Psychological Assessment
2012, Vol. 24, No. 1, 173–186

1040-3590/12/$12.00 DOI: 10.1037/a0025265

Factorial Structure and Age-Related Psychometrics of the MIDUS Personality Adjective Items Across the Life Span

Daniel Zimprich
University of Erlangen–Nuremberg

Mathias Allemand
University of Zurich

Margie E. Lachman
Brandeis University

The present study addresses issues of measurement invariance and comparability of factor parameters of Big Five personality adjective items across age. Data from the Midlife in the United States (MIDUS) survey were used to investigate age-related developmental psychometrics of the MIDUS personality adjective items in 2 large cross-sectional samples (exploratory sample: $N = 862$; analysis sample: $N = 3{,}000$). After having established and replicated a comprehensive 5-factor structure of the measure, increasing levels of measurement invariance were tested across 10 age groups. Results indicate that the measure demonstrates strict measurement invariance in terms of number of factors and factor loadings. Also, we found that factor variances and covariances were equal across age groups. By contrast, a number of age-related factor mean differences emerged. The practical implications of these results are discussed, and future research is suggested.

Keywords: personality traits, five-factor structure, psychometrics, measurement invariance, age differences

Five broad traits (neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness) have been proposed to summarize individual differences in relatively enduring patterns of thoughts, feelings, and behaviors (John & Srivastava, 1999; McCrae & Costa, 2003). Different approaches are suggested to measure individual differences in these Big Five personality traits. Apart from questionnaire formats consisting of sentences or brief behavioral descriptions (e.g., Costa & McCrae, 1992; John, Donahue, & Kentle, 1991), the use of adjective items is common to measure personality traits (e.g., Goldberg, 1992; John, 1990; Trapnell & Wiggins, 1990). Such a personality adjective item measure is also used in the Midlife in the United States (MIDUS) survey (e.g., Brim, Ryff, & Kessler, 2004) to capture the five personality traits (Lachman & Weaver, 1997). The personality adjective items are part of a set of quantitative measures developed for the MIDUS longitudinal survey that consists of a number of modules addressing biomedical, psychological, and social aspects of adult development. Lachman and Weaver (1997) developed the brief adjective item measure on the basis of existing adjective measures of the Big Five (e.g., Goldberg, 1992; John, 1990; Trapnell & Wiggins, 1990). They selected adjectives that were most consistently used across adjective lists and inventories and showed the highest factor loadings or item-to-total correlations. The scale construction was conducted in 1994 with a U.S. probability sample of 1,000 respondents ages 30 to 70. Each scale initially included seven to 10 adjectives. Scale scores were regressed on the items to determine which items would be entered in the model to account for a minimum of 90% of the total scale variance (for more details, see Lachman & Weaver, 1997; Prenda & Lachman, 2001).

In total, there are 25 adjective items in MIDUS that are hypothesized to measure the Big Five. However, the assignment of items to factors is not balanced. As can be seen from Figure 1, Neuroticism, for example, is supposed to be measured by four items, whereas seven items are hypothesized to tap Openness to Experience. Previous studies using the MIDUS personality adjective items particularly relied on the manifest sum or mean scores of the five scales (e.g., Mroczek & Kolarz, 1998; Plaut, Markus, & Lachman, 2002; Prenda & Lachman, 2001). For example, Mroczek and Kolarz (1998) reported substantial convergent validity relations between extraversion and positive affect and neuroticism and negative affect, respectively. To complement previous research using this measure, the present study aimed at modeling the MIDUS personality traits on the latent level. Moreover, this study is the first study to report systematic factorial structure and age-related psychometrics of this personality measure. In a first series of analyses, we set out to test a number of factor models on these 25 items in an exploratory sample to establish

This article was published Online First September 12, 2011.

Daniel Zimprich, Institute of Psychogerontology, University of Erlangen–Nuremberg, Erlangen, Germany; Mathias Allemand, Department of Psychology, University of Zurich, Zurich, Switzerland; Margie E. Lachman, Department of Psychology, Brandeis University.

Parts of the preparation of this article were supported by funding from the Midlife in the United States (MIDUS) Pilot Grant Competition, National Institute on Aging, to Mathias Allemand and by a grant from the Swiss National Science Foundation (SNSF-100014-120625). In addition, this research was supported by a grant from the National Institute on Aging (P01-AG020166) to conduct a longitudinal follow-up of the MIDUS investigation. The original study was supported by the John D. and Catherine T. MacArthur Foundation Research Network on Successful Midlife Development.

Correspondence concerning this article should be addressed to Mathias Allemand, University of Zurich, Department of Psychology, Gerontopsychology, Binzmühlestrasse 14/24, CH-8050 Zurich, Switzerland, or Daniel Zimprich, University of Erlangen–Nuremberg, Institute of Psychogerontology, Nägelsbachstrasse 25, D-91052 Erlangen, Germany. E-mail: m.allemand@psychologie.uzh.ch or daniel.zimprich@geronto.uni-erlangen.de

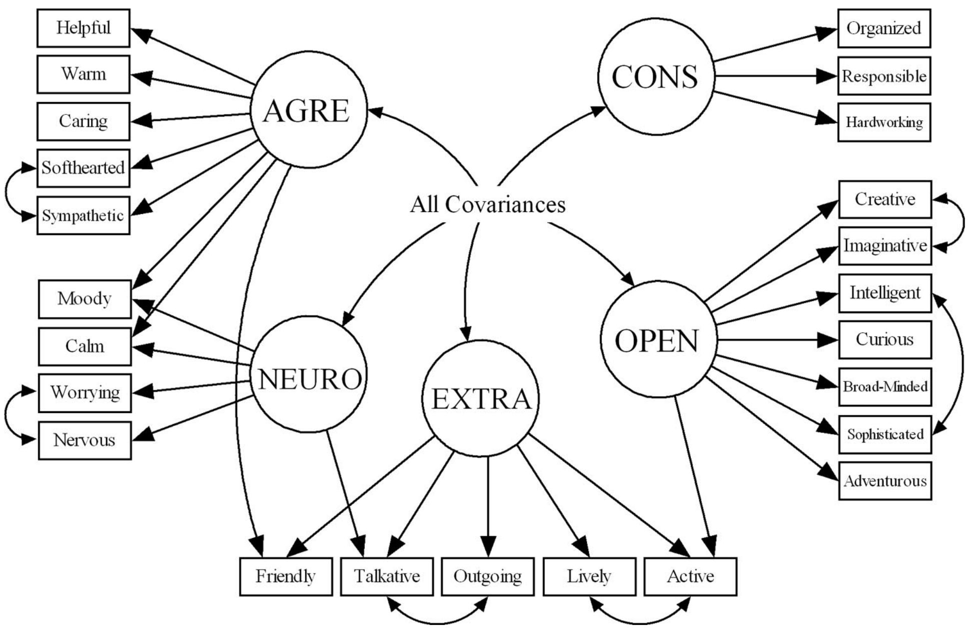

Figure 1. Factor model of personality adjective items. Note that the Conscientiousness item *careless* is excluded (see the results). Model fit in the exploratory sample (siblings) was CFI = 0.973, RMSEA = 0.062; model fit in the analysis sample was CFI = 0.976, RMSEA = 0.057. NEURO = Neuroticism; EXTRA = Extraversion; OPEN = Openness to Experience; AGRE = Agreeableness; CONS = Conscientiousness.

a comprehensive and replicable Big Five factor structure based on the a priori Big Five factor model. In a second series of analyses, we replicated the factor structure in a large and representative analysis sample. In a third series, we examined age-related differences in factor variances, covariances, and means of the personality traits. The main contribution of the present study is to report substantive findings in the context of a strictly age-invariant measurement of personality traits.

Measurement Invariance (MI)

Beyond the establishment of a comprehensive and replicable factor structure of the adjectival personality measures, an important issue in developmental psychometrics touches the question of whether psychological constructs are comparable across different age groups or across measurement occasions. Frequently, in developmental studies it is implicitly assumed that the measurement process of constructs is similar across age. However, there may be age differences in the conceptual frame of reference in interpreting or reacting to a given item of a questionnaire or to stimulus material in experimental studies, thus altering the way the latent construct underlying the item or stimulus is measured. For example, an item such as “I love to wear eccentric or eye-catching clothes” may have a different meaning to different age groups. It might be less indicative of the trait Openness to Experience in young adults as compared with older age cohorts (e.g., the hippie generation). This would indicate that this item does not operate similarly across age groups. Therefore, to ensure that the measures of a construct function equivalently for each group—in the sense that the construct operates in affecting the reactions to items or stimuli in the same way—MI has to be established (Horn & McArdle, 1992).

Borrowing from Meredith’s (1993) terminology, one might distinguish four forms of MI: configural invariance, weak MI, strong MI, and strict MI: (a) Configural invariance ensures that the dimensionality of the measured construct is equivalent across ages. If configural invariance is found, a test for higher levels of MI would follow, and the configural model would be used as a baseline for comparing the fit of more constrained models. (b) Weak MI requires that factor loading matrices be fully invariant across age groups. This form of MI ensures that comparisons of factor variances and covariances across age groups can be made. (c) Strong MI requires that factor loading matrices and intercepts of the manifest indicators be invariant across age groups, which allows for comparing factor means. (d) Strict MI requires that, in addition, unique variances be invariant across age groups. Different degrees of MI can be examined through multiple-group confirmatory factor models with increasingly strict across-group restrictions on parameters (e.g., Martin & Zimprich, 2005; Zimprich, Allemand, & Hornung, 2006).[1]

[1] There are at least two other methods for assessing MI. A typical procedure in personality trait research is to perform exploratory factor analyses or principal component analyses in different groups. To compare the factorial structure of a psychological construct across groups, Procrustean rotation is used, which rotates factors to optimal agreement resulting in a congruence coefficient (cf. McCrae, Terracciano, & Khoury, 2007). The use of item response theory techniques provides another method for assessing MI across groups. Item response theory uses an explicit measurement model to represent the relationship between observed behaviors (i.e., the item response) and the construct to be quantified. Differences in these relationships between groups are termed *differential item functioning* (cf. Morizot, Ainsworth, & Reise, 2007; Reise, Widaman, & Pugh, 1993).

