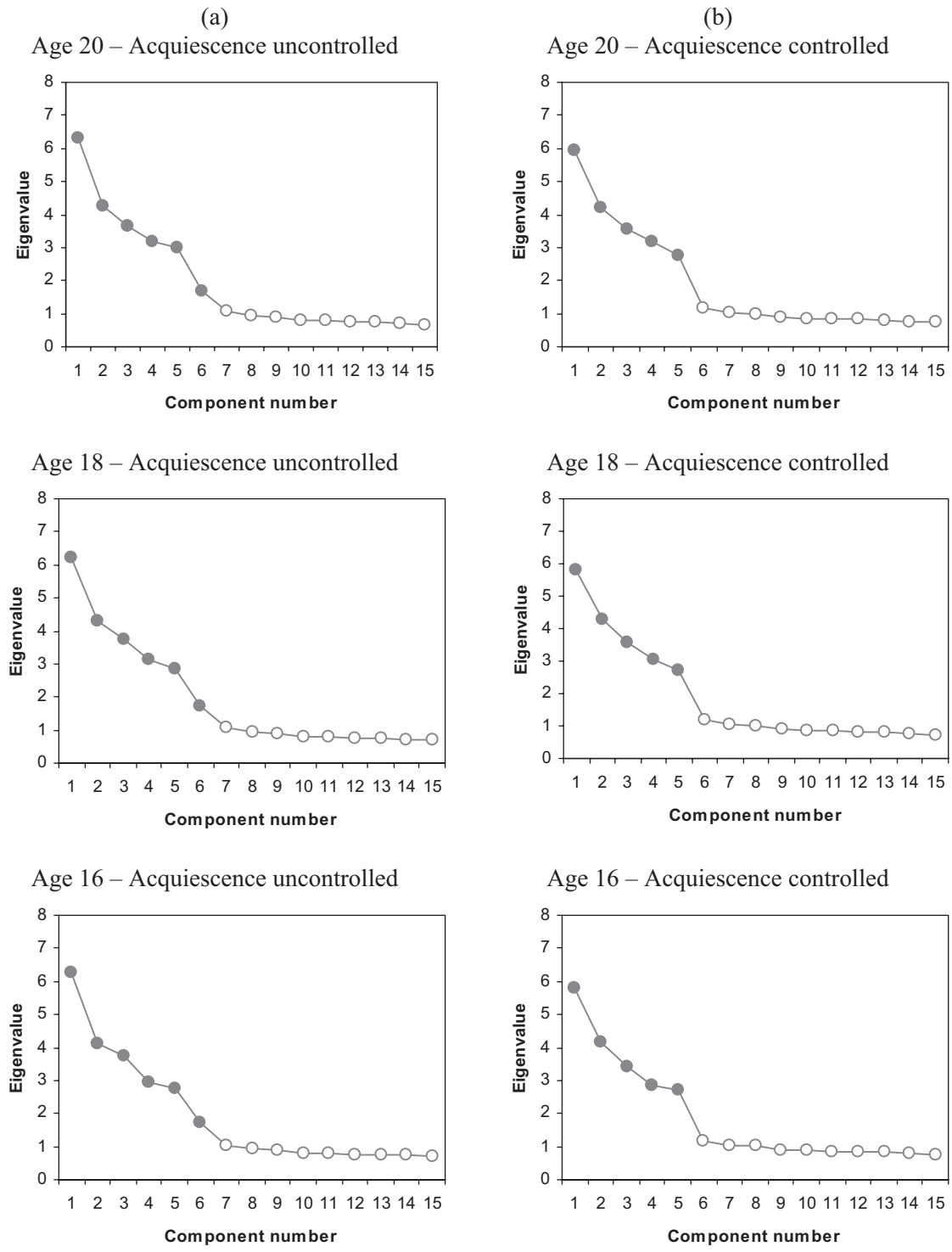
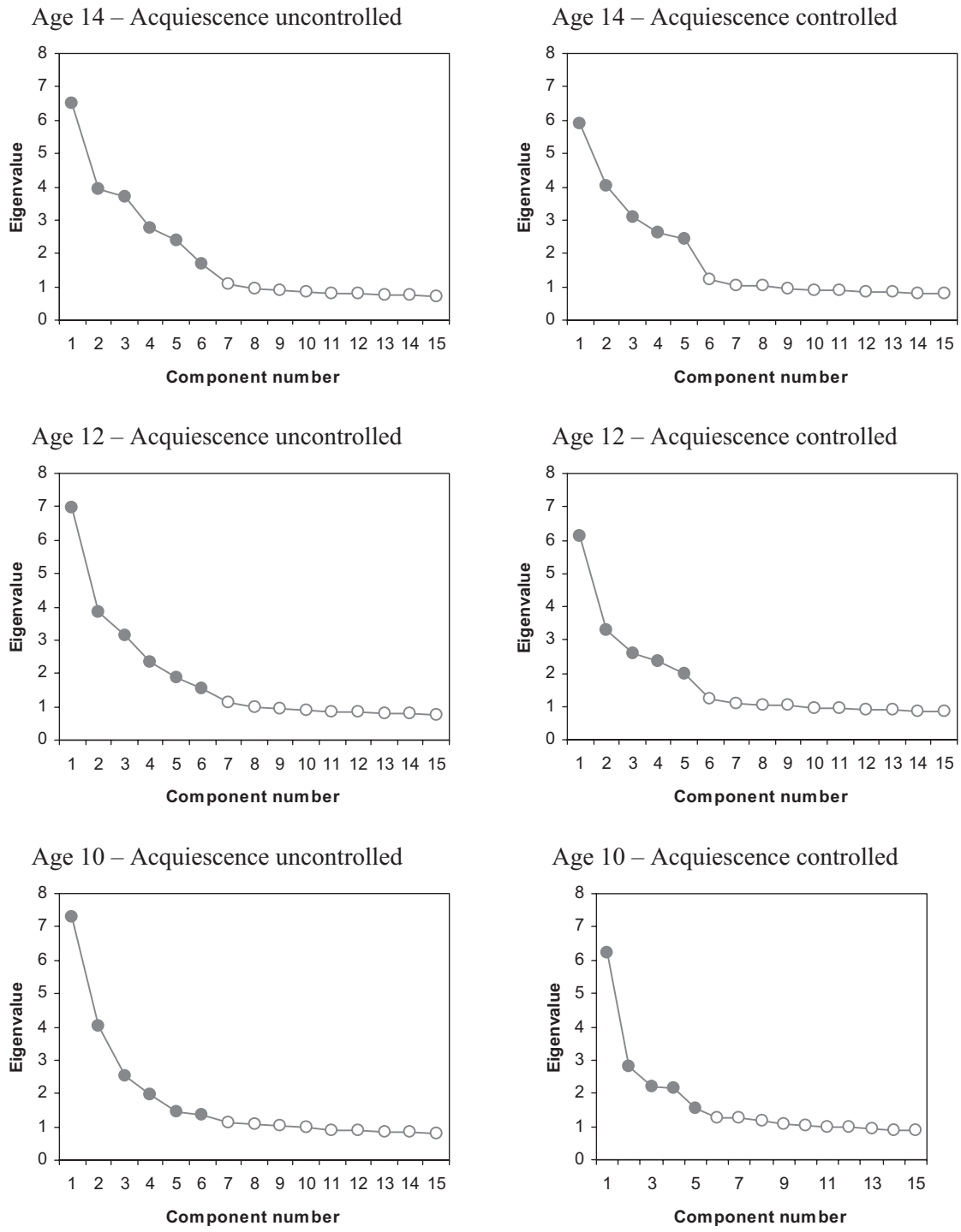


