Personality traits in old age: Measurement and rank-order stability and some mean-level change

Citation for published version:
Mõttus, R, Johnson, W & Deary, IJ 2012, 'Personality traits in old age: Measurement and rank-order stability and some mean-level change' Psychology and Aging, vol. 27, no. 1, pp. 243-249. DOI: 10.1037/a0023690

Digital Object Identifier (DOI):
10.1037/a0023690

Published In:
Psychology and Aging

Personality traits in old age: Measurement and rank-order stability, and some mean-level change

René Mõttus
Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, UK
Department of Psychology, University of Tartu, Estonia

Wendy Johnson
Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, UK
Department of Psychology, University of Edinburgh, UK
Department of Psychology, University of Minnesota – Twin Cities, USA

Ian J. Deary
Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, UK
Department of Psychology, University of Edinburgh, UK

Corresponding author:
Professor Ian J. Deary
Centre for Cognitive Ageing and Cognitive Epidemiology
Department of Psychology
University of Edinburgh
7 George Square
Edinburgh EH8 9JZ
Scotland, UK
Email i.deary@ed.ac.uk

Abstract

Lothian Birth Cohorts 1936 and 1921 were used to study the longitudinal comparability of Five-Factor Model personality traits from ages 69 to 72 years and from ages 81 to 87 years, and cross-cohort comparability between ages 69 and 81 years. Personality was measured using the 50-item International Personality Item Pool (IPIP). Satisfactory measurement invariance was established across time and cohorts. High rank-order stability was observed in both cohorts. Almost no mean-level change was observed in the younger cohort, whereas Extraversion, Agreeableness, Conscientiousness and Intellect declined significantly in the older cohort. The older cohort scored higher on Agreeableness and Conscientiousness. In these cohorts, individual differences in personality traits continued to be stable even in very old age, mean level changes accelerated.

Keywords: personality development; five-factor model; rank-order stability; mean-level changes; measurement invariance.

Introduction

There have been few personality datasets available to date to test the extent to which individual differences in personality remain stable in old age, especially in people aged over 80. We attempted to fill this gap by combining longitudinal and cross-cohort data from two Scottish birth cohorts—the Lothian Birth Cohorts of 1921 and 1936 (LBC1921 and LBC1936)—to test measurement and rank-order stability of personality traits in people aged around 70 and over 80 years. Additionally, we addressed longitudinal and cross-cohort mean-level changes in the traits.

Measurement stability

There is substantial cross-sectional evidence for the structural replicability of the Five-Factor Model (FFM; McCrae & John, 1992) personality traits in various age groups (De Fruyt, De Bolle, McCrae, Terracciano, & Costa, 2009; McCrae & Terracciano, 2005). It is therefore reasonable to start with the assumption that personality instruments tailored to measure the FFM personality traits are likely to be suitable means of summarizing personality differences among aged people. Thus, instead of asking *which* is the best structure underlying the co-variation of all possible personality test items in old age, we assumed that the FFM was appropriate and asked whether the measurements within the intended structure of a particular FFM instrument were time-invariant in old age. Measurement invariance implies that the indicators of latent traits (e.g. items defining the trait) contribute to their scores in the same way over time. If this was not the case, trait scores would not be directly comparable across time; and investigating personality trait development in old age would essentially be impossible.

There is cross-sectional and longitudinal evidence that the FFM traits can be measured invariantly up to the eighth decade (Allemand, Zimprich, & Hendriks, 2008; Allemand, Zimprich, & Hertzog, 2007; Small, Hertzog, Hultsch, & Dixon, 2003). Little is yet known about whether the instruments tailored to FFM traits hold their measurement properties equally well beyond age 70, especially in people aged over 80.

Rank-order stability

The trait perspective on personality is based on the assumption that people's trait rankings are relatively stable across time (Johnson, 1999), notwithstanding that people, on average, show changes in personality traits. Without high rank-order stability, personality traits would not describe what they are expected to describe—enduring individual characteristics.