Although there has been a considerable amount of research on age differences and age-related changes in the Big Five personality traits (cf. McCrae & Costa, 2003; Mroczek & Little, 2006; Roberts & Mroczek, 2008), few studies have rigorously tested for MI in trait measures across age groups or over time by utilizing confirmatory factor analytic techniques and testing for different forms of MI. Moreover, those studies particularly used questionnaire measures of the Big Five instead of adjectival markers. For example, Small, Hertzog, Hultsch, and Dixon (2003) established weak MI of the NEO Personality Inventory (NEO-PI) across a 6-year longitudinal period in older adults. Likewise, Morizot and Le Blanc (2003) found partial weak MI (i.e., the majority of factor loadings remained invariant) of personality scales across two age groups and across time. Allemand, Zimprich, and Hertzog (2007) demonstrated that, both cross-sectionally and longitudinally, strict MI of NEO Five Factor Inventory (NEO-FFI) item parcels held in two adult samples (i.e., middle-aged and older adults) followed across 4 years. Recently, in a large and representative cross-sectional sample of Dutch people ranging in age from 16 to 91 years, Allemand, Zimprich, and Hendriks (2008) established strict MI of Five-Factor Personality Inventory item parcels for six age groups. Instead of testing MI directly at the item-level, in both studies we used aggregate-level indicators (parcels) as the unit of analysis.

Applying the method of congruency coefficients (see footnote 1), other researchers reported MI similar to weak MI for the NEO-FFI (Allik, Laidra, Realo, & Pullmann, 2004). For example, Srivastava, John, Gosling, and Potter (2003) found a large amount of invariance in the pattern of factor loadings of the Big Five Inventory across four age groups. They reported an average congruency coefficient across age groups of .99, reflecting a high degree of similarity of factors. Likewise, Lang, Lüdtke, and Asendorpf (2001) found an invariant factor structure in terms of congruency coefficients of the German version of the Big Five Inventory across three age cohorts (i.e., young, middle-aged, and old adults). To summarize, some limited evidence has been found for MI of personality questionnaire measures across age.

In contrast to studies using questionnaire measures of the Big Five personality traits, such as the NEO-FFI or the Big Five Inventory (e.g., Allemand et al., 2007; Small et al., 2003; Srivastava et al., 2003), in the present study we used personality adjective items. Despite the benefits of adjective items (e.g., adjective items are relatively pure indicators of the factor they are supposed to mark, and they are quick to administer), this approach to measuring personality has been criticized for lacking context and specificity (Briggs, 1992). Definitional ambiguities may lead different groups of respondents to rate the same items differently. Indeed, Zimprich, Allemand, and Huber (2008), for example, recently found that of 20 bipolar adjective items designated to measure the Big Five, one item (*vulnerable–robust*) appeared to be biased, because it apparently elicited a differential meaning of its content in different age groups. It seemed to be interpreted or understood differently by older adults compared with younger and middle-aged adults. Therefore, before investigating age differences or age-related changes in personality as measured by adjectival markers, it is important to demonstrate that these personality adjective items function equivalently across age groups in the sense of being measurement invariant.

Age-Related Differences in Variances, Covariances, and Means

Previous studies have shown that after establishing strict MI, a number of age differences or age-related changes in the mean level of the Big Five personality traits emerged (e.g., Allemand et al., 2007; Allemand, Zimprich, & Hendriks, 2008). Very few studies have rigorously tested for personality factor variance, covariance, and mean differences across age groups or over time after having established MI. For example, Small et al. (2003) found that the Big Five personality factor variances were equal across a 6-year period in a sample of older adults, implying stability in interindividual differences over time. In addition, Allemand et al. (2007) reported that cross-sectionally, but not longitudinally, the Openness to Experience variance in middle-aged participants was significantly larger than in older participants at two measurement occasions across 4 years. Recently, in a large and representative Dutch sample, Allemand, Zimprich, and Hendriks (2008) found that personality factor variances were cross-sectionally equal across six age groups. Regarding age differences in factor covariances, Small et al., for example, found personality factor covariances to be equal longitudinally in older adults. On the basis of strict MI, Allemand et al. (2007) demonstrated invariant covariation patterns cross-group and cross-time in a sample of middle-aged and older adults, indicating that the five-factor personality covariance structure was equivalent. Likewise, Allemand, Zimprich, and Hendriks (2008) reported equal personality factor covariances in six age groups across the adult life span. These findings are in line with Costa and McCrae's (1997) conclusion that cross-sectionally the Big Five personality structure seems to be invariant at different ages. By contrast, other studies reported nonequivalence of personality structure across age groups. For example, there is empirical evidence suggesting that personality structure tends to emerge in late childhood and become clarified in adolescence, implying that the personality trait structure of children becomes differentiated, that is, less correlated, as they age (Allik et al., 2004; del Barrio, Carrasco, & Holgado, 2006; Measelle, John, Ablow, Cowan, & Cowan, 2005). Moreover, there are also studies that raise questions concerning the comparability of personality structure in old age (e.g., Allemand, Zimprich, & Martin, 2008; Mroczek, Ozer, Spiro, & Kaiser, 1998), implying that structural relations might change.

Although there is a large literature on mean-level personality trait development across the entire life span (e.g., Allemand, Gomez, & Jackson, 2010; Costa, Herbst, McCrae, & Siegler, 2000; Donnellan & Lucas, 2008; McCrae et al., 1999; Mroczek & Spiro, 2003; Roberts, Robins, Caspi, & Trzesniewski, 2003; Roberts, Walton, & Viechtbauer, 2006; Terracciano, McCrae, Brant, & Costa, 2005), only a few studies have investigated age differences and age-related changes after having established MI. For example, after establishing weak MI in the form of congruency coefficients Srivastava et al. (2003) reported that Conscientiousness and Agreeableness increased throughout early and middle adulthood at varying rates. After establishing strict MI, Allemand et al. (2007) cross-sectionally found that at the first measurement occasion, middle-aged participants were, on average, more extraverted, more open to experience, and less agreeable than older adults. Longitudinally, in both age groups, an average decline in Neuroticism was observed. Similarly, on the basis of strict MI, Allemand, Zimprich, and Hendriks (2008) found that older adults were, on average, more agreeable and especially more conscientious than middle-aged and younger adults. Together, these findings are in line with the general picture of age differences and age-related changes in personality traits at the mean level, showing that, on average, people become more conscientious and less neurotic through midlife and more agreeable in old age (cf. Roberts et al., 2003, 2006).

The Present Study

In the present study, we followed three broad objectives. First, we aimed at establishing a comprehensive and replicable factor structure model of the MIDUS 25 personality adjective items intended to measure the Big Five personality traits. The structural analyses were based on the a priori model with five latent factors (see Lachman & Weaver, 1997). For this purpose, we conducted analyses in two large samples from the MIDUS survey—in an exploratory and an analysis sample. Second, after having arrived at a comprehensive five-factor personality structure model in the exploratory sample, we cross-validated the factorial structure and investigated the amount of MI of the personality adjective items across 10 age groups in the analysis sample. Specifically, we tested assumptions about increasing levels of MI to show that the adjective items function equivalently across different ages. Finally, we examined variances, covariances, and means of the five personality factors across 10 age groups.

Method

Sample and Procedure

We used data collected in 1995 from the MIDUS survey (for a review of the study, see Brim et al., 2004) conducted by the John D. and Catherine T. MacArthur Foundation's Research Network on Successful Midlife Development. The primary objective of the MIDUS was to investigate the role of behavioral, psychological, social, biological, and neurological factors in understanding age-related differences in physical and psychological health and social responsibility. MIDUS is a national probability sample, drawn with random-digit dialing (RDD) procedures. The sample consists of English-speaking, noninstitutionalized adults ages 25 to 74 years who reside in the 48 contiguous states and whose household includes at least one telephone. The first stage of the multistage sampling design selected households with equal probability through telephone numbers. Disproportionate stratified sampling was used at the second stage, with the largest number of participants being selected in the 40 to 60 age range. The sample was stratified by age and gender, with oversampling of older people and of men. Adults who agreed to participate were administered a computer-assisted telephone interview lasting 45 min on average and were then mailed a two-part self-administered questionnaire requiring about an hour and a half to complete. All participants were offered \$20 and a copy of the final study monograph as incentives for participation. The response rate was 71% with a sample size of N = 3,487 respondents. To deal with occasional missing values in the variables of interest, we included only those participants for whom data for at least 13 of the 25 items designated to measure personality were available. The rationale behind this criterion was that for every person, more data should be available than missing. Applying this criterion resulted in a sample size of N = 3,000. Of the possible 75,000 data points (sample size of 3,000 × 25 items), 287 were missing, implying a missing data proportion of 0.4%. For the present study, we divided the sample into 10 age groups comprising equivalent age bands of 5 years (see Table 1).[2]

Exploratory Sample

We used another existing data set from the MIDUS survey as an exploratory sample to establish a comprehensive factorial structure of the Big Five personality trait measure prior to the confirmatory analyses. Siblings from the RDD national sample respondents were identified (N = 950) and recruited with the cooperation of the national sample respondents, who were asked to provide interviewers with their contact information and to communicate with their siblings about participation prior to the time a recruiter made the contact attempt. Using the same inclusion criteria as described earlier, we arrived at a sample size of N = 862, with 0.5% of the data missing. Table 1 reports the demographic characteristics of the sibling sample.