Figure 2. Scree plots for principal components analyses of the (a) raw data (acquiescence uncontrolled) and (b) ipsatized data (acquiescence controlled) by age group. In Panel a, the circles representing the first six eigenvalues are filled. In Panel b, the circles representing the first five eigenvalues are filled.

increased from a mean of .52 at age 10 to a mean of .63 at age 20; in variance terms, this means that a typical BFI item shared only 27% of its variance with its intended component at age 10 compared with 40% at age 20—a gain of almost half.

Taken together, these PCA results attest to the substantial influence of variation in acquiescent responding on youths' personality self-reports, particularly during late childhood. They also indicate that, once individual differences in response



style were controlled, the overall Big Five structure of the BFI was recognizable even at age 10. However, substantial age differences in the patterns of eigenvalues and the absolute sizes of component loadings point to meaningful developmental trends beyond those due to variability in acquiescent responding. Therefore, we now turn from overall structure to indices of more specific measurement characteristics, namely within-domain coherence and between-domain differentiation.

Within-Domain Coherence

Coherence within each Big Five domain at each age was indexed by the mean interitem correlation, an internal consistency statistic that—unlike Cronbach's alpha coefficient—is unaffected by differences in scale length, allowing for comparisons across the BFI scales.

Influence of variation in acquiescent responding. Because the BFI scales are not fully balanced in terms of item keying, we

Table 1
Congruence Coefficients Comparing Varimax-Rotated Five-Component Solutions at Each Age With the Age-20 Solution

Age (years)	Raw data (acquiescence uncontrolled)					Ipsatized data (acquiescence controlled)				
	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness
19	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
18	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
17	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
16	0.99	1.00	0.99	0.99	1.00	1.00	1.00	0.99	1.00	1.00
15	0.99	1.00	0.99	0.99	0.99	0.99	1.00	0.99	1.00	1.00
14	0.98	0.99	0.98	0.98	0.98	0.99	0.99	0.99	0.99	0.99
13	0.97	0.99	0.98	0.96	0.96	0.98	0.99	0.98	0.99	0.99
12	0.95	0.98	0.97	0.92	0.95	0.98	0.98	0.98	0.98	0.97
11	0.94	0.96	0.93	0.91	0.93	0.96	0.97	0.96	0.97	0.95
10	0.88	0.97	0.83	0.84	0.93	0.97	0.96	0.94	0.95	0.93

Note. Table entries are congruence coefficients comparing a component with the corresponding component at age 20.

expected that the presence of individual differences in acquiescent responding would modestly inflate within-domain coherence. Indeed, Figure 3a shows that overall coherence (averaged across gender and across the five BFI scales) was slightly higher in the raw data than in the ipsatized data, with differences of .01 or .02. At the level of individual scales (not shown in Figure 3a), differences ranged from –.01 to .04 and were largest for Openness—the least balanced BFI scale. Consistent with the hypothesis that individual differences in acquiescent responding would have stronger effects at younger ages, the overall degree of positive bias was larger at younger ages; the correlation of group age with the magnitude of the difference in coherence between the raw and ipsatized data was –.69.

General age trend. As expected, even after controlling for individual differences in response style, overall coherence was higher at older ages. In the ipsatized data (see Figure 3a), coherence increased from .23 at age 10 to .36 at age 20, meaning that a typical pair of items on the same BFI scale shared only 5% of their variance at age 10 but shared 13% at age 20—a gain of 160%. The correlation between group age and coherence was .97. Figure 3a also shows that age gains in coherence were generally large during late childhood and early adolescence, then much smaller at older ages.

Between-Domain Differentiation

Overall differentiation among the Big Five domains at each age was indexed by the mean of the 10 possible pairwise intercorrelations among the five BFI scales; lower values indicate that scores on each scale were better differentiated from scores on other scales. Because interscale correlations are attenuated by measurement error, which differed as a function of age (see Figure 3a), all pairwise interscale correlations were corrected for attenuation due to unreliability.[8]

Influence of variation in acquiescent responding. Because the BFI scales all have more true- than false-keyed items, the presence of individual differences in acquiescent responding should positively bias interscale correlations, hurting differentiation. Indeed, Figure 3b shows that overall differentiation (averaged across gender) was slightly worse in the raw data than in the ipsatized data, with differences ranging from .02 to .06. Supporting the hypothesis that individual differences in acquiescence have stronger effects at younger ages, this bias was larger at younger ages; the correlation of age with the magnitude of difference between differentiation in the raw and ipsatized data was –.98.

General age trend. As predicted, Figure 3b also shows that overall differentiation was better at older ages. In the ipsatized data, the mean interscale correlation decreased from .30 at age 10 to .18 at age 20, meaning that a typical pair of BFI scales shared 9% of their variance at age 10 but only 3% at age 20—a decrease of two-thirds. The correlation between group age and the mean interscale correlation was –.89. Gains in differentiation were generally large during late childhood and early adolescence, then smaller or even nonexistent during late adolescence.

Domain-Specific Age Trends in Coherence and Differentiation

Averaged across personality trait domains, within-domain coherence and between-domain differentiation were better at older ages. However, just as the main effects in a factorial analysis of variance cannot be interpreted without considering interaction effects, these overall age trends in coherence and differentiation need to be qualified by differences among the five personality domains—akin to age by domain interaction effects. Toward this end, Figure 4 presents coherence and differentiation, both computed in the ipsatized data at each age, separately for each of the Big Five domains.[9] As this figure shows, the domains differed substantially from each other in their age trends for both coherence and differentiation.

[8] Age trends were similar for raw correlations and disattenuated correlations. However, the presence of greater error variance at younger ages depressed raw correlations, thereby dampening the trend of better differentiation with age.

[9] The mean interitem correlations plotted in Figure 4a correspond to Cronbach's alpha coefficients for the Extraversion scale of .64 at age 10 and .87 at age 20; for Agreeableness, .76 at age 10 and .78 at age 20; for Conscientiousness, .76 at age 10 and .80 at age 20; for Neuroticism, .73 at age 10 and .82 at age 20; and for Openness, .57 at age 10 and .75 at age 20.

