

Validity of Big Five Personality Judgments in Childhood: A 9 Year Longitudinal Study

JENS B. ASENDORPF^{1*} and MARCEL A. G. VAN AKEN²

¹*Humboldt-Universität zu Berlin, Germany*

²*Utrecht University, The Netherlands*

Abstract

In a 9 year longitudinal study over childhood, the Big Five personality traits were assessed at ages 4–6 by teacher Q-sorts, at age 10 by parental Q-sorts, and at age 12 by parental and friend ratings on bipolar adjective scales. The Big Five Q-sort indices were based on definitions proposed by John, Caspi, Robins, Moffitt, and Stouthamer-Loeber (1994) for adolescent boys. They were related to judgments and behavioural observations of inhibition and aggressiveness, and to antecedents and consequences of school achievement such as IQ and cognitive self-esteem. Neuroticism and low extraversion correlated with social inhibition, low agreeableness and low conscientiousness with aggressiveness, and conscientiousness and/or culture/intellect/openness with antecedents and outcomes of school achievement. These correlations were consistently found throughout childhood. Copyright © 2002 John Wiley & Sons, Ltd.

The Five-Factor Model (FFM) of personality description has attracted much interest over the past two decades. According to this model, five relatively independent, extremely broad dimensions ('Big Five') explain a major portion of judged interindividual differences in personality. These dimensions are extraversion, neuroticism (or the reversely coded dimension emotional stability), agreeableness, conscientiousness, and culture, intellect, or openness to new experience, which is the least agreed-upon factor. This model has been empirically supported by research mainly on North-American and Northern European adults' self- and other-descriptions of personality (John & Srivastava, 1999). The Big Five resulted as the last step of a multi-step lexical approach procedure that started with the full lexicon of personality-descriptive terms, and reduced that information through item elimination and factor analysis (Goldberg, 1990; John, Angleitner, & Ostendorf, 1988). In addition, most of the variance captured by traditional personality questionnaires can be accounted for by the Big Five (McCrae & Costa, 1990).

*Correspondence to: Jens B. Asendorpf, Institut für Psychologie, Humboldt-Universität zu Berlin, Oranienburger Strasse 18, 10178 Berlin, Germany. E-mail: asen@rz.hu-berlin.de

Received 22 November 2001

Accepted 28 May 2002

Building on this model of personality differences in adulthood and on early replications of the FFM through teacher ratings of children (Digman, 1989), researchers have begun to study the emergence of the Big Five structure in parental, teacher, and peer descriptions of children's and adolescents' personality (John, Caspi, Robins, Moffitt, & Stouthamer-Loeber, 1994; Kohnstamm, Halverson, Mervielde, & Havill, 1998; Mervielde, Buyst, & De Fruyt, 1995; Mervielde & De Fruyt, 2000; Scholte, van Aken, & Van Lieshout, 1997; Van Lieshout & Haselager, 1994). According to a recent review by Mervielde and Asendorpf (2000), the FFM has been supported best by parental and teacher ratings of personality. Studying the factor structure of the California Child Q-Set (Block & Block, 1980) instead of the factor structure of ratings, both Van Lieshout and Haselager (1994) and John et al. (1994), in addition to the original five dimensions, found evidence for two other dimensions in Dutch children and adolescents and ethnically diverse adolescent US boys, but these two dimensions were inconsistent across the two studies.

The aim of the present study was to address two validity questions that are, in our view, at least as important. If the Big Five factors are operationalized by reasonable parental, teacher, and peer judgment measures, (i) how valid are such Big Five measures in terms of concurrently assessed important external variables and (ii) how consistent are these correlates of the Big Five across age?

Concerning the question (i), studies of the Big Five in childhood have focused on three main domains of personality and achievement: externalizing tendencies and internalizing tendencies, IQ and school performance, and domain-specific self-esteem. According to John et al. (1994), externalizing problems such as aggression, stealing and lying, inattention, impulsivity, and hyperactivity (Achenbach, Howell, Quay, & Conners, 1991) imply a pattern of low agreeableness and low conscientiousness. In contrast, internalizing problems such as anxiety, somatic complaints, social inhibition, and social withdrawal (Achenbach et al., 1991) imply a pattern of low extraversion and high neuroticism.

Studying extreme groups of boys with externalizing or internalizing problems, John et al. (1994) confirmed the low agreeableness and low conscientiousness of externalizing boys, along with increased scores in extraversion, and also the high neuroticism in boys with internalizing problems, but not the expected low extraversion in these boys. This finding may be specific to the sample because other studies have consistently found an association between both high neuroticism and low extraversion with internalizing difficulties in childhood (Asendorpf, Borkenau, Ostendorf, & van Aken, 2001; Rubin & Asendorpf, 1993).

Several other studies have reported associations between the Big Five and measures of externalizing and internalizing problems, confirming most of the relations between extraversion, agreeableness, and conscientiousness and externalizing problems, and between neuroticism and internalizing problems, but also presenting some incidental findings.

Ehrler, Evans, and McGhee (1999) found that children with low scores on agreeableness and conscientiousness exhibited social problems, conduct problems, attention deficits, and hyperactivity. Children with low scores on openness exhibited problems in social behavior, conduct, and attention, whereas neuroticism was associated with anxiety and depression. Victor (1994) reported various associations between the Big Five and several measures of problem behavior. Conduct disorder was related to extraversion, and low agreeableness and conscientiousness; socialized aggression to extraversion, and low agreeableness and openness; anxiety-withdrawal to low extraversion, neuroticism and low openness; attention problems to low conscientiousness and openness; and motor excess to low agreeableness, to neuroticism and to low openness.

In a study on the development of behavioural and emotional problems in clinic-referred children, Huey and Weisz (1997) found that extraversion and low agreeableness were independent predictors of externalizing problems, whereas only neuroticism predicted internalizing problems. In a similar study, van Aken and Heutinck (1998) found that teacher ratings of low agreeableness at ages 7, 10, and 12 predicted antisocial behaviour at age 20. In several other studies on agreeableness, Graziano and colleagues (Graziano & Ward, 1992; Jensen-Campbell & Graziano, 2001) found that this factor was related to processes and outcomes during interpersonal conflicts, and also to teacher-rated school adjustment.