Measures

The Big Five personality traits were measured using 25 self-descriptive adjective items (Lachman & Weaver, 1997; Prenda & Lachman, 2001). Each of the five personality traits was assessed with between four and seven adjectives: Neuroticism (*moody, worrying, nervous, calm* [reverse scored]), Extraversion (*outgoing, friendly, lively, active, talkative*), Openness to Experience (*creative, imaginative, intelligent, curious, broadminded, sophisticated, adventurous*), Agreeableness (*helpful, warm, caring, softhearted, sympathetic*), and Conscientiousness (*organized, responsible, hardworking, careless* [reverse scored]). Participants indicated how well each adjective describes them on a 4-point scale, ranging from *a lot* (1) to *not at all* (4). For the purpose of calculating Cronbach's alphas, ratings were reverse coded when necessary, so that higher scores reflect higher standing on each trait. Alpha internal consistency coefficients obtained in the exploratory sample and the analysis sample, respectively, were .75 and .74 for Neuroticism, .75 and .78 for Extraversion, .78 and .77 for Openness to Experience, .83 and .80 for Agreeableness, and .53 and .58 for Conscientiousness.

Analytic Procedures

To establish and validate a comprehensive factorial structure model of the personality adjective items, we conducted confirmatory factor analyses based on the a priori model as proposed by Lachman and Weaver (1997). Because the adjective items are Likert-scaled with four categories, we applied factor analysis for ordered-categorical data (B. O. Muthén, 1984) in both the exploratory and the analytic samples. Multiple-group confirmatory factor analysis of ordered-categorical variables (Millsap & Yun-Tein,

[2] Dividing our sample into five groups with a 10-year span as well as into three age groups (i.e., younger adults, middle-aged adults, and older adults) produced similar findings. Hence, the 10 age groups were retained to enable more fine-grained analyses.

Table 1
Sample Description

Characteristic	Exploratory sample (siblings)	Analysis sample
N	862	3,000
Gender (% women)	55.8	51.5
Mean age (*SD*) in years	49.5 (12.7)	47.0 (13.4)
Age range	24–75	25–74
Age groups (sample sizes)		
25–29	37	288
30–34	79	337
35–39	99	375
40–44	124	356
45–49	117	398
50–54	90	323
55–59	96	305
60–64	78	287
65–69	84	158
70–74	58	173

2004) was then utilized to assess MI across age groups (cf. Bollen, 1989). MI was assessed in a series of four steps (cf. Meredith, 1993; Meredith & Horn, 2001). First, configural invariance of each of the Big Five trait scales was examined across age groups. Next, factor loadings were constrained to be equal across age groups, thereby testing weak MI. Afterward, thresholds of the manifest indicators (i.e., the personality adjective items) were constrained to be equal across age groups. This step imposed strong MI across different ages. Finally, strict MI involved additional constraints, namely that residual variances and, subsequently, residual covariances were required to be equivalent across age groups (for details, see Allemand, Zimprich, & Hendriks, 2008; Zimprich et al., 2006). In addition to comparisons of more constrained models with the less constrained models, we also reported model comparisons with the baseline model (i.e., the configural invariance model). After establishing strict MI, factor variances, factor covariances, and factor means were compared across age groups.

All analyses were conducted with Mplus Version 3.0, accounting for the presence of missing data by the full information maximum likelihood algorithm and applying the WLSM estimator, which uses the Satorra–Bentler scale correction; that is, it adjusts the chi-square values for their mean (L. K. Muthén & Muthén, 2004; Satorra & Bentler, 2001). The goodness-of-fit of the models was evaluated with the comparative fit index (CFI) and the root-mean-square error of approximation (RMSEA).[3] CFI values above .95 denote a well-fitting model. RMSEA values less than .06 are indicative of a good model fit, whereas values larger than .08 are considered a nonacceptable fit (cf. Browne & Cudeck, 1993; Hu & Bentler, 1999). As an additional criterion for model evaluation, the Satorra–Bentler adjusted chi-square test statistic is reported (Satorra & Bentler, 2001) and denoted as χ^2_{S-B}. In comparing the relative fit of nested models, we used the Satorra–Bentler adjusted chi-square difference test, which is denoted as $\Delta\chi^2_{S-B}$. To compute a chi-square difference test, the difference of the chi-square values of the two models in question is taken as well as the difference of the degrees of freedom. However, in conjunction with Satorra–Bentler adjusted chi-square values, this procedure is slightly more complicated. First, for each of the two models in question, a scaling correction factor is computed, which is defined as the regular chi-square value divided by the Satorra–Bentler adjusted chi-square value. Next, the scaling correction for the chi-square difference is computed, which is then used to weigh the regular chi-square difference (see the Appendix for a worked example).

Given the large sample size in this study, the alpha level was set to 1% to evaluate statistical significance. Moreover, we mainly relied on the relative model fit (i.e., CFI and RMSEA) to evaluate model fit because chi-square tests become overly sensitive with increasing sample size and a large number of degrees of freedom. According to Cheung and Rensvold (1999), a change in the CFI of less than .01 amounts to a trivial difference in model fit. If the change in the CFI is less than .01, then the set of cross-group constraints is tenable and one can proceed with making further comparisons. As a measure of effect size for mean differences, we report Cohen's *d* (Cohen, 1988, p. 20).

Results

Factorial Structure in the Exploratory Sample

Structural equation modeling started with Model A, a model where each of the 25 items was specified to load on the factor it was designated to measure as described in the MIDUS manual (Lachman & Weaver, 1997). Thus, for example, the items *helpful, warm, caring, softhearted,* and *sympathetic* were specified to load on a common Agreeableness factor. As can be seen from Table 2, Model A did not achieve an acceptable fit according to the χ^2_{S-B} value. However, although both the CFI and the RMSEA were not in the acceptable range, they showed that Model A accounted for a sizeable amount of the covariances among the variables. On inspection, we found that the item *careless,* which should load on the Conscientiousness factor, had a standardized factor loading of only .11 and, hence, was virtually unrelated to the other three Conscientiousness items. Possibly, participants interpreted *careless* as heedless instead of interpreting it as sloppy. As a consequence, we decided to exclude the item *careless* from further analyses.

Next, Model B was estimated, which differed from Model A only by the exclusion of the item *careless*. As shown in Table 2, the CFI and RMSEA were almost unchanged. According to the differences between the predicted and the actual covariance matrix, there were large residual covariances between the items measuring Agreeableness and the items *moody* (Neuroticism), *calm* (Neuroticism), and *friendly* (Extraversion) that remained unaccounted for in Model B. Evidently, then, being moody is not only indicative of being neurotic but also captures the opposite of being agreeable. Similarly, to stay calm represents not only emotional stability (i.e., the converse of Neuroticism), but also aspects of Agreeableness in the sense of being tranquil. Finally, *friendly,* which was designated to measure Extraversion, apparently assesses Agreeableness, too. Subsequently, these three items were allowed to cross-load on the Agreeableness factor. After having incorporated the three additional parameters one at a time, the resulting model (Model C) still produced a statistically significant

[3] Because both the CFI and the RMSEA rely on the Satorra–Bentler adjusted chi-square values, they also reflect adjusted fit indexes.

Table 2
Model Fit in the Exploratory Sample (N = 862, Siblings)

Model	χ^2_{S-B}	*df*	SC	$\Delta\chi^2_{S-B}$	Δdf	$\Delta^B\ \chi^2_{S-B}$	$\Delta^B\ df$	CFI	RMSEA
A	3,051.49*	265	0.690	—	—	—	—	0.905	0.110
B	2,862.58*	242	0.680	—	—	—	—	0.910	0.112
C	2,023.84*	239	0.639	165.55*	3	212.54*	3	0.939	0.093
D	1,869.16*	237	0.629	64.43*	2	248.87*	5	0.944	0.089
E	1,004.85*	231	0.613	449.58*	6	637.55*	11	0.973	0.062

Note. χ^2_{S-B} = Satorra–Bentler adjusted chi-square; SC = scaling correction factor; $\Delta\chi^2_{S-B}$ = Satorra–Bentler adjusted chi-square difference to the previous model; Δdf = degrees of freedom difference; $\Delta^B\ \chi^2_{S-B}$ = Satorra–Bentler adjusted chi-square difference to Model B; $\Delta^B\ df$ = degrees of freedom difference to Model B; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; Model A = as specified in the Midlife in the United States survey manual; Model B = the item *careless* excluded; Model C = cross-loadings of *moody, calm,* and *friendly* on Agreeableness; Model D = cross-loadings of *talkative* on Neuroticism and *active* on Openness; Model E = six residual covariances (see text).
* $p < .01$.