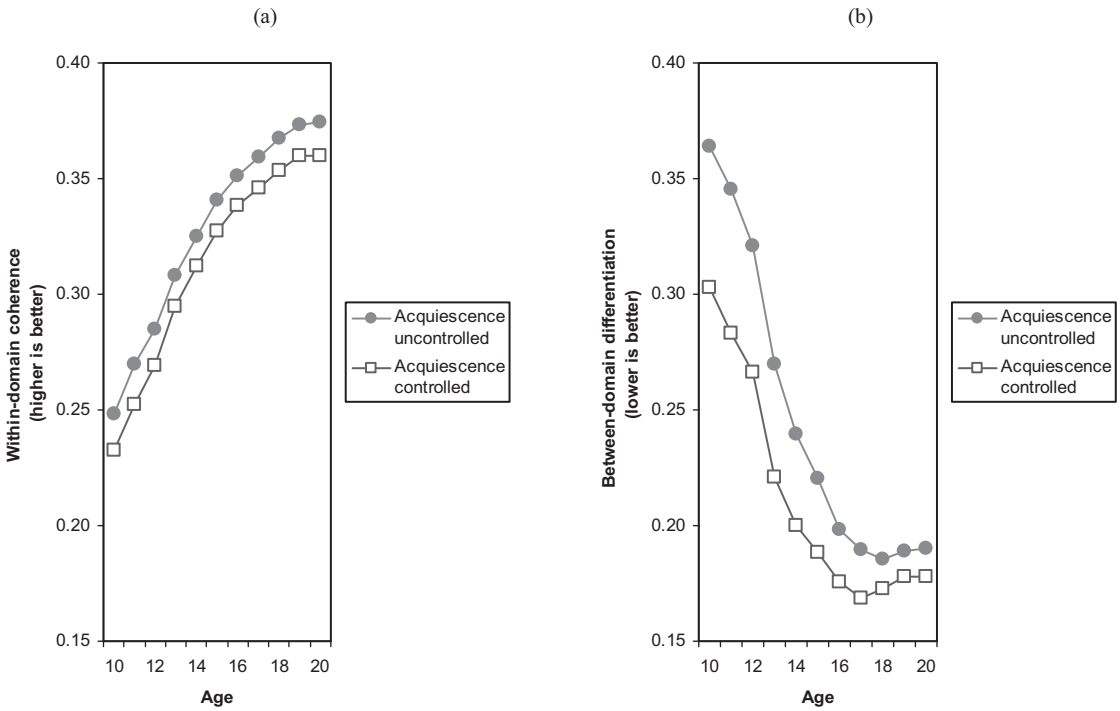


Figure 3. Within-domain coherence (i.e., mean within-domain interitem correlations) and between-domain differentiation (i.e., mean disattenuated interscale correlations) for the raw data (acquiescence uncontrolled) and ipsatized data (acquiescence controlled) by age group. All mean correlations are averaged across the five Big Five Inventory scales by using the Fisher *r*-to-*z* transformation.

Extraversion. As predicted, Extraversion showed the largest increase in coherence from age 10 to age 20, a difference of .28 for male participants and .26 for female participants ($ps < .001$).[10] However, this dramatic age gain was specific to coherence; Extraversion was already well-differentiated from the other BFI scales even at the youngest ages and did not show a significant gain in differentiation with age.

Agreeableness and Conscientiousness. As predicted, Agreeableness and Conscientiousness showed the largest age differences in between-domain differentiation. For Agreeableness (represented by filled triangles in Figure 4b), the difference between the differentiation index at age 10 and at age 20 was –.23 for male participants and –.17 for female participants ($ps < .001$); for Conscientiousness (empty triangles in Figure 4), this difference was –.19 for male participants and –.20 for female participants ($ps < .001$). In fact, of the 10 possible pairwise interscale correlations, that between Agreeableness and Conscientiousness showed the largest decrease from age 10 to age 20—a difference of –.34 for male participants and –.36 for female participants ($ps < .001$). By contrast, Agreeableness did not show a significant age gain in coherence, and Conscientiousness showed only a trivial one; for both male and female participants, the difference between age 10 and age 20 was .05 ($ps < .001$).

Openness and Neuroticism. We hypothesized that, on average, the items of the BFI Openness scale would be relatively difficult to comprehend and therefore that Openness would show particularly large age gains in coherence. Consistent with this expectation, Table 2 shows that Openness items were, on average, the most difficult to comprehend, as judged by both the unfamiliarity and difficulty-of-understanding indices. Planned contrasts comparing Openness with the other four scales were statistically significant for both difficulty of understanding, $F(1, 39) = 5.70$, $p < .05$; and unfamiliarity, $F(1, 35) = 3.93$, $p < .05$. However, as Table 2 also shows, the difficulty of the BFI items also varied considerably within each Big Five domain; for both indices, the ranges of all five scales overlapped with each other.

Regarding age trends in measurement properties, Openness did show substantial age differences in both coherence and differentiation, but the trends for this domain were quite similar to those for Neuroticism, the scale that was judged least difficult to comprehend by both indices. As shown in Figures 4c and 4d, age gains in coherence were moderate in magnitude both for Openness, which increased from age 10 to age 20 by .11 for male participants and .15 for female participants ($ps < .001$), and for Neuroticism,

[10] Here and elsewhere, the statistical significance of age differences in mean interitem correlations and mean interscale correlations was tested with a procedure developed by Feldt (1969) to test the null hypothesis that Cronbach's alpha coefficient is the same for two scales in two different groups. When the number of pairwise intercorrelations for the two scales is the same (as is the case when comparing the same BFI scale—or the correlations of one BFI scale with the other four scales—in two different age groups), this procedure is equivalent to testing the null hypothesis that the mean intercorrelation is the same for the two groups.