Roberts and DelVecchio (2000) reported that mean correlations of personality traits across an average period of 6.7 years were about 0.31 in childhood and increased to around 0.70 in middle adulthood and the early seventies of age. Much less is known about their magnitude thereafter, especially in the ninth decade of life. One of the few studies (Martin, Long, & Poon, 2002) reported that people generally retained their relative positions on personality traits, but rank-order stability was lower in people aged over eighty compared to people in their sixties. Rank-order stability in old age thus calls for more research.

Mean-level changes

Mean-level changes describe age-related differences in trait levels that tend to be more or less universal. Although mean-level changes have no implications for the meaningfulness of traits as such—temporally stable individual differences can co-exist with substantial and even rapid mean-level changes—they reflect an important aspect of personality development, being potentially informative with respect to sources of, or adaptive purposes for, personality changes (cf. Roberts, Walton, & Viechtbauer, 2006).

Among the few relevant cross-sectional studies in people aged over 70, Roepke and colleagues (2001) found that Extraversion was higher after age 80. Lucas and Donnellan (2009) reported that Neuroticism, Extraversion (only in men) and Openness were negatively related to age after 70, whereas Conscientiousness and Agreeableness showed the opposite age differences (the latter only in men). Donnellan and Lucas (2008) found that Openness was negatively associated with age up to the early eighties, whereas there were less systematic age differences in other traits. Weiss and colleagues (2005) observed people (especially men) over 80 being higher on Agreeableness than slightly younger people, but no differences in other FFM traits. Longitudinal studies have reported mixed results. Small et al. (2003) found in a sample of people aged from 55 to 85 years that men increased in Neuroticism, whereas women decreased in Neuroticism and increased in Agreeableness over a 6-year period. Field and Millsap (1991) found an increase in Agreeableness and decrease in Extraversion. Mroczek and Spiro (2003) found in a male sample that Neuroticism declined until the mid-seventies and then increased slightly, but there were no mean-level changes in Extraversion. In sum, the existence and direction of late-life mean-level changes in the FFM personality traits, and potential sex differences in the changes, are far from clear.

Method

Participants and procedure

The LBC1921 was recruited from follow-up participants of the Scottish Mental Survey of 1932 (Deary, Whiteman, Starr, Whalley, & Fox, 2004). The participants of LBC1921 were generally healthy and lived in their own homes in the city of Edinburgh and its surrounding area. The LBC1921 participants were recruited to the study when they were around age 79. Personality traits of 450 people (263 women) were first measured in a follow-up mail-in questionnaire at ages 80 to 81 years (the exact date of completing the questionnaire was not recorded; for convenience, this age is henceforth referred to as 81). At ages about 86 to 87 (henceforth referred to as age 87), 209 participants (110 women; mean age 86.61 ± 0.42 at the second testing) completed the same questionnaire in the course of a comprehensive follow up study (Starr et al., 2010). Comparison of personality traits of those who participated at both testing occasions with those who participated only at the first occasion showed that the latter had significantly lower emotional stability ($t = 2.95$, $df = 441$, $p < 0.01$, Cohen's $d = 0.28$).

The LBC1936 was recruited from surviving participants of the Scottish Mental Survey of 1947. They were first assessed at ages about 69 to 70. Details of the initial recruitment and testing procedures are reported by Deary et al. (2007). At initial testing (henceforth referred to as age 69), 923 people (471 women; mean age 69.52 ± 0.84 years) provided personality data. About three years later (henceforth referred to as age 72), 750 participants who could and agreed to continue in the study (366 women; mean age 72.49 ± 0.73 years) completed the same questionnaire in course of a comprehensive follow up study. Comparison of personality traits of those who participated at both testing occasions with those who participated only at the first occasion showed that the latter had significantly lower emotional stability ($t = 2.76$, $df = 921$, $p < 0.01$, Cohen's $d = 0.18$).

The fact that low Emotional Stability at first testing occasion was related to higher likelihood of drop-out in both cohorts may potentially have influenced our findings. For instance, a common result of such range restriction is understatement of the effect sizes.