Shiner (2000) found that in 8–12 year olds a higher order factor agreeableness (a combined parent and child description) was related to concurrent rule-abiding conduct (combined peer and parent judgments) and peer nominations on social competence. A second higher order factor, surgent engagement (including extraversion, expressiveness, and attention), was also related to social competence. Finally, in a sample of 4–12 year olds, Goedhart, Treffers, and Kohnstamm (1994) found that agreeableness, conscientiousness, extraversion, and openness differentiated between clinically depressive and control children.

John et al. (1994) also studied the relation between their Q-sort based Big Five indices and IQ and school performance. As they expected on the basis of similar findings by Digman (1989), both IQ and school grades were consistently related to conscientiousness and openness. Conscientiousness and openness, or intellect, were also reported in several other studies to have relations with measures of IQ and/or school achievement. Mervielde et al. (1995) found that conscientiousness and openness, but also extraversion, significantly predicted average grades across three primary school age levels. Judge, Higgins, Thoresen, and Barrick (1999) found relations between general mental ability (measured with the Stanford–Binet intelligence test) and openness and conscientiousness, but also neuroticism. Shiner (2000) reported that the higher order factor surgent engagement (including extraversion) was related to concurrent academic functioning and IQ-scores. Regarding motivational school factors, Lay, Kovacs, and Danto (1998) found associations between conscientiousness and teacher ratings of procrastination in a sample of 7–11 year olds.

Concerning domain-specific self-esteem, Graziano and Ward (1992), using the Harter scales, reported that all Big Five factors were clearly related to domains of self-esteem. Agreeableness was significantly related to self-esteem in the domains of academics, appearance, and conduct but not social acceptance or athletics. Extraversion was related to self-esteem regarding academics, appearance, athletics, and appearance but not conduct. Conscientiousness was related to self-esteem regarding academics and athletics. Neuroticism was related to self-esteem in the domains of academics, social acceptance, and athletics, but not appearance or conduct, and openness was related to self-esteem concerning academics, social acceptance, athletics, and appearance, but not conduct.

In this highly diverse literature, no age shifts in the correlates of the Big Five are easily detectable over childhood. Therefore, we assumed with regard to question (ii) that external correlates of the Big Five are constant over childhood. Together, this literature review suggested the following hypothesis: neuroticism and low extraversion correlate with social inhibition, low agreeableness and low conscientiousness with aggressiveness, and conscientiousness and culture/intellect/openness with antecedents and outcomes of school achievement. These correlations are consistently found all throughout childhood.

We tested this hypothesis with available data from the Munich Longitudinal Study on the Genesis of Individual Competencies (LOGIC; Weinert & Schneider, 1999). We

operationalized the Big Five in early to middle childhood through teacher and parental Q-sort indices, following the approach of John et al. (1994), and in middle to late childhood through parental and friend ratings on Big Five scales that were derived from a large pool of German personality-descriptive adjectives (Ostendorf, 1990). Although this design is not optimal for the analysis of the stability of interindividual differences because both the instruments and the judges changed over age, the use of multiple measures and judges and the longitudinal sample provide a particularly strong test for the continuity of external correlates of the personality judgments.

Concerning these correlates, we studied social inhibition and aggressiveness through parental judgments and, if available, also behavioural observations in the preschool and kindergarten peer group. For achievement, we studied IQ, cognitive self-esteem, and deviation from expected grade. In Germany, deviation from expected school grade can be due to late schooling, repetition of a class (a regular intervention of the school in case of poor grades), or advancing to a higher grade (in the case of gifted children). Negative deviations indicate problems with schooling and can be considered a major real-life developmental outcome of German children's cognitive and social-emotional competence.

METHOD

Participants

The participants were part of the Munich Longitudinal Study on the Genesis of Individual Competencies (LOGIC). The LOGIC sample originally consisted of 230 children (119 boys, 111 girls) who were studied every year from their first or second year in preschool until age 12. The sample was reasonably unbiased because the schools were selected from a broad spectrum of neighbourhoods, more than 90% of the parents who were asked for permission gave their consent for their child's participation, and attrition was low (19% over 9 years) and unsystematic (see Weinert & Schneider, 1999). The present study includes those 151 participants who were judged by their main preschool and kindergarten teacher for their personality in Q-sorts at ages 4, 5, and 6 years.

Assessments and measures

Table 1 provides an overview of the assessments. The smaller sample sizes for the parental judgments and behavioural observations in the early assessments were mainly due to the fact that these assessments were only scheduled for children who attended their preschool full-time at age 4; the smaller sample size for the friend judgments was due to refusals by the subjects to nominate a friend, or refusals of the nominated friend to provide judgments. Because most assessments were described in detail by Asendorpf (1990) and Asendorpf and van Aken (1994, 1999), only the most important information for these assessments is provided here.

Teacher Q-sort

The 54-item short version of the California Child Q-Set (CCQ; Block & Block, 1980) was adapted into German (Göttert & Asendorpf, 1989). All LOGIC participants attended a preschool, or kindergarten, from age 4 to 6. At the end of each school year, the child's main teacher provided a Q-sort description of the child according to a fixed, nine-point distribution, ranging from 'extremely uncharacteristic' to 'extremely characteristic'. The teacher was instructed to sort exactly six items into each of the nine categories (forced

Table 1. Ages and numbers of the LOGIC children at the various assessments

Age (years)	<i>n</i>	Assessment	By
4–6	151	California Child Q-Set (54-item version) ^a	Teachers
	84	Social inhibition and aggressiveness scales ^a	Parents
	91	Social inhibition and aggressiveness in preschool ^a	Observers
	141	IQ ^b	IQ-tests
7	142	Deviation of school grade from grade 1	School authorities
10	124	California Child Q-Set (54-item version)	Parents
	83	Social inhibition and aggressiveness scales	Parents
	124	IQ ^c	IQ-tests
	124	Deviation of school grade from grade 4	School authorities
	123	Cognitive and social self-esteem	Self
12	111	Big Five bipolar adjective scales (40 items)	Parents
	89	Big Five bipolar adjective scales (40 items)	Friends
	111	Social inhibition and aggressiveness scales	Parents
	111	IQ ^d	IQ-tests
	111	Deviation of school grade from grade 6	School authorities
	110	Cognitive and social self-esteem	Self

^aAggregate of three assessments at ages 4, 5, and 6.