χ^2_{S-B} value (see Table 2), but compared with the previous model, it had decreased significantly, indicating a better fit. This significantly better fit is also reflected in the CFI and RMSEA values of Model C, which indexed an improvement compared with Model B. The cross-loadings of *calm* and *friendly* on the Agreeableness factor were positive, whereas that of *moody* was negative. Importantly, the cross-loading of *friendly* on Agreeableness was slightly stronger than its loading on the Extraversion factor, showing that *friendly* should be considered a measure of Agreeableness even more than a measure of Extraversion. Because of the still unacceptable fit, we inspected the residual covariance matrix and found that the items *talkative* (Extraversion) and *active* (Extraversion) showed associations to the Neuroticism and Openness to Experience factor, respectively, that were not adequately captured by Model C. Thus, being *talkative* or eager to communicate may also reflect aspects of Neuroticism. Likewise, *active* also is characteristic of being open to new experiences in the sense of, for example, actively striving for experiences.

After having incorporated the two corresponding cross-loadings successively, Model D resulted. Table 2 shows that doing so increased model fit considerably in comparison with Model C, although it did not yet reach acceptable values. The cross-loading of *talkative* on the Neuroticism factor was positive, demonstrating that being talkative is also indicative of Neuroticism. The cross-loading of *active* on the Openness factor was also positive, implying that describing oneself as active does also have a component of Openness. Because Model D did not yet describe the associations among the 24 remaining personality items adequately, a final inspection of the residual covariance matrix revealed that between several individual items, some associations were unaccounted for in Model D. These were the items *worrying* and *nervous* (Neuroticism), *talkative* and *outgoing* (Extraversion), *lively* and *active* (Extraversion), *creative* and *imaginative* (Openness to Experience), *intelligent* and *sophisticated* (Openness to Experience), and *softhearted* and *sympathetic* (Agreeableness).

In Model E, these associations were captured by estimating the corresponding residual covariances. As can be seen from Table 2, Model E did achieve an acceptable fit according to the CFI and an almost acceptable fit as judged by the RMSEA. At the same time, fit had improved considerably compared with Model D. Also, compared with the baseline Model B, it produced a significantly better fit. On balance, we thus regarded Model E as adequately capturing the associations between the 24 items measuring the Big Five. Model E is depicted in Figure 1. Parameter estimates (standardized factor loadings and factor correlations) as based on Model E are given in Table 3. On average, 52% of variance was explained in the personality items, ranging from 21% (*organized*) to 80% (*worrying*). Apart from Neuroticism, factors were relatively strongly intercorrelated, such that they shared 36% of vari-

Table 3
Parameter Estimates (Based on Model E) in the Exploratory Sample (N = 862)

Item or factor	N	E	O	A	C
			Standardized factor loadings		
Moody	0.584			−0.098	
Calm	−0.518			0.445	
Worrying	0.895				
Nervous	0.885				
Friendly		0.404		0.442	
Talkative	0.203	0.614			
Outgoing		0.760			
Lively		0.846			
Active		0.304	0.435		
Creative			0.562		
Imaginative			0.665		
Intelligent			0.613		
Curious			0.739		
Broad-minded			0.553		
Sophisticated			0.662		
Adventurous			0.661		
Helpful				0.793	
Warm				0.858	
Caring				0.879	
Soft-hearted				0.616	
Sympathetic				0.739	
Organized					0.462
Responsible					0.842
Hardworking					0.763
			Factor correlations		
N	—				
E	*−.072*	—			
O	*−.068*	.605	—		
A	.116	.681	.527	—	
C	*−.007*	.498	.603	.646	—

Note. Parameters in italics are not statistically significant at $p < .01$. N = Neuroticism; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness.

ance, on average, ranging from 25% (Extraversion and Conscientiousness) to 46% (Extraversion and Agreeableness). The correlations between residuals were all statistically significant ($p < .01$) and amounted to .23 (*worrying* and *nervous*), .17 (*talkative* and *outgoing*), .27 (*lively* and *active*), .42 (*creative* and *imaginative*), .24 (*intelligent* and *sophisticated*), and .24 (*softhearted* and *sympathetic*).

Factorial Structure and MI in the Analysis Sample

After having established an acceptable model (Model E) of the 24 personality items in the exploratory sample, to validate this model, which was modified in an ad hoc fashion, it was re-estimated in the analysis sample. As shown in Table 4, Model E achieved an acceptable fit as judged by the CFI and the RMSEA, which even was slightly better than the corresponding fit in the exploratory sample. A congruency coefficient calculated for the standardized solution in the exploratory and the analysis sample reached .997, denoting excellent congruency (cf. MacCallum, Widaman, Zhang, & Hong, 1999). On average, 51% of variance was explained in the personality items, ranging from 24% (*organized*) to 77% (*worrying*). Like in the exploratory sample, the six correlations between residuals were statistically significant ($p < .01$) and amounted to .22 (*worrying* and *nervous*), .14 (*talkative* and *outgoing*), .26 (*lively* and *active*), .38 (*creative* and *imaginative*), .22 (*intelligent* and *sophisticated*), and .25 (*softhearted* and *sympathetic*). We regarded Model E as an adequate starting point for investigating the measurement properties of the personality items across age groups. To further asses the fit of Model E, it was estimated as Model E* in every age group separately. As can be seen from Table 4, this resulted in χ^2_{S-B} values ranging from 380 to 621. However, because of differing sample sizes in the age groups, these values are not directly comparable. By contrast, the CFI and the RMSEA may be compared across age groups, and Table 4 shows that they were in the acceptable range, although there was one age group (55–59 years) where the RMSEA exceeded .07. From this result, we concluded that Model E held in every age group separately, not only in the composite sample.

Multigroups confirmatory factor analyses started with an unconstrained model, that is, a configural invariance (CI) model with five factors of personality (Model E) that was estimated simultaneously without any parameter constraints in all 10 age groups (Model CI in Table 4). Factor variances were fixed to 1, and factor means were fixed to 0 to scale the latent variables. As shown in Table 4, the first model (Model CI) demonstrated a good fit as judged by the CFI, which was above .95. The RMSEA indicated that the model fit acceptably. Therefore, configural invariance of the five factors of personality appears to hold across 10 age groups regarding the personality adjective items. Next, in the weak MI model (Model WMI), factor loadings and cross-loadings were constrained to be equal across age groups, whereas factor variances were freely estimated in all age groups apart from the youngest group (i.e., a reference group of those ages 25–29 years). This model also evinced an acceptable fit (see Table 4). Compared with Model CI, Model WMI represented a reduction in relative fit because it had a significantly higher χ^2_{S-B} value. Whereas both CFI statistics were identical, the RMSEA had improved. Hence, one might conjecture that weak MI holds across age groups. In the strong MI model (Model SMI), the additional constraint of equal thresholds of the manifest indicators, implying strong MI, was tested. Factor means were freely estimated in all age groups except for the reference group (i.e., those ages 25–29 years). As Table 4 shows, Model SMI also achieved an acceptable fit. Compared with the former model, Model SMI did not represent a loss in model fit because the χ^2_{S-B} difference was not statistically significant. The CFI was similar to the former model statistic, whereas the RMSEA even showed an improvement. Therefore, we concluded that strong MI holds across the age groups with respect to the five personality dimensions. Finally, the assumption of strict MI was tested in a model of complete MI (Model CMI); that is, residual variances were constrained to be equal across age groups. Model CMI yielded an acceptable fit as well (see Table 4). Compared

Table 4
Model Fit in the Analysis Sample (N = 3,000)

Model	χ^2_{S-B}	*df*	SC	$\Delta\chi^2_{S-B}$	Δdf	$\Delta^{CI}\ \chi^2_{S-B}$	$\Delta^{CI}\ df$	CFI	RMSEA
E	2,514.30*	231	0.614	—	—	—	—	0.976	0.057
E*	380.52* to 621.79*	231	0.587 to 0.697	—	—	—	—	0.956 to 0.983	0.058 to 0.072
CI	5,255.58*	2310	0.620	—	—	—	—	0.972	0.065
WMI	5,461.70*	2526	0.686	350.80*	216	—	—	0.972	0.062
SMI	5,653.54*	2697	0.688	199.17*	171	577.00*	387	0.972	0.060
CMI	5,921.79*	2913	0.750	361.98*	216	947.81*	603	0.971	0.059
CMIr	5,963.75*	2967	0.751	46.50	54	1,007.20*	657	0.971	0.058
CMIrv	5,966.26*	3012	0.794	71.21*	45	1,082.09*	702	0.972	0.057
CMIrvc	5,908.13*	3102	0.938	147.22*	90	1,224.00*	792	0.973	0.055
CMIrvcm	6,247.31*	3147	0.972	269.97*	45	1,447.89*	837	0.970	0.057

Note. χ^2_{S-B} = Satorra–Bentler adjusted chi-square; SC = scaling correction factor; $\Delta\chi^2_{S-B}$ = Satorra–Bentler adjusted chi-square difference; Δdf = degrees of freedom difference; $\Delta^{CI}\ \chi^2_{S-B}$ = Satorra–Bentler adjusted chi-square difference to Model CI; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; Model E = six residual covariances (see text); Model E* = Model E modeled separately in each age group (the range of fit indexes is given); Model CI = multiple-groups model of configural invariance; Model WMI = multiple-groups model of weak measurement invariance; Model SMI = multiple-groups model of strong measurement invariance; Model CMI = multiple-groups model of strict or complete measurement invariance; Model CMIr = Model CMI plus equal residual covariances; Model CMIrv = Model CMIr plus equal factor variances; Model CMIrvc = Model CMIrv plus equal factor covariances; Model CMIrvcm = Model CMIrvc plus equal factor means.
* $p < .01$.

with Model SMI, there was a statistically significant loss of fit as indexed by the χ^2_{S-B} difference test. The improvement in RMSEA, however, suggested that the difference in model fit was not of practical importance, indicating that the hypothesis of strict MI should not be rejected. Consequently, the model of strict MI seemed to adequately capture our data.[4]

As a final model of this set of analyses, in addition to constraining residual variances to be equal across age groups, the six covariances between residuals were required to be equal in all age groups (Model CMIr). As Table 4 shows, the fit of Model CMIr was not statistically different from that of the previous model, and according to the RMSEA fit had even improved slightly. Hence, we concluded that an extended model of strict MI, where residual variances and covariances were equal, held for the data of the analysis sample.