Measure

The 50-item International Personality Item Pool (IPIP; Goldberg, 1999), freely downloadable from the internet at http://ipip.ori.org, contains ten items for each of the five FFM personality traits—Emotional Stability (often called its inverse, Neuroticism), Extraversion, Agreeableness, Conscientiousness, and Intellect (similar to Openness often tapped in other FFM questionnaires, but see also DeYoung, Shamosh, Green, Braver, & Gray, 2009). According to Gow, Whiteman, Patty and Deary (2005), the IPIP scales correlate moderately to highly ($r = 0.49$ to 0.83) to the corresponding scales of the NEO Five-Factor Inventory (Costa & McCrae, 1992). Participants were requested to rate how well they believed each of the 50 items described them on a 5-point Likert-type scale (very inaccurate to very accurate). In the present data, internal consistencies of the IPIP scales were adequate in both cohorts and testing occasions, with Cronbach's alphas ranging from 0.68 to 0.88 (median = 0.80).

Analytic procedures

Measurement invariance. To estimate measurement invariance of personality traits across two testing occasions and cohorts, a series of multiple-group confirmatory factor analyses (CFAs) was carried out. Longitudinal invariance was estimated separately within the two cohorts, including only those participants who provided data at both testing occasions. Cross-cohort invariance was estimated using the data from the first testing occasions. In the CFA models, each item was allowed to load only on its intended factor as we imposed the intended structure on the data. Although the assumption of the lack of cross-loadings is not fully realistic (Church & Burke, 1994), this made sense in the context of estimating measurement invariance: normally personality scales are scored so that each item can contribute to only one factor. Latent factors were allowed to covary. Some items were allowed to have correlated residuals, because they had extremely similar content (e.g. "I seldom feel blue" and "I often feel blue"). The items for which residuals were allowed to correlate (see Appendix I) were determined on the basis of item content and by observing the largest residual covariances that were consistent across the cohorts.

Importantly, in all models throughout the study identical item pairs were allowed to have correlated residuals.

Measurement invariance was tested on the basis of Meredith's (1993) guidelines. In Model 1, factor loadings, residual variances and intercepts were free to differ across the testing occasions/cohorts. Models 2, 3 and 4 introduced increasingly strict constraints on model parameters, first forcing only factor loadings to be equal across time/cohorts, then adding equality constraints to intercepts, and finally also to residual variances. If imposing loading constraints had resulted in a significant deterioration of the model fit, it would have indicated that the items differed in relative importance to the trait over time or across cohorts. Deteriorations of model fit resulting from equality constraints on intercepts would have meant that, everything else equal, the average item responses differed over time. Deteriorations of model fit resulting from equality constraints on residual variances would have indicated that the proportions of variance in item responses that were independent of the latent factors differed over time or across cohorts. Deteriorations of model fit were first estimated by means of the chi-square differences test. However, as the chi-square test is known to be highly dependent on sample size, we relied more heavily on three other information theoretic indices designed for estimating overall model fit or for comparing competing models that take into consideration sample size and model parsimony: root mean square error of approximation (RMSEA; Steiger & Lind, 1980); Akaike information criterion (AIC; Akaike, 1974); and Bayesian information criterion (BIC; Schwarz, 1978).

Rank-order stability and longitudinal mean-level changes. In both cohorts, we constructed a separate model for each FFM domain, with all items at the first testing occasion loading on one factor and all items at the second testing occasion loading on the second factor. Factor loadings, residual variances and intercepts were constrained to be equal over time. Several items (determined by seeking maximum model fit on the basis of AICs and BICs) were allowed to have correlated residuals across testing occasions as they apparently had some age-related variance over and above the latent traits, which were our main focus. The correlations between corresponding latent factors at first and second testing occasion were interpreted as estimates of rank-order stability. To estimate longitudinal mean differences, we fixed latent factor scores at first testing to zero and only estimated means for the second testing occasion, which were then interpreted as showing the directions and effect sizes of the mean-level changes. To estimate possible sex differences in the mean-level changes, we additionally ran models with men and women specified as separate groups and tested for differences in the change parameters.