^bAggregate of four assessments at ages 4, 5, and 6.

^cAggregate of two assessments at ages 9 and 10.

^dAggregate of two assessments at age 12.

equal distribution). To increase the reliability of the judgments, the three Q-sorts at ages 4–6 were averaged itemwise.

Parental Q-sort

The same Q-sort instrument was answered at age 10 by the child's main caregiver (nearly always the mother).

Parental bipolar adjective Big Five scales

At age 12, the child's main caregiver (nearly always the same as at age 10) judged the child on 60 age-appropriate bipolar adjective pairs that were derived from the highest-loading items on the first five factors of a pool of 179 bipolar adjective pairs by Ostendorf (1990) in a multi-step procedure. Five eight-item Big Five scales were derived from these 60 items on the basis of a factor analysis (see Asendorpf & van Aken, 1999, for details, and the Appendix for the items of the scales). The items were answered on a five-point response scale. The reliabilities of the scales were satisfactory (median $\alpha = 0.86$, range 0.83–0.91). The fifth factor was labelled culture instead of openness or intellect because of a focus on knowledge and intelligence rather than creativity, imagination, and curiosity (openness) or intellectual competence alone (intellect). To make the scoring of the scale that was labelled emotional stability by Asendorpf and van Aken (1999) consistent with the neuroticism scale based on the Q-sort (see below), we scored it in the opposite direction and labelled it neuroticism.

Friend bipolar adjective Big Five scales

The same Big Five scales were answered by a same-sex friend nominated by the child. The parents of these friends were asked by post for permission for their child to rate the LOGIC participant; 81% of these families cooperated. The reliabilities of the scales were sufficiently high (median $\alpha = 0.85$, range 0.78–0.87).

Parental inhibition and aggressiveness scales

The main caregiver (nearly always the mother) answered a questionnaire at ages 4, 5, 6, 10, and 12. The aggressiveness scale consisted of four items, and the inhibition scale consisted of eight items at ages 4–10 and four items at age 12. Item content was identical across age and referred to aggressiveness with peers and inhibition toward strangers. At ages 4–10, the four inhibition items were answered by the parent separately for inhibition toward adults and toward peers; at age 12, two of them referred to adults and the other two to peers. The items of the two scales were randomly mixed with other items and answered on a seven-point response scale. The reliabilities were satisfactory for both scales at all assessments, α at least 0.83 in each case. The three assessments at ages 4–6 were averaged.

Peer group observations

At 4, 5, and 6 years, children's contact initiation behaviour was observed for eight to 12 10 min periods on at least five different days during regular free play periods in their preschool or kindergarten class. Analysed in the present study were (i) inhibited approach (when a child approached an individual or a group the child stopped and looked at the individual or group for at least 3 s without speaking) and (ii) aggressive approach (including physical aggression, verbal aggression, and taking toys from a child without asking). Intercoder agreement among the seven trained coders was checked each year by parallel observations; it was satisfactory for all three years of observation (see Asendorpf, 1990, for details). The percentages of inhibited and aggressive approaches among all approaches served as the child's behavioural scores for inhibition and aggressiveness.

Intelligence tests

Verbal intelligence was assessed with the German versions of the Wechsler scales for preschool children (ages 4 and 5 years: HAWIVA; Eggert, 1978) or school-aged children (ages 9 and 12: HAWIK-R; Tewes, 1983). Nonverbal intelligence was assessed with the Columbia Mental Maturity Scale (Burgemeister, Blum, & Lorge, 1972) at ages 4 and 6, and with the German version of the Culture Fair Intelligence Test (CFT-20; Weiß & Osterland, 1979) at ages 10 and 12. Because the tests varied necessarily across assessments, and the available test norms referred to different birth cohorts and cultures, total IQ scores were computed for ages 4–6, 10, and 12 by averaging z -transformed test scores (computed for the full LOGIC sample) of the four assessments until age 6, the two assessments at 9 and 10 years, and the two assessments at 12 years; these scores were then transformed into IQ scores. The reliabilities of the three total IQ variables were satisfactory ($\alpha > 0.82$ in each case).

Deviation from expected school grade

At ages 7, 10, and 12, an expected school grade was determined on the basis of the child's age and the Bavarian rules of age-appropriate schooling. Deviations from expected grade were computed by subtracting children's expected school grades at ages 7, 10, and 12 (grades 1, 4, and 6) from their actual grades at these ages. Thus, except for a few gifted children who attend higher grades than expected, these deviation scores are zero or negative, and more negative scores indicate cognitive and/or social-emotional problems in school.

Social and cognitive self-esteem

At 10 and 12 years, children's domain-specific self-esteem was assessed with the German version of Harter's (1985) Self-Perception Profile for Children (Asendorpf & van Aken, 1993). This inventory includes six-item scales for assessing children's social self-esteem ('social acceptance') and cognitive self-esteem ('scholastic competence'); each item is rated

on a four-point scale. The internal consistencies were sufficiently high for both ages ($\alpha > 0.72$ in each case). For cognitive (but not for social) self-esteem there was a strong reference group effect due to different school tracks in Bavaria after grade 4. Children who continued in the lower school track showed no change in cognitive self-esteem, but children who continued in the higher school track (the 'Gymnasium') showed a drop in cognitive self-esteem. This well established effect (see e.g. Schwarzer, Lange, & Jerusalem, 1982) is attributed to the change in the reference group of the higher-achieving children. They compare themselves in the higher track only with higher-achieving children, and hence judge themselves less favourably than before the change. Therefore the age 12 data for cognitive self-esteem were corrected for this effect (see Asendorpf & van Aken, 1994).