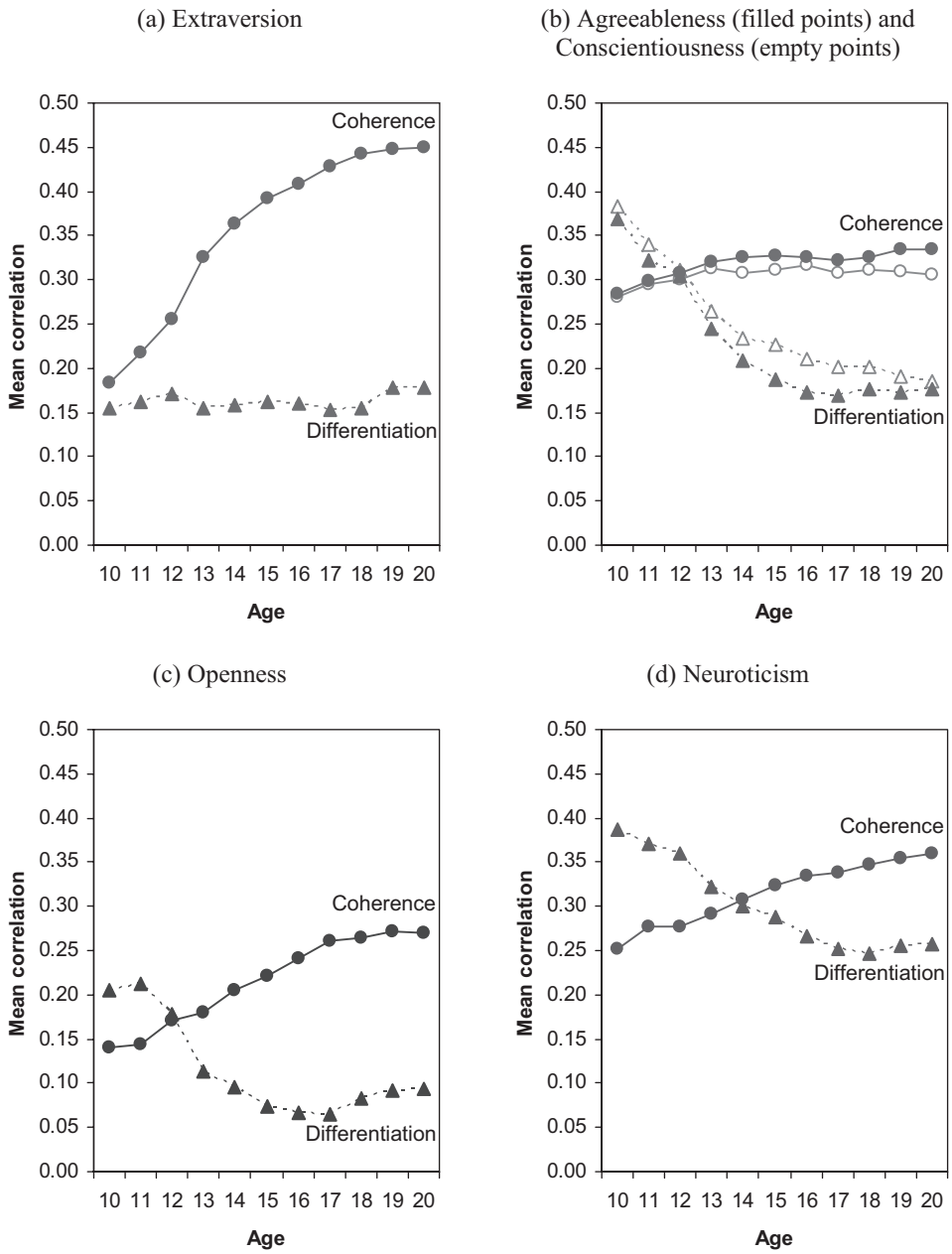


Figure 4. Within-domain coherence (i.e., mean within-domain interitem correlations) and between-domain differentiation (i.e., mean disattenuated interscale correlations) for each of the five Big Five Inventory scales by age group. In each panel, within-domain coherence is represented by the circles and solid lines, whereas between-domain differentiation is represented by the triangles and dotted lines. In Panel b, Agreeableness is represented by the filled points, whereas Conscientiousness is represented by the empty points. All values are computed from the ipsatized data to control for individual differences in acquiescence. All mean correlations are computed by using the Fisher *r*-to-*z* transformation.

which increased by .11 for male participants and .10 for female participants ($ps < .001$). Age trends in differentiation were also similar for these domains; Openness differed between ages 10 and 20 by –.16 for male and –.06 for female participants ($ps < .001$); Neuroticism differed by –.15 for male and –.12 for female participants ($ps < .001$).

Relation Between Item-Level Comprehension Difficulty and Coherence

The finding that the scales most difficult (Openness) and least difficult (Neuroticism) to comprehend showed very similar age trends does not support the idea that developmental trends in the

Table 2
Two Indices for Comprehension Difficulty of the Big Five Inventory Items: Means, Standard Deviations, Minimums, and Maximums by Big Five Domain

	Difficulty of understanding for 10-year-olds				Unfamiliarity of trait words for adults			
Big Five domain	*M*	*SD*	Minimum	Maximum	*M*	*SD*	Minimum	Maximum
Openness	3.92	1.31	2.00	5.38	2.65	0.84	1.88	3.80
Conscientiousness	3.13	0.96	1.85	4.92	1.78	0.58	1.08	2.58
Agreeableness	2.93	1.19	1.62	5.15	1.92	0.80	0.76	3.20
Extraversion	2.86	1.33	1.23	4.77	1.66	0.38	0.92	2.12
Neuroticism	2.79	0.93	1.46	3.62	1.53	0.52	1.04	2.60
All items	3.16	1.19	1.23	5.38	1.93	0.75	0.76	3.80

Note. Ratings of "Difficulty of understanding for 10-year-olds" ranged from 1 (*very easy to understand*) to 7 (*very difficult to understand*), with a midpoint of 4. Ratings of "Unfamiliarity of trait words for adults" were derived from Saucier (1997) and ranged from 0 (*I know the meaning of the word and think it is used extremely often to describe people*) to 9 (*I don't know the meaning of the word*), with a midpoint of 4.5.

coherence of personality self-reports are simply due to verbal comprehension. However, comparisons at this level of analysis could be made only with an effective sample size of five—the five domain scales. To investigate more sensitively the relation between comprehension difficulty and coherence, we also conducted analyses at the item level. Specifically, we tested whether more difficult BFI items showed (a) less within-domain coherence at young ages and (b) larger age gains in coherence than easier items showed. To conduct these item-level tests, we indexed coherence as the corrected item–total correlation, a statistic that reflects the extent to which an individual item correlates with the mean of the other items on its scale.

Table 3 presents correlations between (a) item-level within-domain coherence at age 10 and at age 20 (separately for male and female participants), (b) gain in coherence from age 10 to age 20 (separately for male and female participants), and (c) the two indices of item-comprehension difficulty. Correlations involving unfamiliarity were computed across the 40 items for which unfamiliarity ratings were available; all other correlations were computed across the full set of 44 BFI items. These correlations illustrate three key points. First, the robustness of the item-level coherence index is evidenced by its strong cross-gender generalizability correlations at age 10 ($r = .92$), at age 20 ($r = .95$), and for the gain from age 10 to age 20 ($r = .90$). These high correlations indicate that the most and least coherent items, and the items that showed the largest and smallest age gains in coherence, were very similar for male and female participants.

Second, item-level coherence showed negative associations with the two indices of comprehension difficulty at both age 10 and age 20; these correlations, presented in the two bottom rows of Table 3, were modest to moderate in size, ranging from –.11 to –.34. Although these correlations did not all reach conventional levels of statistical significance with samples of 40 or 44 items, together they imply that both children and adults had somewhat more trouble responding coherently to difficult items than to easy items.