Cross-cohort differences. Cross-cohort mean-level differences were studied using a multiple-group CFA model similar to that used for testing measurement invariance (Model 4), with the only difference that models were run separately for each FFM domain. We also estimated possible sex differences in the mean-level changes by testing whether the effect of cohort on the factor scores from the fully constrained models differed significantly across sexes.

Details on all models are available on request from the corresponding author.

Results

Measurement invariance

According to RMSEAs, all models fit the data at least satisfactorily (Table 1), suggesting that the FFM, as embodied in the IPIP, described personality differences in these two samples adequately well. Importantly, RMSEAs differed very little between models with different numbers of constraints, with all the changes taking place within 90% confidence intervals of competing models. CFIs did not meet the threshold for satisfactory model fit, likely due to lack of allowed cross-loadings. Importantly, however, the differences in CFIs between models with different constraints were minimal. The fit indices that recognize model parsimony—AIC and especially BIC, which rewards model simplicity to an even greater degree—tended to be relatively favourable for models with consecutively greater constraints.

To illustrate the differences in measurement properties over time and across cohorts, we describe the differences in the loadings of items on the latent traits in unconstrained models. The median differences in standardized loadings over time were 0.06 (varying from 0.00 to 0.24) in LBC1921, and 0.04 (varying from 0.00 to 0.14) in LBC1936. In cross-cohort comparisons, the median difference in the loadings was 0.05 (varying from 0.00 to 0.22). The three items with differences in loadings greater than 0.20 belonged exclusively to the Intellect domain and described cognition (e.g., “I have difficulty understanding abstract ideas“). Given that cognitive decline in old age tends to be normative, slight differences in the role of the cognition-related items in defining the broad personality domain Intellect are perhaps not surprising. In summary, there was an acceptable degree of measurement invariance for the five IPIP factors across the two testing occasions in both cohorts and across cohorts. Thus, the IPIP appeared to measure the five broad FFM traits relatively robustly across time and groups of people in these elderly samples.

Rank-order stability

In LBC1921, the correlations between latent factors ranged from 0.78 to 0.89 with a median of 0.83 (Table 2). In LBC1936, the correlations ranged from 0.83 to 0.94 with a mean of 0.88. The slightly higher rank-order stability in the younger cohort probably reflected in part the shorter interval between testing occasions (3 years as opposed to 6 years in the older cohort).

Mean-level changes

In LBC1921, mean scores on all traits declined from ages 81 to 87 years, with the changes being significant for Extraversion, Agreeableness, Conscientiousness and Intellect (Table 2). There was a significant age x sex interaction in Agreeableness, with the general decrease being driven only by women ($d = -0.64$, $p < 0.01$) and men showing no decline at all. In LBC1936, the changes were very small, and significant only for Extraversion, which slightly increased during the three-year period. There were age x sex interactions, however. Men increased in Agreeableness ($d = 0.14$, $p < 0.01$), whereas women decreased ($d = -0.20$, $p < 0.001$). In addition, women but not men decreased in Emotional Stability ($d = -0.08$, $p < 0.05$). LBC1921 members at age 81 had higher mean scores on Agreeableness and Conscientiousness than LBC1936 members at age 69. Sex had no significant effect on cross-cohort differences.

Discussion

The data from LBC1921 and LBC1936 showed that personality traits retained a satisfactory degree of measurement and rank-order stability from the beginning of the eighth to the second half of the ninth decades of people's lives. This suggests that individual differences in the five FFM personality traits met the fundamental requirements for being credible units of description of enduring human variation in these aged cohorts.

The use of a particular personality questionnaire did not allow us to investigate whether the FFM, and not any other model, was the most optimal personality model for describing personality differences in old age—our analyses were intended to measure the continuity in the content of these five broad traits only. However, since we found evidence for adequate measurement invariance in the traits, we believe that, within the framework of the broad FFM traits, there is structural continuity in personality traits up to very old age. We note that CFIs of our models were often lower than would be considered satisfactory—although the RMSEAs were in the acceptable range. This was especially the case in the models that included multiple FFM domains (Table 1). The low CFI's were largely caused by substantial correlations between items belonging to different domains, which is commonly observed in all instruments used to measure the FFM. At the same time, in our single-domain models (Table 2) the CFIs were not always at levels that would usually be considered satisfactory either, which indicates that restricting cross-loading was not the only reason for the poor CFI's. The primary reason was that CFIs estimate the degree to which a specified model that shows relatively large differences in correlations among variables is better than an independence model that assumes no (and therefore even) covariance among variables. In the cases of our models, the covariances among variables tended to be more even than presumed by the specified FFM models, resulting in large CFI penalties due to the unrecognized residual item correlations and cross-loadings.