RESULTS

Sample attrition

Attrition was studied by comparing children's scores at age 5 (when the full LOGIC sample with $N = 230$ was available) in verbal IQ, nonverbal IQ, and the parental inhibition and aggressiveness scales between the 151 participants of the present study and the remaining 79 children with t tests. In each case, not even a marginally significant difference was found ($t < 1.5$, $p > 0.10$). Thus, the initial sample of the present study was not biased for the main variables. Similar analyses did not reveal significant selection biases for (i) the subsample that was observed in preschool and judged by a parent in early childhood, or (ii) the subsample that was rated by a friend. Dropout was not significantly related to sex either.

Q-sort indices for the Big Five

Derivation of Q-sort indices

John et al. (1994) defined Big Five scales on the basis of the full 100-item CCQ with 48 items. The number of items per scale varied between seven (openness to experience) and 13 (agreeableness), with reliabilities ranging from 0.53 (openness) to 0.83 (agreeableness). When we used the same items for defining Big Five scales on the basis of the 54-item German short version of the CCQ, 30 of the 54 items were retained. The number of items per scale varied between two (extraversion) and nine (agreeableness), with reliabilities ranging from 0.61 (openness) to 0.84 (conscientiousness). Because inspection of the corrected item-scale correlations indicated that deletion of some items would substantially increase the reliability of the scale, 24 items with corrected item-scale correlations above 0.30 were retained. Inspection of the correlations of the items of a scale with other scales indicated problems with two agreeableness items and one conscientiousness item because of high correlations with other scales. After deletion of these three items, two scales were very short (extraversion, two items; openness, three items). Therefore, two new extraversion items and one new openness item were added from the remaining items, resulting in scales that varied from four to six items per scale, with reliabilities ranging from 0.72 (neuroticism) to 0.89 (conscientiousness) (see Table 3). Thus, despite their brevity, these Q-sort scales showed a sufficient reliability. These final scales were highly similar to those used by John et al. (1994) because 17 of the 20 items were also used by John et al. (1994). The items of these Big Five scales and their corrected item-scale correlations are presented in Table 2.

Table 2. Q-sort scales for the Big Five

CCQ item number and item	Corrected item–scale correlation ^a	
	Ages 4–6 (teacher)	Age 10 (parent)
Extraversion		
28. Is vital, energetic, lively	0.61	0.34
35. Is inhibited and constricted (R)	0.64	0.49
44. Tends to yield and give in in conflicts (R)	0.53	0.48
82. Is self-assertive	0.56	0.51
Neuroticism		
23. Is fearful and anxious	0.59	0.37
24. Tends to brood and ruminate or worry	0.30	0.38
43. Can recoup or recover after stress (R)	0.45	0.26
46. Tends to go to pieces under stress	0.46	0.35
60. Becomes anxious in unpredictable environments	0.65	0.39
78. Is easily offended, sensitive to criticism	0.33	0.33
Agreeableness		
2. Is considerate and thoughtful of other children	0.84	0.52
4. Gets along well with other children	0.73	0.50
6. Is helpful and cooperative	0.72	0.56
80. Teases other children (R)	0.73	0.38
Conscientiousness		
41. Is persistent in activities; does not give up easily	0.73	0.52
47. Has high standards of performance for self	0.56	0.45
66. Is attentive and able to concentrate	0.80	0.64
67. Is planful, thinks ahead	0.74	0.68
76. Can be trusted, is dependable	0.69	0.45
89. Is competent, skillful	0.73	0.54
Openness to experience		
36. Is resourceful in initiating activities	0.45	0.34
40. Is curious and exploring; open to new experiences	0.51	0.56
74. Becomes strongly involved in what s/he does	0.47	0.31
54. Is creative in perception, thought, work, or play	0.67	0.44

Items not used by John et al. (1994) in boldface. (R) = item is reversed.

^aCorrelation between the item and the remainder of the scale.

Inspection of the internal consistencies and cross-scale correlations indicated that the same scales could be used also for the parental Q-sort judgments at age 10. As Table 2 indicates, the corrected item–scale correlations were again above 0.30 with one exception, but in most cases lower than at ages 4–6, and the scale reliabilities were therefore approximately 0.10 lower for each scale, ranging from 0.61 (neuroticism) to 0.79 (conscientiousness) (see Table 3). This result can be mainly attributed to the fact that the teacher Q-sorts were averaged across three assessments, which increases the reliability of the aggregate, whereas there was only one parental Q-sort. All in all, the reliabilities were acceptable for both ages and judges.

The intercorrelations of the scales were higher at ages 4–6 than at age 10, which can be mainly attributed to the higher reliability of the teacher scales. The correlations between extraversion and neuroticism (negative) and the correlations between conscientiousness and openness were particularly high. Inspection of these four correlations at the item level showed that the correlations could not be reduced through revisions of the scales.

Table 3. Reliabilities and intercorrelations of the Q-sort Big Five scales

Teacher Q-sorts at ages 4–6	Parental Q-sorts at age 10				
	Extraversion	Neuroticism	Agreeableness	Conscientiousness	Openness
Extraversion	0.78/0.67	–0.47***	–0.18*	0.10	0.27**
Neuroticism	–0.64***	0.72/0.61	–0.03	–0.38***	–0.33***
Agreeableness	–0.18*	–0.23**	0.88/0.69	0.20*	0.05
Conscientiousness	–0.10	–0.25**	0.53***	0.89/0.79	0.49***
Openness	0.23**	–0.42***	0.36***	0.66***	0.72/0.63

Intercorrelations at ages 4–6 ($n = 151$) are presented below the diagonal, intercorrelations at age 10 ($n = 124$) above the diagonal, reliabilities (Cronbach's α for ages 4–6/age 10) on the diagonal in boldface.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Concurrent correlates

The concurrent external correlates of the Big Five Q-sort scales at ages 4–6 are shown in Table 4. The correlations in Table 4 suggest substantial external validity for the teacher Q-sort scales for extraversion, agreeableness, and conscientiousness. For significant correlates of extraversion the correlations with neuroticism were smaller than for extraversion, and for significant correlates of conscientiousness the correlations with openness were smaller than for conscientiousness. An even clearer picture emerged when the incremental validity of the Big Five Q-sort scales was studied by hierarchical multiple regression. For each Big Five scale, a significant external correlate was first regressed on all four other Big Five scales, and subsequently on the scale of interest, controlling the influence of the other scales on the external variable. Nine of the 18 significant external correlates survived this stringent test (printed in boldface in Table 4), and all were correlates of extraversion, agreeableness, or conscientiousness.