To summarize, the tests of different degrees of MI revealed that the measurement properties of the adjective items used to operationalize the Big Five personality appear to be equal across age groups in the sense that the adjective items measure the same construct or, in the case of multiple loadings, the same constructs across the 10 age groups. In other words, because an extended model of strict MI fit the data, all age-related differences in means, variances, and covariances of the 24 personality items are attributable to differences in parameters of the latent variables underlying the items, that is, differences in factor means, factor variances, and factor covariances.

Table 5
Parameter Estimates (Based on Model CMIrvc) in the Analysis Sample (N = 3,000)

Item or factor	N	E	O	A	C
		Standardized factor loadings			
Moody	0.633			−0.149	
Calm	−0.595			0.357	
Worrying	0.873				
Nervous	0.878				
Friendly		0.462		0.411	
Talkative	0.221	0.646			
Outgoing		0.791			
Lively		0.891			
Active		0.404	0.357		
Creative			0.570		
Imaginative			0.678		
Intelligent			0.657		
Curious			0.730		
Broad-minded			0.513		
Sophisticated			0.642		
Adventurous			0.716		
Helpful				0.804	
Warm				0.850	
Caring				0.887	
Softhearted				0.573	
Sympathetic				0.675	
Organized					0.540
Responsible					0.833
Hardworking					0.771
		Factor correlations			
N					
E	−.122				
O	*−.035*	.654			
A	.162	.686	.514		
C	*.030*	.460	.547	.606	

Note. Parameters in italics are not statistically significant at $p < .01$. N = Neuroticism; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness.

Factorial Differences in the Analysis Sample

First, to test for age-related differences in the latent variables, personality factor variances were constrained to be equal to those in the reference group (i.e., 25–29 years); that is, they were constrained to be one in all age groups. The resulting model (Model CMIrv) still yielded an acceptable fit (Table 4). Albeit, compared with Model CMIr, there was a statistically significant decrease in model fit; both additional criteria for absolute model fit suggested that model fit was indistinguishable. We therefore concluded that individual differences in the Big Five personality factors as measured by the adjective markers were equally pronounced in all age groups across the adult life span. Second, factor covariances were constrained to be equal across age groups (Model CMIrvc). This model also achieved an acceptable fit (Table 4). Compared with the model of strict MI plus equal factor variances, this model did, again, represent a statistically significant decrease in model fit as judged from the chi-square differences, whereas the CFI and the RMSEA showed improvements. Therefore, we concluded that equal factor covariances could be assumed in all age groups. Parameter estimates (standardized factor loadings and factor correlations) based on Model CMIrvc are given in Table 5. Eventually, we constrained all factor means to be equal across age groups (Model CMIrvcm). Doing so, however, led to a decrement in model fit (see Table 4). Compared with the previous model fit differences, we considered this decrement to be substantial because per degree of freedom, there was a loss of Δchi-square/Δ*df* = 6 points in chi-square. For the previous chi-square differences, the Δchi-square/Δ*df* ratio had always been smaller than 2 (cf. footnote 4). Moreover, for Model CMIrvcm, for the first time both fit indexes (CFI and RMSEA) consistently indicated a loss of fit. Together, we regarded this as sufficient evidence for not only statistically significant but also substantial mean differences between age groups. To locate statistically significant factor mean differences between age groups, we calculated 84% inferential confidence intervals for each factor mean (Tryon, 2001). Results are shown in Figure 2. The finding of equal factor variances allows for interpreting factor mean differences directly in terms of Co-

[4] We tested different degrees of MI only for Model E, which represented the best fitting model. One may wonder whether strong invariance would also have held for other simpler models in particular. Although beyond the scope of the present article, as an example case, we specified Model A, which was based on the MIDUS manual, both as a configural invariance model and as a strong MI model. The corresponding Satorra–Bentler corrected chi-square values were 20,709 (*df* = 2650) and 25,815 (*df* = 3235). These numbers are easier to interpret if the chi-square values are weighed by their *df*s, resulting in 7.81 and 7.97, respectively, showing that per degree of freedom, fit became slightly worse by imposing strict MI. For comparison, the according chi-square/*df* values for Model E were 2.27 and 2.03, respectively, indicating that per degree of freedom, fit slightly increased by imposing strict MI. Note that this result should not be generalized but rather represents a snapshot. We are not aware of any systematic inquiry of this intriguing issue of MI research.

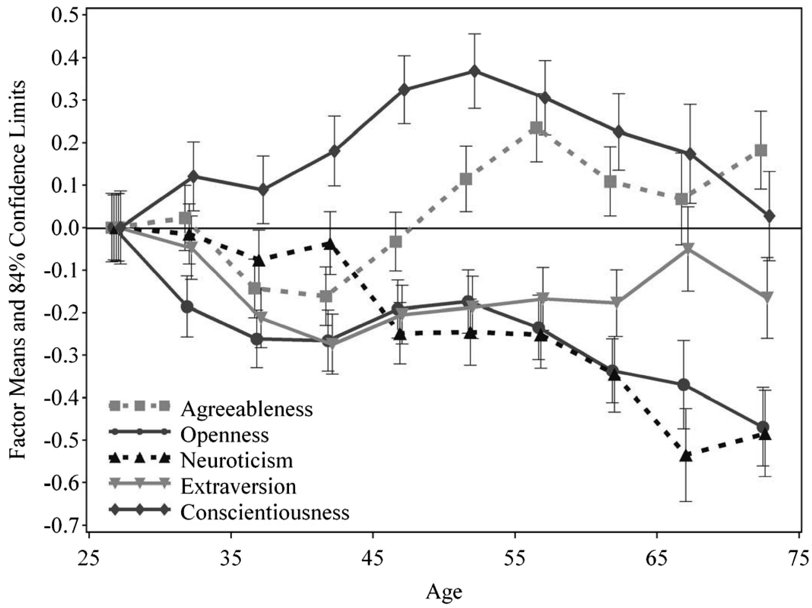

Figure 2. Factor means across age groups based on Model CMIrvc (multiple-groups model of strict or complete measurement invariance plus equal residual covariances and equal factor variances). Figure 2 is to be read as follows: If the 84% confidence interval of a factor mean in one age group overlaps with the 84% confidence interval of the corresponding factor mean in another age group, factor means are not significantly different at the 5% level. In turn, if the 84% confidence interval in one age group does not overlap with the 84% confidence interval of the corresponding factor mean in another age group, factor means should be considered as being significantly different at the 5% level. Error bars represent the 84% confidence interval.

hen's *d*. The picture that emerges with respect to the means of the Big Five personality traits may be described as follows. Neuroticism showed a decrease across age groups, with the decrease being significant in the two oldest groups as compared with participants younger than 59 years. Also, those participants ages 45 to 64 years were significantly more neurotic than the younger age groups. Extraversion exhibited a gradual decrease in the early midlife and then plateaued thereafter. Openness to Experience tended to show a decrease with age in early midlife and then in late midlife into old age. Agreeableness slightly decreased up to the early forties and then showed a gradual increase to the late fifties. Finally, Conscientiousness tended to show a curvilinear pattern of change with the highest factor mean levels in those participants ages 50 to 54. This age group significantly differed from those ages 44 downward (see Figure 2).

To summarize, these results suggest that after establishing strict MI, both stability and change mark personality traits across the adult life span. The Big Five personality traits were stable across the life span with respect to factor variances and covariances. At the same time, a number of age-related differences in the factor means of the Big Five personality traits emerged.

Discussion

The purpose of this research was to present analyses of the factorial structure and age comparative psychometric analyses of adjective items intended to measure the Big Five personality traits. Specifically, we reported substantive findings for the five personality traits in the context of strict MI, which allows for an unambiguous quantitative comparability of the MIDUS Big Five composite scores. In the present article, we followed three goals. First, we wanted to examine the factorial structure of the 25 Likert-type items included in MIDUS to measure five dimensions of personality. To do so, in the exploratory sample of 862 persons, five different models were estimated. From these models, a number of noteworthy findings emerged. The item *careless* designated to assess Conscientiousness only shared about 1% of variance with the common factor defined by the remaining three Conscientiousness items. Hence, it appears that this specific item is not a good measure of Conscientiousness. There are several reasons for this, one of which may be that *careless* was the only negatively worded Conscientiousness item. Hence, it might be that participants based their answers to this item on a certain response style (e.g., acquiescence; cf. Marsh, 1996). However, there are a number of other personality items in MIDUS that are also negatively worded but did share a substantial amount of variance with their respective factor. Thus, without further analyses—for example, by applying the random intercept factor analysis model developed by Maydeu-Olivares and Coffmann (2006)—the issue of a response style in conjunction with the *careless* item remains speculative. From a more substantive point of view, it might be that the connotations invoked by *careless* are different from those that come to one's mind by answering the remaining three Conscientiousness items used in MIDUS (*organized, responsible, hardworking*). These three items describe someone you can count on, who puts in a lot of effort and tries hard. In contrast, someone who is careless may be so, on the one hand, because of doing things in a haphazard manner without much effort. On the other hand, this person could also be someone who makes a lot of errors and mistakes, not because of low Conscientiousness, but because of poor concentration or attentional control. Thus, in the case of someone who is careless, unreliable performance may not be intentional. Because of this limitation in the original Conscientiousness items, in the second wave of the study 10 years later in 2005 (MIDUS II), a new item was added to the Conscientiousness scale. Future analysis will show whether with this new item the Conscientiousness scale is more homogeneous.