There were no substantial differences in the overall magnitudes of the rank-order stability coefficients for personality traits between the LBC1921 and LBC1936. The coefficients were also in the same range as those reported for younger age groups (e.g. Allemand et al., 2007), suggesting that there was no decrease in the rank-order stability of personality in these cohorts that extended to the ninth decade of life. Whether the small variations in the rankings of people result from idiosyncratic changes or could be ascribed to identifiable factors is a question for future research.

We assessed age-related mean-level differences from two perspectives, longitudinal and cross-generational. According to the longitudinal results, the mean-level changes in personality traits accelerated in old age: there were no substantial changes in the beginning of the eighth decade but more pronounced changes within the ninth decade. Although the testing interval was longer in the older cohort, the effect sizes of changes were much larger, suggesting that the more pronounced changes in the ninth decade compared to the previous decade probably did not reflect only the longer testing interval. This conclusion is consistent with there being effects on personality from the relatively normative changes in people's lives in later old age, such as deterioration of health and cognition, narrowing of social interaction and increasingly limited ability to function independently. For instance, to retain a comparable level of well-being (Emotional Stability), as functionality sharply declines, people have to be more selective in their investments in social and intellectual life (decreasing Extraversion and Intellect), and how much effort they can put into taking care of other people or being orderly and dutiful (decreasing Agreeableness and Conscientiousness; cf. Freund & Baltes, 2002).

The *existence* of sex differences in the mean-level changes in Agreeableness tended to be in line with previous findings, but the directions of the changes were only partly consistent. Although the previously reported sex differences in age trends of Agreeableness have also been somewhat inconsistent, Agreeableness has more often shown increases in men (Lucas & Donnellan, 2009; Weiss et al., 2005) than in women (Small et al., 2003). In the present study, the men from the younger cohort also increased slightly in Agreeableness. On the other hand, our relatively stronger decline of Agreeableness in women is a novel finding.

Comparing the results from the first testing occasions, the older cohort had higher Agreeableness and Conscientiousness than the younger cohort. However, these cross-cohort differences may result from sources other than age. First, the cohorts represented two generations born 15 years apart and it is possible that observed cross-cohort differences reflected secular trends in personality traits. Second, members of LBC1921 were in the second half of their ninth decade at second testing, far beyond the average life-expectancy. This fact alone makes the sample potentially different from the younger sample. That is, assuming that personality traits may be associated with mortality, samples in their early seventies and late eighties are likely to be inherently different in lifelong average personality trait levels—regardless of pure age effects. Indeed, there is evidence that low Conscientiousness and antisocial personality traits predict mortality (Kern & Friedman, 2008; Nabi et al., 2008). Therefore, age differences in cross-cohort comparisons may have been confounded by sampling differences resulting from differential mortality. Future studies on personality development in old age are likely to benefit more from longitudinal designs, especially those started early in the life-course and therefore better able to address the sampling issues.

References

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

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.

Church, A. T., & Burke, P. J. (1994). Exploratory and confirmatory tests of the big five and Tellegen's three- and four-dimensional models. *Journal of Personality and Social Psychology*, *66*, 93-114.

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

De Fruyt, F., De Bolle, M., McCrae, R. R., Terracciano, A., & Costa, P. T. (2009). Assessing the universal structure of personality in early adolescence: The NEO-PI-R and NEO-PI-3 in 24 cultures. *Assessment*, *16*, 301-311.

Deary, I. J., Gow, A. J., Taylor, M. D., Corley, J., Brett, C., Wilson, V., Campbell, H., et al. (2007). The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatrics*, *7*, 28-28.