As expected, neuroticism and low extraversion correlated significantly with the parental judgments and the behavioural observations of social inhibition; both low agreeableness and low conscientiousness correlated significantly with aggressiveness; and

Table 4. Correlates of the Q-sort Big Five scales at ages 4–6

External variable	Teacher Q-sort scale				
	Extraversion	Neuroticism	Agreeableness	Conscientiousness	Openness
Parental inhibition scale	–0.55***	0.36***	0.15	0.25*	–0.07
Parental aggressiveness scale	0.09	0.00	–0.35***	–0.18(*)	–0.06
Peer group observations					
inhibition	–0.48***	0.20(*)	0.26*	0.26*	0.07
aggressiveness	0.30**	–0.11	–0.38***	–0.37***	–0.15
IQ	0.11	–0.12	0.19*	0.44***	0.32***
Deviation from expected grade	–0.02	–0.07	0.27**	0.32***	0.16

n Varied between 84 and 141 (see Method section). Deviation from expected school grade refers to age 7; all other variables were aggregated across ages 4, 5, and 6 years. Correlations in boldface indicate relations with a Big Five scale that remained significant ($p < 0.05$) when all other Big Five scales were statistically controlled by hierarchical multiple regression.

(*) $p < 0.07$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

conscientiousness correlated significantly with both IQ and deviation from expected school grade (see Table 4; because these were *a priori* predictions, the correlations can be tested with one-tailed tests for significance such that even the marginal correlations with $p < 0.07$, two tailed, confirmed the hypothesis significantly). Thus, our hypotheses was fully confirmed for early childhood.

However, the correlations were considerably smaller for neuroticism and openness than for extraversion and conscientiousness. When all remaining Big Five indices were statistically controlled in multiple regressions, only extraversion and conscientiousness, not neuroticism and openness, showed unique contributions to inhibition and achievement, respectively.

Intercorrelations of the bipolar adjective Big Five scales at age 12

The intercorrelations of the parental and friend bipolar adjective Big Five scales, and their consistency across judges, are presented in Table 5. Parent-judged neuroticism correlated significantly with all other parent scales, and friend-judged culture correlated significantly with all other friend scales. Thus, generalized positive versus negative evaluations of a child were reflected by the parental neuroticism scale and the friend culture scale.

In addition, neuroticism was not significantly correlated between the two judges (see Table 5). However, the discriminant validity of the scales with regard to their cross-judge consistency was good because the ten correlations between a parental scale and different friend scales, and the ten correlations between a friend scale and a different parent scale, were not significant in each case, and therefore these correlations are not presented in a separate table.

Concurrent correlates of the Big Five scales at ages 10 and 12

The concurrent external correlates of the parental and friend Big Five scales are shown in Table 6. As in the analyses for ages 4–6, significant external correlates of a Big Five scale were additionally controlled for the influence of other Big Five scales by hierarchical multiple regression.

Table 6 indicates that social inhibition and aggressiveness were again specifically related to low extraversion and agreeableness, respectively. Also, conscientiousness showed consistently lower correlations with low parent-judged aggressiveness than agreeableness (the same difference was found at ages 4–6; see Table 4), and neuroticism

Table 5. Intercorrelations of the parental and friend Big Five scales at age 12

Parental scales	Friend scales				
	Extraversion	Neuroticism	Agreeableness	Conscientiousness	Culture
Extraversion	0.36***	−0.23*	0.14	−0.11	0.32**
Neuroticism	−0.35***	0.20	−0.21*	−0.12	−0.30**
Agreeableness	0.23*	−0.33***	0.22*	0.46***	0.35***
Conscientiousness	0.08	−0.27**	0.30***	0.41***	0.34***
Culture	0.17	−0.39***	0.12	0.18	0.35***

Intercorrelations of the parental scales ($n = 111$) are presented below the diagonal, intercorrelations of the friend scales ($n = 89$) above the diagonal, consistencies between parents and friends on the diagonal in boldface.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 6. Concurrent correlates of the Big Five scales at ages 10 and 12

Big Five scale	Parental scales			Deviation from expected grade	Self-esteem	
	Inhibition	Aggressiveness	IQ		Cognitive	Social
Parent, age 10						
Extraversion	-0.42***	0.33**	0.30***	0.13	0.10	-0.05
Neuroticism	0.22*	-0.04	-0.37***	-0.12	-0.20*	-0.02
Agreeableness	0.03	-0.52***	-0.11	0.10	0.02	0.19*
Conscientiousness	0.17	-0.24*	0.46***	0.26**	0.25**	0.17
Culture	0.04	-0.19	0.41***	0.18*	-0.01	-0.04
Parent, age 12						
Extraversion	-0.53***	0.07	-0.05	0.11	-0.11	0.27**
Neuroticism	0.18(*)	0.20(*)	-0.19*	-0.19*	-0.22*	-0.31***
Agreeableness	0.07	-0.56***	-0.13	0.00	-0.18	0.07
Conscientiousness	0.17	-0.22*	0.03	0.04	0.19*	0.04
Culture	0.01	-0.13	0.52***	0.15	0.26**	0.08
Friend, age 12						
Extraversion	-0.30**	0.15	0.18	-0.18	0.15	0.28**
Neuroticism	0.17	0.12	0.04	0.30**	-0.05	-0.15
Agreeableness	-0.04	-0.31**	0.09	-0.04	-0.01	-0.05
Conscientiousness	0.04	-0.03	0.01	-0.03	0.09	-0.13
Culture	-0.17	0.05	0.28**	-0.02	0.35***	0.04

n Varied between 83 and 124 (see Method section). Correlations refer to concurrent assessments of the Big Five and the external variables. Correlations in boldface indicate relations with a Big Five scale that remained significant ($p < 0.05$) when all other Big Five scales of the same judge were statistically controlled by hierarchical multiple regression.