In total, five statistically significant cross-loadings emerged in the exploratory sample, showing that some items tapped not one but two of the five personality traits. This is a common finding in trait psychology and led to the development of the Abridged Big-Five Dimensional Circumplex model (AB5C; Hofstee, de Raad, & Goldberg, 1992), which explicitly recognizes and represents combinations or blends of the Big Five factors. Most cross-loadings (three) were on the Agreeableness factor, whereas Neuroticism and Openness to Experience had one cross-loading each. In all cases, these cross-loadings were interpretable based on the item content—for example, *friendly* as a designated measure of Extraversion also loaded on Agreeableness. A similar finding was reported by Goldberg (1990) with respect to the opposite pole of the trait adjective *friendly,* namely, *unfriendly.* Moreover, the adjective *calm* primarily loaded on Neuroticism and secondarily on Agreeableness. Using the Abridged Big-Five Dimensional Circumplex model, *calmness* similarly refers to both Neuroticism and Agreeableness in the International Personality Item Pool scales (Goldberg et al., 2006). Likewise, the other cross-loadings might reflect blends of the Big Five factors (Hofstee et al., 1992).

Also, in the exploratory sample we found six correlated residuals. Within the neuroticism domain, *worrying* and *being nervous* were two related adjectives that might describe a susceptibility to intrusive and anxious thoughts, feelings, and habits. Residual covariances were also estimated between two pairs of extraversion adjective items. The residual between the adjectives *talkative* and *outgoing* might reflect the fact that both items describe individual differences in gregariousness. People who are outgoing and talkative enjoy social events and the company of others. Second, the shared residual variance between *lively* and *active* reflects the fact that both items refer to individuals' energy level (cf. John & Srivastava, 1999). The Openness trait adjectives *creative* and *imaginative* may have shared common residual variance because both items measure an artistic tendency or a strong fantasy. Similarly, *sophisticated* and *intelligent* may be seen as reflecting openness with respect to intellectual material and might also reflect typical intellectual engagement (see Dellenbach & Zimprich, 2008; Mascherek & Zimprich, in press). Eventually, for the Agreeableness domain, the residual covariance between *soft-hearted* and *sympathetic* may indicate a tendency to be compliant. Importantly, these residual correlations were all interpretable and did not cross factor boundaries. Rather, they should be seen as specific factors, that is, factors that tap side aspects of the common factors and thereby lead to residual covariances. Importantly, they did not compromise the principal solution of five personality factors. In evaluating the five-factor solution, one has to keep in mind that we directly factored the 25 and 24 items, respectively, which most likely contributed to the fact that cross-loadings and covariances between residual emerged. That is, whereas idiosyncratic and unsystematic influences tend to cancel out in scale scores or in item parcels (e.g., Allemand et al., 2007), individual items are much more susceptible to such unwanted sources of variance. To summarize, a five-factor model provided a good fit for the personality adjective items in the exploratory sample, although with some deviations.

What are the practical implications of our factorial structure results for future users of the MIDUS personality adjective item measure? We have two recommendations. First, if one is interested in modeling the personality traits at the latent level, we suggest grouping the items on the basis of the modified factor structure reported in this article (i.e., including cross-loadings). Second, if future users are primarily interested in using a short instrument for measuring the Big Five traits, which is often the case in applied settings, the sum or mean scores of the five scales could also be computed on the basis of the findings from the present study. Here, however, it might be sufficient to take into account only the cross-loadings for those items with loadings of almost equal size on two factors, that is, *friendly, active,* and *calm.* These items measuring two traits could enter sum scores of the Big Five with a weight of .5 for each of the two traits.

As a second goal of our study, we aimed at examining the amount of age invariance of the five-factor solution found in the exploratory sample. To do so, we re-estimated the best-fitting model from the exploratory sample (Model E) in the analysis sample of 3,000 MIDUS RDD participants. Consistent with the results of the exploratory sample, the model showed a good fit, and congruency of the standardized solutions between both samples was excellent. Afterward, different degrees of MI of the five-factor model with respect to age were imposed by constraining factor loadings, thresholds, residual variances, and residual covariances across 10 age groups. As it turned out, there were no significant differences in these parameters across age groups, implying that the 24 personality items used in MIDUS are invariant to age as far as their measurement properties are concerned (cf. Horn & McArdle, 1992). Specifically, we established strict MI of the items with respect to age (Meredith, 1993). Such a finding has three consequences, the first of which directly touches the comparison of factor parameters (variances, covariances, and means), which are unambiguously interpretable once at least strong invariance has been shown to hold. In comparing factor variances and covariances across age groups, we found that these were also equal, which, in conjunction with strict invariance, necessarily implies that factor correlations and reliabilities of the individual items were equal across age. Also, equal factor variances allowed for interpreting factor mean differences directly in terms of Cohen's *d*. The second consequence of strict MI with respect to age is that differences between age groups are attributable to the same sources as individual differences within age groups (Lubke, Dolan, Kelderman, & Mellenbergh, 2003). In other words, cross-sectional age differences in observed variables cannot be due to other factors than individual differences. Third, as Lubke et al. (2003) have demonstrated, strict MI with respect to age implies weak MI of the 24 personality items with respect to all selection variables related to age (e.g., different levels of physical health). Taking into account the severity of restrictions that must obtain and considering the relatively large sample size, the finding of strict factorial invariance with respect to age appears remarkable. A limitation of our model comparisons could be raised given that we mainly relied on the CFI and RMSEA to judge the degree of MI. Although this is in accordance with the fact that the chi-square test becomes excessively powerful in large samples, a shortcoming of current model fit indexes is that they do not have effect size equivalency. The problem is that the same difference between, for example, two RMSEA values reflects different effect sizes dependent on the values of the RMSEA (MacCallum, Browne, & Sugawara, 1996).

As a third goal of our study, we tested the invariance of factor variances, covariances, and means of the Big Five personality traits across age groups. These tests extend beyond MI and refer to issues of structural invariance. The estimated factor variances were equal across age groups, implying that the amount of individual differences with respect to the five personality dimensions did not change systematically with age—similar to what has been reported by Allemand, Zimprich, and Hendriks (2008). A strong interpretation of this finding is qualified by the fact that it stems from cross-sectional comparisons—albeit, elsewhere, we have demonstrated that longitudinally variances in the Big Five hardly changed in old age (Allemand et al., 2007; Allemand, Zimprich, & Martin, 2008; Small et al., 2003). Moreover, the fact that personality factors were measured with only a few ordered-categorical items might have contributed to the feasibility of finding equal factor variances because measurements were less fine graded than they would have been if we had used full questionnaires including facets scales (e.g., the NEO-PI-R; Costa & McCrae, 1992). Notwithstanding, if interpreting these results developmentally, equal factor variances across age would imply the absence of a so-called *Matthew effect* (Bast & Reitsma, 1997). The Matthew effect denotes the phenomenon that "the rich get richer and the poor get poorer," which, transferred to personality, would mean that those scoring high in a personality dimension would increase more across time (or show less of a decrease). In another vein of developmental research into old age, namely, cognitive aging research, proponents of the model of fluid and crystallized intelligence have argued that constant variances across age would be a sign of a biologically driven process of change in fluid intelligence (Horn, 1988). Such a process would serve to maintain individual differences because different individuals would be affected equally by increasing biological constraints of aging (Horn & Hofer, 1992). If one borrows from this argument in conjunction with personality development, the finding of equal factor variances would be indicative of a biologically driven change process. This does not exclude the possibility that other factors (e.g., certain events, social roles) also come into play, but keeping the focus on samples rather than individuals, it appears that individual differences are homogeneous across age. Of course, this interpretation is hampered by the fact that, in total, the amount of age difference in personality is much smaller and less systematic than in fluid intelligence. However, one might argue that constant individual differences should not be interpreted in terms of a change process but rather in light of continuity or stability of personality (cf. Allemand et al., 2007). Yet, the issue of change in factor variances is, after all, preferably and less ambiguously examined using longitudinal data (e.g., Allemand, Zimprich, & Martin, 2008).

The estimated factor covariances were also equal across age groups, implying that the factorial relations of the Big Five personality traits did not change with age. According to this finding, across 10 age groups the Big Five personality structure is invariant, which supports Costa and McCrae's (1997) argument that, after adolescence, the structure of personality is stable across age. The finding of highly stable interrelations among the five personality traits across age in adulthood is consistent with findings from other cross-sectional (Allemand, Zimprich, & Hendriks, 2008; Lang et al., 2001; Srivastava et al., 2003) and longitudinal studies (Allemand et al., 2007; Small et al., 2003). In contrast to these studies, Allemand, Zimprich, and Martin (2008) recently reported change in the structural relations among the five personality traits over 12 years in old age. Specifically, the pattern of covariation between Conscientiousness and three other traits (i.e., Extraversion, Openness to Experience, and Agreeableness) showed an increase at the second measurement occasion, indicating that the relative significance of Conscientiousness with respect to these three other personality traits seemed to become stronger over time. This suggests that personality might become less differentiated or, in turn, more dedifferentiated over time in old age.