Third, correlations (shown in the lower right corner of Table 3) between gain in coherence from age 10 to age 20 and comprehen-

Table 3
Item-Level Correlations Between Coherence at Ages 10 and 20, Gain in Coherence from Ages 10 to 20, and Comprehension Difficulty

	Item-level coherence at age 10		Item-level coherence at age 20		Gain from age 10 to age 20	
Variable	Male	Female	Male	Female	Male	Female
Coherence at age 10 years						
Male	—					
Female	*.92**	—				
Coherence at age 20 years						
Male	.47*	.56*	—			
Female	.36*	.46*	*.95**	—		
Gain from age 10 to 20 years						
Male	−.68*	−.52*	.34*	.41*	—	
Female	−.66*	−.65*	.24	.38*	*.90**	—
Comprehension difficulty						
Difficulty of understanding for 10-year-olds	−.26*	−.30*	−.18	−.11	.14	.23
Unfamiliarity of trait words for adults	−.21	−.21	−.34*	−.20	−.03	.05

Note. Cross-gender generalizability correlations are set in italics. Gain in coherence from age 10 to age 20 is keyed such that positive correlations indicate that more difficult items showed larger age gains in coherence.
$^*p < .05$, one-tailed.

sion difficulty were uniformly modest in magnitude and not statistically significant. In other words, items that were more difficult to comprehend did not show much larger age gains in coherence than did easy items.[11] Taken together, the correlations presented in Table 3 imply that, although comprehension plays a role in determining the coherence of youths' personality self-reports, developmental trends in coherence cannot simply be attributed to children's difficulty in understanding the items.

General Discussion

What Have We Learned? Findings and Implications

The influence of individual differences in acquiescent responding on personality self-reports is particularly strong in childhood and early adolescence. Psychometricians have long cautioned that uncontrolled variation in acquiescent responding can have adverse effects on the validity of personality questionnaires (Jackson & Messick, 1958). Going beyond previous research, the present findings indicate that the consequences of acquiescence variation are particularly dire for ratings provided by children and early adolescents. These age groups showed much greater individual differences in acquiescence than did older adolescents and adults: There was fully twice as much variance in acquiescence scores at age 10 as at age 20. Moreover, prior to age 14, acquiescence variation had especially strong effects on the coherence, differentiation, and overall factor structure of the BFI. Indeed, when individual differences in acquiescence were not controlled, the BFI's intended factor structure could not be recovered at age 10.

In view of these findings, we recommend that when researchers examine self-reports provided by children and early adolescents, they control for individual differences in acquiescent responding. This is particularly important when conducting analyses at the item level or when using scales with an imbalance of true- and false-keyed items. This control can be accomplished by way of ipsatization (as we have done here), within-person centering, or partialling the within-person mean response (see Ten Berge, 1999). Whichever method is chosen, however, care must be taken to ensure that acquiescence is assessed and removed while preserving the substantive content of youths' self-reports (e.g., by computing acquiescence scores from a set of semantically opposite item pairs, rather than across all available items) and that validity is thus enhanced rather than reduced.

Some groups of respondents as young as age 10 can provide structurally valid Big Five self-reports. We found that, after controlling for variation in acquiescence, the intended five-factor structure of the BFI was recognizable in each specific age group from 10 to 20 years. Previous studies have recovered versions of the NEO-FFI's intended structure from the self-reports of smaller samples as young as age 12 (Allik et al., 2004; McCrae et al., 2002; Parker & Stumpf, 1998). At ages younger than this, special interview techniques have been needed to obtain reliable reports (Measelle, John, Ablow, Cowan, & Cowan, 2005). The present research, using a large and diverse sample of respondents, as well as a non-NEO instrument, extends to age 10 the lower bound at which the Big Five structure can be recovered from questionnaire self-reports.

There are important age differences in the characteristics of youths' personality self-reports—beyond those attributable to acquiescence—and these age differences vary substantially across personality domains. We tested whether there were age differences in the coherence and differentiation of youths' personality self-reports after individual differences in response style were controlled and—going beyond previous research—whether age trends in coherence and differentiation varied by personality domain. Averaging across the Big Five domains, we found substantial age gains in both coherence and differentiation. Analyses of coherence showed that a typical pair of items on the same BFI scale shared more than twice as much variance at age 20 as at age 10. Similarly, analyses of differentiation showed that a typical pair of BFI scales overlapped three times as much at age 10 as at age 20. These pronounced age differences strongly caution against the assumption that, from a measurement perspective, youths' self-reports can be considered equivalent to those of adults (cf. McCrae et al., 2002).

Looking more closely, we found that these aggregate age trends in coherence and differentiation (i.e., main effects of age) masked substantial variation among the age trends of different Big Five domains (i.e., age by domain interaction effects). Extraversion showed the largest age gain in coherence: Two typical Extraversion items shared more than six times as much variance at age 20 as they did at age 10. Yet Extraversion was already well differentiated from the other four domains at the youngest ages studied. By contrast, Agreeableness and Conscientiousness showed only trivial age gains in coherence but became much better differentiated with age, particularly from each other: The overlap between these two scales was eight times as large at age 10 as it was at age 20. Finally, Openness and Neuroticism showed moderate age gains in both coherence and differentiation.

Note that most of the Big Five domains showed rather different age trends for coherence than they did for differentiation. Indeed, some domains showed an increase only in coherence or only in differentiation. Importantly, this pattern of findings indicates that age differences in these two characteristics of youths' personality self-reports are not both caused by a single developmental process; multiple factors must be at work.

Developmental trends in the coherence of youths' personality self-reports are not simply due to verbal comprehension. Another novel contribution of the present research is its consideration of the comprehension difficulties that personality questionnaires may pose for children and adolescents. By identifying BFI items and scales that are particularly easy or difficult for younger respondents to understand, we were able to test whether age differences in the coherence of BFI self-reports were simply due to

[11] To check whether the modest size of these associations was an artifact of using difference scores, we also performed GLM analyses with age and gender as within-subject (i.e., within-BFI-item) factors and comprehension difficulty as a between-subject (i.e., between-item) factor; in each of these analyses, the interaction between age and comprehension difficulty was not significant, indicating that the effect of age did not differ strongly for easy versus difficult items. Also, redoing these correlational and GLM analyses by using Flesch–Kincaid reading grade levels as the index of comprehension difficulty (see Footnote 6) produced similar but weaker results.

verbal comprehension difficulties. At the scale level, we found that Openness—the domain with the most difficult items—did show a substantial age gain in coherence, yet Extraversion, a scale with substantially easier items, showed a much larger gain. At the item level, more difficult items showed somewhat less coherence than did easier items—but that was true at both age 10 and age 20—and more difficult items did not show especially large age gains in coherence. These results imply that, at least for a generally easy-to-comprehend measure like the BFI, comprehension is not the major determinant of developmental trends in the coherence of youths' personality self-reports.

Limitations and Future Directions

Generalizability across populations, measures, and methods of administration. In order to obtain the large sample size needed to compare coherence, differentiation, and factor structure across specific age groups, participants in the present research were recruited passively and completed a particular personality measure, the BFI, over the Internet. To what extent are the present findings generalizable to other populations, measures, and methods of administration?

Passive recruitment, although necessary to obtain the large sample size required for the present research, introduces the possibility of self-selection effects. The present participants, especially those in the younger age groups (where sample sizes were smaller), may not be fully representative of the general population in terms of their capacity to provide valid self-reports. Instead, passive recruitment of our sample probably selected for, at least to some extent, children who were particularly motivated and able to respond to a personality questionnaire. Note, however, that this kind of selection bias would not create the kind of age effects reported here; instead, it would dampen any age differences in measurement characteristics. Thus, the age trends in acquiescence variability, within-domain coherence, and between-domain differentiation observed in the present sample may underestimate the population trends.