(*) $p < 0.07$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

showed lower correlations with inhibition than extraversion. Thus, the correlates of the Big Five in terms of social inhibition and aggressiveness were completely consistent over the whole of childhood. Social self-esteem showed no consistent relations with the Big Five judgments.

However, a systematic shift occurred for the achievement-related variables. At age 10, IQ, deviation from expected grade, and cognitive self-esteem were specifically related to conscientiousness, which was completely consistent with the findings for early childhood. However, at age 12, both IQ and cognitive self-esteem were specifically related to culture for both types of judge, and no longer at all to conscientiousness. Because IQ correlated 0.85 between ages 10 and 12, this shift can be largely attributed to a shift in the meaning of the conscientiousness and openness/culture scales.

A curious, unexpected shift occurred for the Big Five correlates of deviation from expected grade. Whereas above-average school grade was positively related to conscientiousness at age 10, replicating the age 4–6 finding (see Table 4), it lost its associations with conscientiousness completely and became associated with neuroticism for both types of judge at age 12, but in different directions. Whereas friends judged children with above-average school grade as high in neuroticism ($r = 0.30$), parents perceived them as low in neuroticism ($r = -0.19$). The difference between the correlations was significant according to Steiger's (1980) test for correlations in the same sample, $t(79) = 3.54$, $p < 0.001$. This finding once more attested to the different perspective of parents and friends on children's neuroticism.

Table 7. Heterotypic stability of the Big Five scales over childhood

From	To age (years)		
	Age (years)	10	12
Extraversion	4–6	0.19*	0.25**
Neuroticism	4–6	0.23*	0.20*
Agreeableness	4–6	0.28**	0.34***
Conscientiousness	4–6	0.45***	0.19*
Openness	4–6	0.24**	0.24*
Extraversion	10		0.34***
Neuroticism	10		0.41***
Agreeableness	10		0.43***
Conscientiousness	10		0.47***
Culture	10		0.50***

Reported are uncorrected Pearson correlations between ages. $n = 124$ for stabilities from ages 4–6 to 10, $n = 111$ for age 4–6 to 12, $n = 102$ for age 10 to 12. The stabilities underestimate the stability of personality because the judges or the instrument was different between each of the assessments.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Stability of the Big Five judgments

Finally, the stability of the Big Five judgments was studied by correlating them between ages. It should be noted that these stability coefficients provide a lower limit to the true stability of the Big Five because of (i) unreliability, particularly of the later-assessed scores that were provided by only one judge, and (ii) heterotypicity of the assessments (between any two assessments, the type of judge, the instrument, or even both were different). Despite these problems, all stability coefficients that spanned an age-interval of up to 9 years were significant for all Big Five factors (see Table 7). The average stabilities were 0.28 (ages 4–6 to 10), 0.24 (ages 4–6 to 12), and 0.43 (ages 10–12).

DISCUSSION

This longitudinal study tested hypotheses on the validity of Big Five measures for children. We assumed that, consistently over childhood, neuroticism and low extraversion correlate with social inhibition, low agreeableness and low conscientiousness correlate with aggressiveness, and conscientiousness and openness/culture correlate with antecedents and outcomes of school achievement such as IQ, deviation from expected grade, and cognitive self-esteem.

The results fully confirmed this hypothesis for our indices of internalizing and externalizing tendencies, inhibition and aggressiveness. This was a particularly strong finding because it was replicated with two different methods at both ages 4–6 (teacher judgments were related to parental judgments and behavioural observations) and 12 (parental judgments were related to both parental and friend judgments).

The results also confirmed the hypothesis for the achievement-related variables, but with a clear shift in the validity of our indices of conscientiousness and openness/culture between ages 10 and 12. Up to age 10, conscientiousness predicted all achievement-related variables, even when all other Big Five indices were statistically controlled. At age 12, however, IQ and cognitive self-esteem were predicted by both the parental and the friend

ratings of culture, not conscientiousness. Because of the high stability of IQ over the period of 2 years where this shift occurred, we attribute the shift to a different meaning of the conscientiousness and openness/culture scales.

Indeed, inspection of the items indicated a focus on planfulness and concentration ability in the Q-sort conscientiousness scale (see Table 2), which was in line with the original scale of John et al. (most items referred to planfulness and concentration ability; see John et al., 1994), whereas the bipolar conscientiousness adjectives focused on carefulness and orderliness (see Appendix), which was in line with Ostendorf's (1990) conscientiousness factor. Although all four traits are clearly related to conscientiousness, planfulness and concentration ability appear to be more closely related to IQ and school achievement than carefulness and orderliness.

Furthermore, the Q-sort openness scale focuses on creativity (see Table 2), which was again in line with the original scale of John et al. (most items referred to creativity and imagination), whereas the bipolar culture scale focused on knowledge and intelligence (only two of the eight items were related to creativity; see the Appendix). Although all three variables are clearly related to the fifth factor of the five-factor model of personality, knowledge and intelligence are more closely related to IQ and school achievement than creativity. It should be noted that the original culture factor in the analysis by Ostendorf (1990) was more balanced with regard to creativity, but the knowledge- and intelligence-related items survived the item selection procedures better than the creativity items, probably because they were clearer in meaning for parents. Thus, the two Big Five instruments differed in their focus on subfactors of conscientiousness and openness/culture, which explains the different external correlates of conscientiousness and openness/culture at ages 10 and 12.

These problems with the operationalization and meaning of the conscientiousness and the culture factor remind us of the jingle-jangle problem (Block, 2000). The 'jingle fallacy' (Thorndike, 1904) refers to cases where the same term (e.g. conscientiousness) is used for different entities (e.g. focus of an operationalization of conscientiousness on planfulness versus orderliness). In contrast, the 'jangle fallacy' (Kelley, 1927) refers to cases where different terms (e.g. intellect, culture, and openness) refer to the same entity (e.g. the same factor of the FFM or even the same scale). Both have to be avoided in adequate interpretations of empirical findings, but sometimes it is difficult to do so.