The present results of substantial cross-sectional interrelations among most of the Big Five might also be reflected with respect to higher order factors. For example, Digman (1997) demonstrated the emergence of two consistent higher order factors. One factor (alpha) involves the common aspects of Emotional Stability (inverse Neuroticism), Agreeableness, and Conscientiousness and might be regarded as a social desirability factor, in the sense that socialization processes would shape socially acceptable levels of personality traits. The second higher order factor consisting of Extraversion and Openness to Experience might be interpreted as a factor of personal growth, which appears to reflect the tendency to explore or to engage voluntarily with novelty and may, in consequence, be associated with plasticity in behavior and cognition. However, Ashton, Lee, Goldberg, and de Vries (2009) recently concluded that correlations between personality factor scales can be explained without postulating any higher order dimensions of personality. Moreover, contrary to Digman's (1997) assumption, Neuroticism was virtually unrelated to Conscientiousness and was weakly correlated with Agreeableness in the present study.

Finally, the estimated factor means were not equal across age groups, suggesting age-related differences across the 10 age groups. Neuroticism tended to slightly decrease with age, especially in old adulthood. This negative age trend of neuroticism is consistent with other findings (Costa et al., 2000; Roberts et al., 2003, 2006). Extraversion showed a slight decrease in early midlife and then plateaued thereafter. Previous results were generally mixed for Extraversion, unless one distinguishes between two facets of Extraversion: traits related to independence and dominance (labeled *social dominance*) versus traits related to positive affect, activity level, and sociability (labeled *social vitality*; cf. Helson & Kwan, 2000). If one organizes the cross-sectional and longitudinal literature around these two categories, then the patterns of development of Extraversion suggest that people increase in measures of social dominance and decrease in measures of social vitality with age (Roberts et al., 2003, 2006). The adjective items intended to measure extraversion in the present study seem to resemble the social vitality aspect of Extraversion; thus, the cross-sectional decrease is consistent with expectations. Pertaining to age differences in the levels of Openness to Experience, the present results demonstrate a negative age trend, especially in early midlife and then from late midlife into old age. Other cross-sectional studies also reported a decrease of Openness with age (Donnellan & Lucas, 2008; Roberts et al., 2003; Srivastava et al., 2003; Zimprich, Allemand, & Dellenbach, 2009). For example, McCrae et al. (1999) found a negative relationship between Openness and age in samples drawn from six different cultures. Mixed results were found with respect to Agreeableness. It tended to show a slight decrease in early midlife and an increase from age 40 to 59. Although average levels of Agreeableness are generally positively associated with age throughout adulthood (e.g., Allemand, Zimprich, & Hendriks, 2008; McCrae et al., 1999; Srivastava et al., 2003), there are exceptions in the literature such that some studies have found little or no change in traits related to agreeableness (e.g., Costa et al., 2000). Finally, contrasting other cross-sectional and longitudinal research showing an increase of Conscientiousness throughout adulthood (Allemand, Zimprich, & Hendriks, 2008; Roberts et al., 2006; Srivastava et al., 2003), we found a slight increase in Conscientiousness up to 54 years and thereafter a decline, reflecting a curvilinear trend. However, roughly similar patterns were also observed in other cross-sectional studies. Terracciano et al. (2005) conducted cross-sectional and longitudinal analyses examining links between age and mean levels of the Big Five personality traits. They found curvilinear patterns for Conscientiousness with the exception that the cross-sectional peak for Conscientiousness was around age 50, whereas the longitudinal peak average level is near age 70. Similarly, Donnellan and Lucas (2008) recently reported that average levels of Conscientiousness were highest for middle-aged participants around age 50 in two large national samples from Great Britain and Germany. However, because of the cross-sectional nature of the present study, differential sampling by age and cohort differences are both potential sources of confounds. That is, differences (e.g., in mean levels of the Big Five) found across age groups can be attributed, in part, to the culture, climate, or historical context that an individual was born into and lived through, which is unique to each cohort (Hofer & Sliwinski, 2001).

To conclude, our age comparative analyses of variances, covariances, and means of the five personality factors across 10 age groups based on strict MI demonstrate a picture of stability and change of personality traits across age. The use of a brief list of 25 adjective ratings to assess the Big Five holds up well in terms of structure across the adult age span. Findings are consistent with other types of personality instruments in terms of the number of factors and the age invariance of factor loadings, variances, and covariances. The structure developed with confirmatory factor analysis with model modification deviated from the a priori, simple structure solution based on content validation and regression-based scale development methods (Lachman & Weaver, 1997). The modified structure allows more flexibility for items to load on multiple factors and takes correlated errors into account to achieve the best model fit. The results showing the strictest form of age invariance in factor structure are reassuring for researchers who plan to make age comparisons using the MIDUS personality measure.

References

Allemand, M., Gomez, V., & Jackson, J. J. (2010). Personality trait development in midlife: Exploring the impact of psychological turning points. *European Journal of Ageing, 7,* 147–155.

Allemand, M., Zimprich, D., & Hendriks, A. A. J. (2008). Age differences in five personality domains across the life span. *Developmental Psychology, 44,* 758–770. doi:10.1037/0012-1649.44.3.758

Allemand, M., Zimprich, D., & Hertzog, C. (2007). Cross-sectional age differences and longitudinal age changes of personality in middle adulthood and old age. *Journal of Personality, 75,* 323–358. doi:10.1111/j.1467-6494.2006.00441.x

Allemand, M., Zimprich, D., & Martin, M. (2008). Long-term correlated change in personality traits in old age. *Psychology and Aging, 23,* 545–557. doi:10.1037/a0013239

Allik, J., Laidra, K., Realo, A., & Pullmann, 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. doi:10.1002/per.524

Ashton, M. C., Lee, K., Goldberg, L. R., & de Vries, R. E. (2009). Higher order factors of personality: Do they exist? *Personality and Social Psychology Review, 13,* 79–91. doi:10.1177/1088868309338467

Bast, J., & Reitsma, P. (1997). Mathew effects in reading: A comparison of latent growth curve models and simplex models with structured means. *Multivariate Behavioral Research, 32,* 135–167. doi:10.1207/s15327906mbr3202_3

Bollen, K. A. (1989). *Structural equations with latent variables.* New York, NY: Wiley.

Briggs, S. R. (1992). Assessing the five-factor model of personality description. *Journal of Personality, 60,* 253–293. doi:10.1111/j.1467-6494.1992.tb00974.x

Brim, O. G., Ryff, C. D., & Kessler, R. C. (2004). The MIDUS national survey: An overview. In O. G. Brim, C. D. Ryff, & R. C. Kessler (Eds.), *How healthy are we? A national study of well-being at midlife* (pp. 1–36). Chicago, IL: University of Chicago Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25,* 1–27. doi:10.1177/014920639902500101

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Costa, P. T., Jr., Herbst, J. H., McCrae, R. R., & Siegler, I. C. (2000). Personality at midlife: Stability, intrinsic motivation, and responses to life events. *Assessment, 7,* 365–378. doi:10.1177/107319110000700405

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

Costa, P. T., Jr., & McCrae, R. R. (1997). Longitudinal stability of adult personality. In R. Hogan, J. Johnson & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 269–290). San Diego, CA: Academic Press.

del Barrio, V., Carrasco, M. Á., & Holgado, F. P. (2006). Factor structure invariance in the Children's Big Five Questionnaire. *European Journal of Psychological Assessment, 22,* 158–167. doi:10.1027/1015-5759.22.3.158

Dellenbach, M., & Zimprich, D. (2008). Typical intellectual engagement and cognition in old age. *Aging, Neuropsychology, and Cognition, 15,* 208–231. doi:10.1080/13825580701338094

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology, 73,* 1246–1256. doi:10.1037/0022-3514.73.6.1246

Donnellan, M. B., & Lucas, R. E. (2008). Age differences in the Big Five across the life span: Evidence from two national samples. *Psychology and Aging, 23,* 558–566. doi:10.1037/a0012897

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59,* 1216–1229. doi:10.1037/0022-3514.59.6.1216

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4,* 26–42. doi:10.1037/1040-3590.4.1.26

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40,* 84–96. doi:10.1016/j.jrp.2005.08.007

Helson, R., & Kwan, V. S. Y. (2000). Personality development in adulthood: The broad picture and processes in one longitudinal sample. In S. Hampton (Ed.), *Advances in personality psychology* (Vol. 1, pp. 77–106). London, United Kingdom: Routledge.

Hofer, S. M., & Sliwinski, M. J. (2001). Understanding aging. *Gerontology, 47,* 341–352. doi:10.1159/000052825

Hofstee, W. K. B., de Raad, B., & Goldberg, L. R. (1992). Integration of the Big Five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology, 63,* 146–163. doi:10.1037/0022-3514.63.1.146

Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 645–685). New York, NY: Plenum Press.

Horn, J. L., & Hofer, S. M. (1992). Major abilities and development in the adult period. In R. J. Sternberg & C. A. Berg (Eds.), *Intellectual development* (pp. 44–99). New York, NY: Cambridge University Press.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18,* 117–144.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. doi:10.1080/10705519909540118

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

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., & 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* (Vol.2, pp. 102–138). New York, NY: Guilford Press.