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The Impact of Childhood Intelligence on Later Life: Following Up the Scottish Mental Surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, *86*, 130-147.

DeYoung, C. G., Shamosh, N. A., Green, A. E., Braver, T. S., & Gray, J. R. (2009). Intellect as distinct from openness: Differences revealed by fMRI of working memory. *Journal of*

Personality and Social Psychology, *97*, 883-892.

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.

Field, D., & Millsap, R. E. (1991). Personality in advanced old age: continuity or change? *Journal of Gerontology*, *46*, 299-308.

Freund, A. M., & Baltes, P. B. (2002). Life-management strategies of selection, optimization and compensation: Measurement by self-report and construct validity. *Journal of Personality and Social Psychology*, *82*, 642-662.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. J. Deary, F. D. Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe* (Vol. 7, pp. 7–28). Tilburg: Tilburg University Press.

Gow, A. J., Whiteman, M. C., Pattie, A., & Deary, I. J. (2005). Goldberg's IPIP Big-Five factor markers: Internal consistency and concurrent validation in Scotland. *Personality and Individual Differences*, *39*, 317-329.

Johnson, J. A. (1999). Persons in situations: Distinguishing new wine from old wine in new bottles. *European Journal of Personality*, *13*, 443-453.

Kern, M. L., & Friedman, H. S. (2008). Do Conscientious Individuals Live Longer? A Quantitative Review. *Health Psychology*, *27*, 505–512.

Lucas, R. E., & Donnellan, M. B. (2009). Age differences in personality: Evidence from a nationally representative Australian sample. *Developmental Psychology*, *45*, 1353-1363.

Martin, P., Long, M. V., & Poon, L. W. (2002). Age changes and differences in personality traits and states of the old and very old. *The Journals of Gerontology. Series B, Psychological*

Sciences and Social Sciences, *57*, 144-152.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*, 175-215.

McCrae, R. R., & Terracciano, A. & 78 Members of the Personality Profiles of Cultures Project. (2005). Universal Features of Personality Traits From the Observer's Perspective: Data From 50 Cultures. *Journal of Personality and Social Psychology*, *88*, 547–561.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525-543.

Mroczek, D. K., & Spiro, A. (2003). Modeling intraindividual change in personality traits: findings from the normative aging study. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *58*, 153-165.

Nabi, H., Kivimäki, M., Zins, M., Elovainio, M., Consoli, S. M., Cordier, S., Ducimetière, P., et al. (2008). Does personality predict mortality? Results from the GAZEL French prospective cohort study. *International Journal of Epidemiology*, *37*, 386 -396.

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

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.

Roepke, S., McAdams, L. A., Lindamer, L. A., Patterson, T. L., & Jeste, D. V. (2001). Personality profiles among normal aged individuals as measured by the NEO-PI-R. *Aging & Mental Health*, *5*, 159-164.

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. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *58*, 166-176.

Starr, J. M., Kilgour, A., Pattie, A., Gow, A. J., Bates, T. C., & Deary, I. J. (2010). Height and intelligence in the Lothian Birth Cohort 1921: a longitudinal study. *Age Ageing, 39*, 272-275.

Weiss, A., Costa, P. T., Karuza, J., Duberstein, P. R., Friedman, B., & McCrae, R. R. (2005). Cross-Sectional Age Differences in Personality Among Medicare Patients Aged 65 to 100. *Psychology and Aging*, *20*, 182-185.

Author Note

Wendy Johnson and Ian Deary are members of and René Mõttus is a visiting research fellow in The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (G0700704/84698). René Mõttus is also at the Department of Psychology, University of Tartu, Estonia. Wendy Johnson and Ian Deary are also at the Department of Psychology, University Of Edinburgh, UK. Wendy Johnson is also at the Department of Psychology, University of Minnesota – Twin Cities, USA. René Mõttus' fellowship is supported by European Social Fund (grant no MJD44 of the Mobilitas program, coordinated by Estonian Science Foundation). Wendy Johnson is a Research Fellow of the United Kingdom. The original recruitment of the Lothian Birth Cohort 1921 (LBC1921) was funded by the UK's Biotechnology and Biological Sciences Research Council. The personality questionnaire data for the LBC1921 on both occasions was funded by the Chief Scientist Office of the Scottish Government's Health Directorates. The original recruitment of and initial personality questionnaire data collection from the Lothian Birth Cohort 1936 (LBC19936) was funded by Research Into Ageing, and the follow-up personality data were funded by Age UK as part of The Disconnected Mind project. Funding from the BBSRC, EPSRC, ESRC and MRC is gratefully acknowledged.