Contrary to our hypotheses, deviation from expected grade became associated with neuroticism instead of conscientiousness or culture, but in a different direction for the parental and friend judgments. Whereas parents judged children who were late schoolers or had to repeat a grade as higher in neuroticism, friends judged the same children lower in neuroticism. A *post hoc*, somewhat speculative interpretation refers to the Bavarian system of streamlining children into three different types of school, which occurs after grade 4, i.e. between 10 and 12 years of age.

First, children's success or failure in this major school transition may have induced a generalized positive or negative bias in their parents' perception of them that would both explain the significant correlations of the parental neuroticism scale with all other parental Big Five scales (see Table 5) and its positive correlation with below-average school grade.

Second, many friends of the children in the highest school track (the Gymnasium) continued in a different, lower track (because high acquaintance was the criterion for nominating a friend, and children often nominated long-term friends; unfortunately, we did not assess the school track of the friend judges). The different peer culture in these tracks and a cultural stereotype that children of the highest track are often awkward 'nerds'

might have contributed to the impression of these friends from lower tracks that the high-track friends are more neurotic.

Our interpretation of both the parental and the friend neuroticism correlation refers to a reference group effect induced by the differential streamlining of the children after age 10 into high-, medium-, and low-achieving children. Other streamlining effects are well established in the German educational–psychological literature. Repeatedly found has been, for example, a mean decrease in cognitive self-esteem in the children who advance to the Gymnasium after age 10 because they compare themselves with children of high ability in the Gymnasium (Asendorpf & van Aken, 1994; Schwarzer et al., 1982). If it is true that such reference group effects affect not only self-esteem measures, but also neuroticism judgments by others, such a result warns us to take the fact seriously that a substantial portion of the variance in Big Five judgments, particularly for the less-observable neuroticism factor, is in the eye of the beholders rather than in the judged persons.

All in all, then, the present study shows substantial validity for Big Five judgments of children, not only in terms of cross-judge consistencies, but also in terms of observed behaviour, IQ, and important school-related outcomes. A clear continuity of the external correlates of the Big Five judgments was found, despite the fact that different judges and instruments were used for different ages. Last but not least, the findings also suggest that Big Five judgments may reflect perceiver effects such as reference group effects.

ACKNOWLEDGEMENTS

We wish to thank the rest of the LOGIC crew for many years of cooperation, and Iain Glen for comments and stylistic corrections.

APPENDIX: BIPOLAR ADJECTIVE BIG FIVE SCALE ITEMS FOR CHILDREN

Extraversion	Neuroticism	Agreeableness	Conscientiousness	Culture
sociable–withdrawn (R)	helpless–self-assured (R)	touchy–good-natured	careless–thorough	uneducated–knowledgeable
withdrawn–gregarious	balanced–tense	stubborn–gentle	careful–careless (R)	intelligent–unintelligent (R)
unsociable–sociable	insecure–self-assertive (R)	inconsiderate–soft-hearted	negligent–careful	uninterested–interested
talkative–quiet (R)	tense–relaxed (R)	peaceful–quarrelsome (R)	orderly–disorderly (R)	astute–narrow-minded (R)
open–withholding (R)	relaxed–oversensitive	warm–cold (R)	sloppy–conscientious	informed–ignorant (R)
inhibited–spontaneous	self-assertive–vulnerable	conciliatory–revengeful (R)	lazy–diligent	stupid–intelligent
silent–talkative	nervous–relaxed (R)	kind–unfriendly (R)	imprecise–picky	creative–simple (R)
lively–sober (R)	worried–calm (R)	rigid–tolerant	serious–rash (R)	unimaginative–imaginative
0.90/0.87	0.83/0.78	0.84/0.85	0.91/0.86	0.86/0.82

Presented are English translations of the items, ordered by their corrected item–scale correlations for the parental judgments at age 12, and their reliabilities for parents/friends (Cronbach's α). (R) = item is reversed.