Regarding generalizability across measures, some of our findings may not apply to personality instruments very different from the BFI. Specifically, the effect of verbal comprehension on coherence may be stronger for measures that include items more difficult for young respondents to understand. Similarly, we would expect less clear factor structures for measures that include many difficult items, or items irrelevant to young respondents (e.g., questions about marriage, career, or civic duties). Moreover, commonly used Big Five measures differ somewhat in their construct definitions, especially for Openness and Extraversion (John et al., in press; John & Srivastava, 1999). Such differences may affect the exact nature of domain-specific developmental trends in coherence and differentiation.

Regarding methods of administration, there is now good evidence that Internet, paper-and-pencil, and face-to-face assessment procedures produce similar results in college-aged and adult samples (Buchanan & Smith, 1999; Gosling et al., 2004; McGraw, Tew, & Williams, 2000; Robins, Trzesniewski, Tracy, Gosling, & Potter, 2002; Skitka & Sargis, 2006; Srivastava, John, Gosling, & Potter, 2003) and that paper-and-pencil, proctored computer, and unproctored Internet versions of personality questionnaires have very similar measurement properties (Chuah, Drasgow, & Roberts, 2006). Much less is known, however, about the characteristics of questionnaires administered to young respondents with different methods.

These concerns about generalizability are ameliorated by the fact that some of our results are consistent with previous research that has examined the characteristics of adolescents' personality self-reports—research that has sampled from very different populations, focused mostly on a different measure, namely the NEO-FFI, and used paper-and-pencil administration. For example, our finding that most measurement properties reach asymptote by late adolescence is consistent with the results of a previous study that administered an Estonian translation of the NEO-FFI via paper and pencil in a secondary school setting (Allik et al., 2004). Moreover, the integrity of our Internet administration is evidenced by the very small percentage of cases (less than 3% of all those available) eliminated from analysis due to evidence of indifferent or random responding. Nevertheless, results from the present research should be taken to indicate that some groups of children and adolescents are able to provide coherent and differentiated personality self-reports with some instruments. They do not imply, of course, that all questionnaires are appropriate for use with all child and adolescent samples, nor that Internet and paper-and-pencil administration are exactly equivalent for young respondents.

Further research is needed to examine variables, beyond chronological age and item comprehension, that define groups of youths who can and cannot provide coherent and differentiated personality self-reports. One study from the adult personality literature suggested that the hierarchical versus non-hierarchical structure of concepts may be one such variable (Toomela, 2003; see also Allik & McCrae, 2004). In this study, the five-factor structure of the NEO PI-R was clearly recovered in groups of Estonian men who showed "scientific concepts" thinking (i.e., hierarchical classification of concepts) but was not recognizable among those who showed "everyday concepts" thinking (i.e., non-hierarchical classification).

An even lower age limit for personality questionnaire self-reports. The present research provides evidence that some children as young as age 10 can provide coherent and differentiated personality self-reports on a questionnaire that—although relatively easy to understand—was developed primarily for use with adult samples. Results from one previous study suggest that, by using a special puppet-interview procedure, reliable and valid Big Five self-reports begin to emerge between ages 5 and 7 (Measelle et al., 2005). Taken together, these findings raise the question of whether children even younger than age 10 can provide valid self-reports on personality questionnaires—especially on those developed specifically to assess personality in childhood (e.g., Mervielde & De Fruyt, 2002).

Continued efforts to develop child-appropriate personality self-report measures are thus needed. Such research can help investigate the features of personality questionnaires, including overall ease of comprehension and relevance for children and adolescents, that make questionnaires particularly well or poorly suited for younger populations (cf. Angleitner et al., 1986; Goldberg, 1963) as well as directly test the equivalence of Internet versus paper-and-pencil administration of questionnaires to young respondents.

Causes of age differences in measurement properties. Now that more is known about the nature of age differences in the coherence, differentiation, and factor structure of youths' person-

ality self-reports, longitudinal research is needed to directly test hypotheses about the developmental mechanisms underlying these trends. For example, longitudinal data are needed to rule out cohort effects as an explanation for age differences in measurement characteristics.

The findings of the present research can inform such future work in at least three ways. First, more research is needed to study how youths' understanding of personality trait terms is related to their capacity to provide valid personality self-reports. Across the BFI items, the present research found only a modest link between trait-word familiarity and the coherence of personality self-reports in late childhood and adolescence. This result, however, is bounded both by the age range studied and by the content of the BFI and should thus be regarded as the first word—rather than the last—on the link between comprehension and self-report validity.

Second, the finding that coherence and differentiation showed different developmental trends for most personality domains suggests that more than a single developmental factor is at work, and that the most important influences on coherence and differentiation are likely domain-specific ones. In particular, further research is needed to test whether the dramatic age gains documented here in the coherence of Extraversion self-reports are due to changes in the underlying structure of this trait domain from childhood to adulthood. Research is also needed to further investigate the hypothesis that age gains in the differentiation of youths' self-reports on Agreeableness and Conscientiousness are influenced by their increasingly complex understandings of different ways to be good or bad, to do right or wrong.

Third, the present research found that age differences in measurement properties were much larger from ages 10 to 14 than at later ages. This finding indicates that late childhood and early adolescence are key periods in the development of the capacity to report about one's own personality. Future research should focus specifically on this age range.

The basic structure of personality traits in childhood and adolescence. The present research was designed, most centrally, to address questions about developmental trends in the characteristics of personality self-reports from late childhood to adulthood. To best meet this aim, we obtained self-reports by using the same personality instrument in each age group of our sample—an instrument that is generally easy to understand, assesses traits relevant to children and adolescents, and has a clear multidimensional structure by early adulthood. We did not use different instruments at different ages. Put another way, we adopted the developmental analogue of an etic, rather than emic, measurement approach (Berry, 1969; Paunonen, 2000). For our purposes, this etic approach had the necessary feature of separating age effects from method effects: If we instead had adopted an emic approach and used different measures at different ages, we could not be sure that any observed age trends in measurement characteristics reflected developmental changes rather than the idiosyncrasies of the different age-specific instruments.

By contrast, the present approach is less well suited for addressing questions about whether the basic structure of personality traits differs by age. More specifically, basic trait structure is only one of several factors that likely influenced the structure of BFI self-reports documented here; other influences likely include age-related cognitive capacities as well as the "built-in" Big Five structure of the BFI itself. For these reasons, our conclusions regarding basic structure must be cautious: First, certain of the present findings (e.g., the pronounced age gain in coherence for Extraversion) provide some support for the view that the basic structure of personality traits changes over the course of childhood and adolescence (Eaton, 1994; John et al., 1994; van Lieshout & Haselager, 1994). Second, the rather clear five-factor structure that we obtained even at age 10 does not prove that the Big Five constitutes a comprehensive set of basic personality dimensions in late childhood. Thus, there is a need for more research designed to directly address the issue of basic structure (e.g., Goldberg, 2001; Halverson et al., 2003; John et al., 1994; Kohnstamm et al., 1998; van Lieshout & Haselager, 1994).