Lachman, M. E., & Weaver, S. L. (1997). *The Midlife Development Inventory (MIDI) Personality Scales: Scale construction and scoring*. Retrieved from http://www.brandeis.edu/departments/psych/lachman/pdfs/midi-personality-scales.pdf

Lang, F. R., Lüdtke, O., & Asendorpf, J. B. (2001). Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen, mittelalten und alten Erwachsenen [Validity and psychometric equivalence of the German version of the Big Five Inventory in young, middle-aged and old adults]. *Diagnostica, 47,* 111–121. doi:10.1026//0012-1924.47.3.111

Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). Weak measurement invariance with respect to unmeasured variables: An implication of strict factorial invariance. *British Journal of Mathematical and Statistical Psychology, 56,* 231–248. doi:10.1348/000711003770480020

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1,* 130–149. doi:10.1037/1082-989X.1.2.130

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4,* 84–99. doi:10.1037/1082-989X.4.1.84

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology, 70,* 810–819. doi:10.1037/0022-3514.70.4.810

Martin, M., & Zimprich, D. (2005). Cognitive development in midlife. In S. L. Willis & M. Martin (Eds.), *Middle adulthood: A lifespan perspective* (pp. 179–206). Thousand Oaks, CA: Sage.

Mascherek, A., & Zimprich, D. (in press). Age-related differences in typical intellectual engagement in young and old adults. *Experimental Aging Research.*

Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods, 11,* 344–362. doi:10.1037/1082-989X.11.4.344

McCrae, R. R., & Costa, P. T., Jr. (2003). *Personality in adulthood: A five-factor theory perspective.* New York, NY: Guilford Press. doi:10.4324/9780203428412

McCrae, R. R., Costa, P. T., Jr., de Lima, M. P., Simones, A., Ostendorf, F., Angleitner, A., . . . Piedmont, R. L. (1999). Age differences in personality across the adult life: Parallels in five cultures. *Developmental Psychology, 35,* 466–477. doi:10.1037/0012-1649.35.2.466

McCrae, R. R., Terracciano, A., & Khoury, B. (2007). Dolce far niente: The positive psychology of personality stability and invariance. In A. Ong & M. van Dulmen (Eds.), *Handbook of methods in positive psychology* (pp. 176–188). New York, NY: Oxford University Press.

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. doi:10.1037/0022-3514.89.1.90

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika, 58,* 525–543. doi:10.1007/BF02294825

Meredith, W., & Horn, J. L. (2001). The role of measurement invariance in modeling growth and change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 203–240). Washington DC: American Psychological Association. doi:10.1037/10409-007

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research, 39,* 479–515. doi:10.1207/S15327906MBR3903_4

Morizot, J., Ainsworth, A. T., & Reise, S. P. (2007). Towards modern psychometrics: Application of item response theory models in personality research. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 407–423). New York, NY: Guilford Press.

Morizot, J., & Le Blanc, M. (2003). Continuity and change in personality traits from adolescence to midlife: A 25-year longitudinal study comparing representative and adjudicated men. *Journal of Personality, 71,* 705–755. doi:10.1111/1467-6494.7105002

Mroczek, D. K., & Kolarz, C. M. (1998). The effect of age on positive and negative affect: A developmental perspective on happiness. *Journal of Personality and Social Psychology, 75,* 1333–1349. doi:10.1037/0022-3514.75.5.1333

Mroczek, D. K., & Little, T. D. (2006). *Handbook of personality development.* Mahwah, NJ: Erlbaum.

Mroczek, D. K., Ozer, D. J., Spiro, A., III, & Kaiser, R. T. (1998). Evaluating a measure of the five-factor model of personality. *Assessment, 5,* 287–301. doi:10.1177/107319119800500308

Mroczek, D. K., & Spiro, A., III. (2003). Modeling intraindividual change in personality traits: Findings from the normative aging study. *Journals of Gerontology: Series B. Psychological Sciences, 58,* 153–165. doi:10.1093/geronb/58.3.P153

Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered-categorical, and continuous latent variable indicators. *Psychometrika, 49,* 115–132. doi:10.1007/BF02294210

Muthén, L. K., & Muthén, B. (2004). *Mplus user's guide.* Los Angeles, CA: Authors.

Plaut, V. C., Markus, H., & Lachman, M. E. (2002). Place matters: Consensual features and regional variation in American well-being and self. *Journal of Personality and Social Psychology, 83,* 160–184. doi:10.1037/0022-3514.83.1.160

Prenda, K. M., & Lachman, M. E. (2001). Planning for the future: A life management strategy for increasing control and life satisfaction in adulthood. *Psychology and Aging, 16,* 206–216. doi:10.1037/0882-7974.16.2.206

Reise, S. P., Widaman, K. F., & Pugh, R. E. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114,* 552–566. doi:10.1037/0033-2909.114.3.552

Roberts, B. W., & Mroczek, D. (2008). Personality trait change in adulthood. *Current Directions in Psychological Science, 17,* 31–35. doi:10.1111/j.1467-8721.2008.00543.x

Roberts, B. W., Robins, R. W., Caspi, A., & Trzesniewski, K. H. (2003). Personality trait development in adulthood. In J. T. Mortimer & M. Shanahan (Eds.), *Handbook of the life course* (pp. 579–595). New York, NY: Plenum. doi:10.1007/978-0-306-48247-2_26

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132,* 1–25. doi:10.1037/0033-2909.132.1.1

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66,* 507–514. doi:10.1007/BF02296192

Small, B. J., Hertzog, C., Hultsch, D. F., & Dixon, R. A. (2003). Stability and change in adult personality over 6 years: Findings from the Victoria Longitudinal Study. *Journals of Gerontology: Series B. Psychological Sciences, 58,* 166–176. doi:10.1093/geronb/58.3.P166

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. doi:10.1037/0022-3514.84.5.1041

Terracciano, A., McCrae, R. R., Brant, L. J., & Costa, P. T., Jr. (2005). Hierarchical linear modeling analyses of the NEO-PI-R scales in the Baltimore Longitudinal Study of Aging. *Psychology and Aging, 20,* 493–506. doi:10.1037/0882-7974.20.3.493

Trapnell, P. D., & Wiggins, J. S. (1990). Extension of the Interpersonal Adjective Scales to include the Big Five dimensions of personality. *Journal of Personality and Social Psychology, 59,* 781–790. doi:10.1037/0022-3514.59.4.781

Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests. *Psychological Methods, 6,* 371–386. doi:10.1037/1082-989X.6.4.371

Zimprich, D., Allemand, M., & Dellenbach, M. (2009). Openness to experience, fluid intelligence, and crystallized intelligence in middle-aged and old adults. *Journal of Research in Personality, 43,* 444–454. doi:10.1016/j.jrp.2009.01.018

Zimprich, D., Allemand, M., & Hornung, R. (2006). Measurement invariance of the abridged Sense of Coherence Scale in adolescents. *European Journal of Psychological Assessment, 22,* 280–287. doi:10.1027/1015-5759.22.4.280

Zimprich, D., Allemand, M., & Huber, S. (2008). *Measurement invariance in ordered-categorical variables: The case of Big Five personality markers in adulthood.* Unpublished manuscript.

Appendix

Worked Example for the Adjusted Chi-Square Difference Test

The adjusted chi-square difference between two nested models can be calculated according the formulas given on the Mplus website (http://www.statmodel.com/chidiff.shtml). First, a scaling correction factor *SCD* for the chi-square difference has to be calculated (cf. Satorra & Bentler, 2001):

$$SCD = \frac{df_0 \times SC_0 - df_1 \times SC_1}{df_0 - df_1},$$

where df_0 is the degrees of freedom in the nested model, SC_0 is the scaling correction factor in the nested model, df_1 is the degrees of freedom in the comparison model, and SC_1 is the scaling correction factor in the comparison model. For example, the scaling correction factor for the chi-square difference between Models B and C (Table 2) was calculated as

$$SCD = \frac{242 \times 0.680 - 239 \times 0.639}{242 - 239} = 3.94633.$$

Next, the unscaled chi-square values of both models have to be calculated as

$$\chi^2 = \chi^2_{S-B} \times SC.$$

where χ^2_{S-B} is the Satorra–Bentler corrected chi-square value and *SC* is the scaling correction factor. Continuing the example, the unscaled chi-square values of Models B and C (Table 2) are

$$\chi^2_B = \chi^2_{S-B_B} \times SC_B = 2{,}862.58 \times 0.680 = 1{,}946.55,$$

$$\chi^2_C = \chi^2_{S-B_C} \times SC_C = 2{,}023.84 \times 0.639 = 1{,}293.23.$$

Finally, the adjusted chi-square difference ($\Delta\chi^2_{S-B}$) can be computed by dividing the unscaled chi-square difference by the scaling correction factor:

$$\Delta\chi^2_{S-B} = \frac{\chi^2_0 - \chi^2_1}{SCD}.$$

For the example with Models B and C (Table 2), the adjusted chi-square difference ($\Delta\chi^2_{S-B}$) is

$$\Delta\chi^2_{S-B} = \frac{\chi^2_0 - \chi^2_1}{SCD} = \frac{1{,}946.55 - 1{,}293.23}{3.94633} = \frac{653.32}{3.94633}$$
$$= 165.55.$$

Received March 23, 2009
Revision received July 5, 2011
Accepted July 20, 2011 ■