REFERENCES

- Achenbach, T. M., Howell, C. T., Quay, H. C., & Conners, C. K. (1991). National survey of problems and competencies among four- to sixteen-year-olds. *Monographs of the Society for Research in Child Development*, 56 (3, Serial No. 225).
- Asendorpf, J. B. (1990). Development of inhibition during childhood: evidence for situational specificity and a two-factor model. *Developmental Psychology*, 26, 721–730.
- Asendorpf, J. B., Borkenau, P., Ostendorf, F., & van Aken, M. A. G. (2001). Carving personality description at its joints: confirmation of three replicable personality prototypes for both children and adults. *European Journal of Personality*, 15, 169–198.
- Asendorpf, J. B., & van Aken, M. A. G. (1993). Deutsche Versionen der Selbstkonzeptskalen von Harter (German versions of Harter's self-concept scales). *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 25, 64–86.
- Asendorpf, J. B., & van Aken, M. A. G. (1994). Traits and relationship status: stranger versus peer group inhibition and test intelligence versus peer group competence as early predictors of later self-esteem. *Child Development*, 65, 1786–1798.
- Asendorpf, J. B., & van Aken, M. A. G. (1999). Resilient, overcontrolled, and undercontrolled personality prototypes in childhood: replicability, predictive power, and the trait-type issue. *Journal of Personality and Social Psychology*, 77, 815–832.
- Block, J. (2000). Three tasks for personality psychology. In L. R. Bergman, R. B. Cairns, L.-G. Nilsson, & L. Nystedt (Eds.), *Developmental science and the holistic approach* (pp. 155–164). Mahwah, NJ: Erlbaum.
- Block, J. H., & Block, J. (1980). The role of ego-control and ego-resiliency in the organization of behavior. In W. A. Collins (Ed.), *Minnesota Symposium on Child Psychology* (Vol. 13, pp. 39–101). Hillsdale, NJ: Erlbaum.
- Burgemeister, B., Blum, L., & Lorge, J. (1972). *Columbia Mental Maturity Scale*. New York: Harcourt Brace Jovanovich.
- Digman, J. M. (1989). Five robust trait dimensions: development, stability, and utility. *Journal of Personality*, 57, 195–214.
- Eggert, D. (Ed.). (1978). *Hannover-Wechsler Intelligenztest für das Vorschulalter (HAWIVA) [Wechsler intelligence scale for preschool children]*. Bern: Huber.
- Ehrler, D. J., Evans, J. G., & McGhee, R. L. (1999). Extending the Big Five theory into childhood: a preliminary investigation into the relationship between Big Five personality traits and behavior problems in children. *Psychology in the Schools*, 36, 451–458.
- Goedhart, A. W., Treffers, P. D., & Kohnstamm, G. A. (1994). The Five-Factor Model in child psychiatry: parents' free personality descriptions of their children. In C. F. Halverson, G. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 399–412). Hillsdale, NJ: Erlbaum.
- Goldberg, L. R. (1990). An alternative 'description of personality': the Big-Five factor structure. *Journal of Personality and Social Psychology*, 59, 1216–1229.
- Götttert, R., & Asendorpf, J. (1989). Eine deutsche Version des California-Child-Q-Sort, Kurzform [A German short version of the California Child Q-Set]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 21, 70–82.
- Graziano, W. G., & Ward, D. (1992). Probing the Big Five in adolescence: personality and adjustment during a developmental transition. *Journal of Personality*, 60, 425–439.
- Harter, S. (1985). *Manual for the Self-Perception Profile for Children* (Technical Report). Denver, CO: University of Denver.
- Huey, S., & Weisz, J. R. (1997). Ego control, ego resiliency, and the Five-Factor Model as predictors of behavioral and emotional problems in clinic-referred children and adolescents. *Journal of Abnormal Psychology*, 106, 404–415.
- Jensen-Campbell, L. A., & Graziano, W. G. (2001). Agreeableness as a moderator of interpersonal conflict. *Journal of Personality*, 69, 323–362.
- John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: a historical review of trait taxonomic research. *European Journal of Personality*, 2, 171–203.
- 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., & Srivastava, S. (1999). The Big Five trait taxonomy: history, measurement and theoretical perspectives. In L. A. Pervin, & O. P. John (Eds.), *Handbook of personality: Theory and Research* (2nd ed., pp. 102–138). New York: Guilford.
- Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The Big Five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*, *52*, 621–652.
- Kelley, T. L. (1927). *Interpretation of educational measurements*. Yonkers, NY: World Book.
- Kohnstamm, G. A., Halverson, C. F., Jr., Mervielde, I., & Havill, V. (Eds.). (1998). *Parental descriptions of child psychology: Developmental antecedents of the Big Five?* Mahwah, NJ: Erlbaum.
- Lay, C., Kovacs, A., & Danto, D. (1998). The relation of trait procrastination to the Big-Five factor conscientiousness: an assessment with primary-junior school children based on self-report scales. *Personality and Individual Differences*, *25*, 187–193.
- McCrae, R. R., & Costa, P. T., Jr. (1990). *Personality in adulthood*. New York: Guilford.
- Mervielde, I., & Asendorpf, J. B. (2000). Variable-centred and person-centred approaches to childhood personality. In S. E. Hampson (Ed.), *Advances in Personality Psychology* (Vol. 1, pp. 37–76). Hove, U.K.: Psychology Press.
- Mervielde, I., Buyst, V., & De Fruyt, F. (1995). The validity of the Big Five as a model for teachers' ratings of individual differences among children aged 4 to 12. *Personality and Individual Differences*, *18*, 525–534.
- Mervielde, I., & De Fruyt, F. (2000). The Big Five personality factors as a model for the structure of children's peer nominations. *European Journal of Personality*, *14*, 91–106.
- Ostendorf, F. (1990). *Sprache und Persönlichkeitsstruktur: Zur Validität des Fünf-Faktoren-Modells der Persönlichkeit*. [Language and personality structure: Validity of the five-factor model of personality.] Regensburg: Roderer.
- Rubin, K. H., & Asendorpf, J. B. (Eds.). (1993). *Social withdrawal, inhibition, and shyness in childhood*. Hillsdale, NJ: Erlbaum.
- Scholte, R. H. J., van Aken, M. A. G., & Van Lieshout, C. F. M. (1997). Adolescent personality factors in self-ratings and peer nominations and their prediction of peer acceptance and peer rejection. *Journal of Personality Assessment*, *69*, 534–554.
- Schwarzer, R., Lange, B., & Jerusalem, M. (1982). Selbstkonzeptentwicklung nach einem Bezugsgruppenwechsel [Self-concept shift after a change of the reference group]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *14*, 125–140.
- Shiner, R. (2000). Linking childhood personality with adaptation: evidence for continuity and change across time into late adolescence. *Journal of Personality and Social Psychology*, *78*, 310–325.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245–251.
- Tewes, U. (1983). *HAWIK-R: Hamburg-Wechsler Intelligenztest für Kinder (Revision 1983)* [Hamburg-Wechsler intelligence test for children]. Bern: Huber.
- Thorndike, E. L. (1904). *An introduction to the theory of mental and social measurements*. New York: Columbia University, Teachers College.
- van Aken, M. A. G., & Heutink, C. (1998). Persoonlijkheid tijdens de basisschoolleeftijd als voorspeller van anti-sociaal gedrag in de late adolescentie [Personality in elementary school as a precursor of antisocial behavior in late adolescence]. In M. Junger, & J. Dronkers (Eds.), *Daders, slachtoffers en andere tegenslag*. *Mens & Maatschappij*, *73*, 109–132.
- 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, G. A., Jr., Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 293–318). Hillsdale, NJ: Erlbaum.
- Victor, J. B. (1994). The Five-Factor Model applied to individual differences in school behavior. In C. F. Halverson, G. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 355–366). Hillsdale, NJ: Erlbaum.
- Weinert, F. E., & Schneider, W. (Eds.). (1999). *Individual development from 3 to 12: Findings from the Munich Longitudinal Study*. Cambridge: Cambridge University Press.
- Weiß, R., & Osterland, J. (1979). *Grundintelligenztest CFT [Culture Fair Intelligence Test]*. Braunschweig: Westermann.