The external validity of youths' self-reports. The present research carefully examined whether youths' Big Five self-reports conform to the familiar five-factor structure (they do, once individual differences in acquiescence are taken into account) and whether the characteristics of these self-reports differ by age and domain (they do, even beyond the effects of acquiescence). Now, knowing that personality self-reports show some structural validity even by age 10, future studies should examine external validity. Adult descriptions of youths' personalities are known to predict important social, academic, and mental health outcomes (e.g., Halverson et al., 2003; John et al., 1994). Are youths' self-reports also predictive of these outcomes? Is some of the predictive power provided by youths' self-reports unique from that of adults' descriptions? Does the external validity of youths' self-reports increase with age? Future research could obtain external data from smaller samples to answer these questions.

Conclusion

Taken as a whole, the present findings provide encouragement for researchers interested in obtaining psychological self-reports from children and adolescents. At the same time, they caution against the assumption that, from a measurement perspective, youths' self-reports are equivalent to those of adults. Our own view is that questionnaire self-reports from children and adolescents—obtained while safeguarding the welfare of young respondents and interpreted in light of age differences in the characteristics of such reports—are a source of information that is too often neglected. Because descriptions provided by parents or teachers are based on limited information (i.e., only the child's overt actions in the classroom or at home) and may be biased by an adult-centric construal of child behavior, youths' self-reports provide a unique and valuable window into psychological development during late childhood and adolescence.

References

Achenbach, T. M. (1966). The classification of children's psychiatric symptoms: A factor-analytic study. *Psychological Monographs: General and Applied, 80,* 1–37.

Achenbach, T. M. (1978). The Child Behavior Profile: I. Boys aged 6–11. *Journal of Consulting and Clinical Psychology, 46,* 478–488.

Achenbach, T. M. (1979). The Child Behavior Profile: II. Boys aged 12–16 and girls aged 6–11 and 12–16. *Journal of Consulting and Clinical Psychology, 47,* 223–233.

Allik, J., Laidra, K., Realo, A., & Pullman, H. (2004). Personality development from 12 to 18 years of age: Changes in mean levels and structure of traits. *European Journal of Personality, 18,* 445–462.

Allik, J., & McCrae, R. R. (2004). Escapable conclusions: Toomela (2003) and the universality of trait structure. *Journal of Personality and Social Psychology, 87,* 261–265.

Angleitner, A., John, O. P., & Löhr, F. (1986). It's what you ask and how you ask it: An itemmetric analysis of personality questionnaires. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaire: Current issues in theory and measurement* (pp. 61–108). Berlin, Germany: Springer-Verlag.

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Di Blas, L., et al. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology, 86,* 356–366.

Benet-Martínez, V., & John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait–multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology, 75,* 729–750.

Berry, J. W. (1969). On cross-cultural comparability. *International Journal of Psychology, 4,* 119–128.

Block., J., & Block, J. H. (1980). *The California Child Q-set.* Palo Alto, CA: Consulting Psychologists Press.

Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology, 90,* 125–144.

Byrne, B., & Shavelson, R. (1996). On the structure of social self-concept for pre-, early, and late adolescents: A test of the Shavelson, Hubner, & Stanton (1976) model. *Journal of Personality and Social Psychology, 70,* 599–613.

Caspi, A., & Shiner, R. L. (2006). Personality development. In W. Damon & R. M. Lerner (Series Eds.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 300–365). Hoboken, NJ: Wiley.

Chuah, S. C., Drasgow, F., & Roberts, B. W. (2006). Personality assessment: Does the medium matter? No. *Journal of Research in Personality, 40,* 359–376.

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.

DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant study. *Journal of Personality and Social Psychology, 91,* 1138–1151.

Digman, J. M. (1994). Child personality and temperament: Does the Five-Factor Model embrace both domains? In C. F. Halverson, Jr., G. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 323–338). Hillsdale, NJ: Erlbaum.

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

Donahue, E. M. (1994). Do children use the Big Five, too? Content and structural form in personality description. *Journal of Personality, 62,* 45–66.

Eaton, W. O. (1994). Temperament, development, and the five-factor model: Lessons from activity level. In C. F. Halverson, Jr., & G. A. Kohnstamm (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 173–187). Hillsdale, NJ: Erlbaum.

Erikson, E. H. (1968). *Identity: Youth and crisis.* New York: Norton.

Feldt, L. S. (1969). A test of the hypothesis that Cronbach's alpha or Kuder–Richardson coefficient twenty is the same for two tests. *Psychometrika, 34,* 363–373.

Flavell, J. H., Miller, P. H., & Miller, S. A. (1993). *Cognitive development* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Garner, R. (1990). Children's use of strategies in reading. In D. F. Bjorklund (Ed.), *Children's strategies: Contemporary views of cognitive development.* Hillsdale, NJ: Erlbaum.

Goldberg, L. R. (1963). A model of item ambiguity in personality assessment. *Educational and Psychological Measurement, 23,* 467–492.

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 data: Derivation of Big-Five Factor scores from each of six samples. *Journal of Personality, 69,* 709–743.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist, 59,* 93–104.

Halverson, C. F., Jr., Havill, V. L., Deal, J., Baker, S. R., Victor, J. B., Pavlopoulos, V., et al. (2003). Personality structure as derived from parental ratings of free descriptions of children: The inventory of child individual differences. *Journal of Personality, 71,* 995–1026.

Hampson, S. E., & Goldberg, L. R. (2006). A first large cohort study of personality trait stability over the 40 years between elementary school and midlife. *Journal of Personality and Social Psychology, 91,* 763–779.

Harter, S. (1999). *The construction of the self: A developmental perspective.* New York: Guilford Press.

Harter, S. (2006). The self. In W. Damon & R. M. Lerner (Series Eds.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 505–570). Hoboken, NJ: Wiley.

Harter, S., & Monsour, A. (1992). Developmental analysis of conflict caused by opposing attributes in the adolescent self-portrait. *Developmental Psychology, 28,* 251–260.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* New York: Basic Books.

Jackson, D. N., & Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin, 55,* 243–252.

John, O. P. (1990). The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 66–100). New York: Guilford Press.

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., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory—Versions 4a and 54.* Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

John, O. P., Naumann, L., & Soto, C. J. (in press). The Big Five trait taxonomy: Discovery, measurement, and theoretical issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed.). New York: Guilford Press.

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. 102–139). New York: Guilford Press.

Knowles, E. S., & Condon, C. A. (1999). Why people say "yes": A dual-process theory of acquiescence. *Journal of Personality and Social Psychology, 77,* 379–386.

Kohnstamm, G. A., Halverson, C. F., Jr., Mervielde, I., & Havill, V. (Eds.). (1998). *Parental description of child personality: Developmental antecedents of the Big Five?* Mahwah, NJ: Erlbaum.

Lamb, M. E., Chuang, S. S., Wessels, H., Broberg, A. G., & Hwang, C. P. (2002). Emergence and construct validation of the Big Five factors in early childhood: A longitudinal analysis of their ontogeny in Sweden. *Child Development, 73,* 1517–1524.

Markey, P. M., Markey, C. N., Tinsley, B. J., & Ericksen, A. J. (2002). A preliminary validation of preadolescents' self-reports using the five-

factor model of personality. *Journal of Research in Personality, 36,* 173–181.

Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early adulthood. *Journal of Educational Psychology, 81,* 417–430.

Marsh, H. W., & Ayotte, V. (2003). Do multiple dimensions of self-concept become more differentiated with age? The differential distinctiveness hypothesis. *Journal of Educational Psychology, 95,* 687–706.

McCrae, R. R., Costa, P. T., Jr., Terracciano, A., Parker, W. D., Mills, C. J., De Fruyt, F., & Mervielde, I. (2002). Personality trait development from age 12 to age 18: Longitudinal, cross-sectional and cross-cultural analyses. *Journal of Personality and Social Psychology, 83,* 1456–1468.

McCrae, R. R., Herbst, J. H., & Costa, P. T., Jr. (2001). Effects of acquiescence on personality factors structures. In R. Riemann, F. M. Spinath, & F. Ostendorf (Eds.), *Personality and temperament: Genetics, evolution, and structure* (pp. 217–231). Berlin, Germany: Pabst Science Publishers.

McGraw, K. O., Tew, M. D., & Williams, J. E. (2000). The integrity of Web-delivered experiments: Can you trust the data? *Psychological Science, 11,* 502–506.

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., & De Fruyt, F. (2002). Assessing children's traits with the Hierarchical Personality Inventory for Children. In B. De Raad & M. Perugini (Eds.), *Big Five assessment* (pp. 129–142). Seattle, WA: Hogrefe & Huber.

Montemayor, R., & Eisen, M. (1977). The development of self-conceptions from childhood to adolescence. *Developmental Psychology, 13,* 314–319.

Osherson, D. N., & Markman, E. (1975). Language and the ability to evaluate contradictions and tautologies. *Cognition, 3,* 213–226.

Parker, W. D., & Stumpf, H. (1998). A validation of the five-factor model of personality in academically talented youth across observers and instruments. *Personality and Individual Differences, 25,* 1005–1025.

Paunonen, S. V. (2000). Construct validity and the search for cross-cultural consistencies in personality. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 123–140). Norwell, MA: Kluwer Academic.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41,* 203–212.

Robins, R. W., John, O. P., & Caspi, A. (1994). Major dimensions of personality in early adolescence: The Big Five and beyond. In C. F. Halverson, Jr., G. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 267–291). Hillsdale, NJ: Erlbaum.

Robins, R. W., Trzesniewski, K. H., Tracy, J. L., Gosling, S. D., & Potter, J. (2002). Global self-esteem across the life span. *Psychology and Aging, 17,* 423–434.

Saucier, G. (1997). Effects of variable selection on the factor structure of person descriptors. *Journal of Personality and Social Psychology, 73,* 1296–1312.

Shiner, R. L. (2006). Temperament and personality in childhood. In D. K. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (pp. 213–230). Mahwah, NJ: Erlbaum.

Siegler, R. S. (1998). *Children's thinking* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Skitka, L. J., & Sargis, E. G. (2006). The Internet as psychological laboratory. *Annual Review of Psychology, 57,* 529–555.

Smith, M. K. (1941). Measurement of the size of general English vocabulary through the elementary grades and high school. *Genetic Psychology Monograms, 24,* 311–345.

Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology, 84,* 1041–1053.

Ten Berge, J. M. F. (1999). A legitimate case of component analysis of ipsative measures, and partialling the mean as an alternative to ipsatization. *Multivariate Behavioral Research, 34,* 89–102.

Toomela, A. (2003). Relationships between personality structure, structure of word meaning, and cognitive ability: A study of cultural mechanisms of personality. *Journal of Personality and Social Psychology, 85,* 723–735.

United States Department of Health and Human Services. (2005). *Federal Code of Regulations, Title 45, Part 46: Protection of human subjects.* Rockville, MD: Author.

Van Lieshout, C. F. M., & Haselager, G. J. T. (1994). The Big Five personality factors in Q-sort descriptions of children and adolescents. In C. F. Halverson, Jr., & G. A. Kohnstamm (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 293–318). Hillsdale, NJ: Erlbaum.

Yuill, N. (1993). Understanding of personality and dispositions. In M. Bennett (Ed.), *The development of social cognition: The child as psychologist* (pp. 87–110). New York: Guilford Press.

Appendix

Computing the Content-Balanced Acquiescence Index and Ipsatizing the BFI Items

As described in the Method section, we indexed each person's level of acquiescence as their mean response to a set of 16 pairs of BFI items with opposite implications for personality (e.g., item 1, "Is talkative," vs. item 21, "Tends to be quiet"). Standard BFI item numbers (Benet-Martínez & John, 1998; John et al., in press; John & Srivastava, 1999) for the 16 pairs of opposite items are as follows: 1 and 21, 6 and 16, 31 and 36, 2 and 17, 7 and 12, 27 and 42, 32 and 37, 3 and 43, 8 and 13, 18 and 33, 23 and 28, 9 and 19, 24 and 29, 34 and 39, 5 and 35, and 30 and 41. The SPSS syntax presented below first computes, for each person, an acquiescence score (the mean of their 16 × 2 = 32 item responses) and a response extremeness score (the standard deviation of their 32 item responses). These two scores are then used to ipsatize the person's full set of 44 BFI item responses. The syntax assumes that the variables for the 44 BFI items are named bfi1 to bfi44 in standard order.

```
* Compute within-person response means (bfiave) and standard deviations (bfistd).
COMPUTE bfiave = mean(bfi1, bfi6, bfi16, bfi21, bfi31, bfi36, bfi2, bfi7, bfi12, bfi17, bfi27, bfi32, bfi37, bfi42, bfi3, bfi8, bfi13, bfi18, bfi23, bfi28, bfi33, bfi43, bfi9, bfi19, bfi24, bfi29, bfi34, bfi39, bfi5, bfi30, bfi35, bfi41).
COMPUTE bfistd = sd(bfi1, bfi6, bfi16, bfi21, bfi31, bfi36, bfi2, bfi7, bfi12, bfi17, bfi27, bfi32, bfi37, bfi42, bfi3, bfi8, bfi13, bfi18, bfi23, bfi28, bfi33, bfi43, bfi9, bfi19, bfi24, bfi29, bfi34, bfi39, bfi5, bfi30, bfi35, bfi41).
EXECUTE.
* Compute ipsatized BFI items (zbfi).
COMPUTE zbfi1 = (bfi1-bfiave)/bfistd.
COMPUTE zbfi2 = (bfi2-bfiave)/bfistd.
COMPUTE zbfi3 = (bfi3-bfiave)/bfistd.
.
.
.
COMPUTE zbfi44 = (bfi44-bfiave)/bfistd.
EXECUTE.
```

Received December 16, 2005
Revision received October 16, 2007
Accepted November 12, 2007 ■