Table 1. Multiple-group CFAs for within and between-cohort comparisons: Fit indices for progressively more constrained models.

	χ^2	df	$\Delta\chi^2$	Δdf	CFI	AIC	BIC	RMSEA [90% CIs]
Longitudinal measurement invariance in LBC1921 (N = 209)								
Model 1	3998.36***	2304	-	-	0.73	56832.40	58228.68	0.059 [0.056; 0.062]
Model 2	4032.15***	2349	33.79	45	0.73	56776.20	57990.88	0.059 [0.055; 0.062]
Model 3	4075.05***	2394	42.90	45	0.73	56729.10	57762.18	0.058 [0.055; 0.061]
Model 4	4230.19***	2444	155.14***	50	0.72	56784.23	57615.54	0.059 [0.056; 0.062]
Longitudinal measurement invariance in LBC1936 (N = 750)								
Model 1	6682.73***	2304	-	-	0.82	195698.83	197537.20	0.050 [0.049; 0.052]
Model 2	6731.08***	2349	48.35	45	0.82	195657.18	197256.46	0.050 [0.048; 0.051]
Model 3	6809.17***	2394	78.09**	45	0.82	195645.27	197005.45	0.050 [0.048; 0.051]
Model 4	6906.23***	2444	97.06***	50	0.82	195642.33	196736.85	0.049 [0.048; 0.051]
Cross-cohort measurement invariance between LBC1921 (N = 450) and LBC1936 (N = 923)								
Model 1	6567.58***	2304	-	-	0.80	183600.52	185408.29	0.052 [0.050; 0.053]
Model 2	6651.93***	2349	84.35***	45	0.80	183594.87	185167.52	0.052 [0.050; 0.053]
Model 3	6825.17***	2394	173.24***	45	0.79	183678.11	185015.65	0.052 [0.050; 0.053]
Model 4	7034.98***	2444	209.805***	50	0.78	183787.92	184864.22	0.052 [0.051; 0.054]

NOTE: Model 1 = no equality constraints across testing occasions/cohorts; Model 2 = only factor loadings constrained to be equal across testing occasions/cohorts; Model 3 = factor loadings and intercepts constrained to be equal across testing occasions/cohorts; Model 4 = factor loadings, intercepts and residual variances constrained to be equal across testing occasions/cohorts; χ^2 = chi-square; df = degrees of freedom; $\Delta\chi^2$ = chi-square difference; Δdf = degrees of freedom difference; CFI = Comparative fit index; AIC = Akaike information criterion; BIC = Bayesian information criterion; RMSEA = root mean square error of approximation; CI = confidence interval. *** $p < 0.001$; ** $p < 0.01$.

Table 2. Rank-order stability and mean-level differences between the latent personality traits along with the fit indices of respective models.

	Longitudinal correlations and mean-level changes in LBC1921 (ages 81 to 87 years)				Longitudinal correlations and mean-level changes in LBC1936 (ages 69 to 72 years)				Cross-cohort mean-level differences (ages 69 to 81 years)		
	r	d	RMSEA [CIs]	CFI	r	d	RMSEA [CIs]	CFI	d	RMSEA [CIs]	CFI
ES	0.83***	-0.05	0.060 [0.049; 0.071]	0.92	0.83***	0.02	0.054 [0.049; 0.059]	0.94	-0.08	0.068 [0.061; 0.075]	0.94
E	0.89***	-0.23***	0.070 [0.060; 0.081]	0.88	0.94***	0.05*	0.057 [0.052; 0.062]	0.93	-0.11	0.077 [0.070; 0.084]	0.91
A	0.78***	-0.28***	0.063 [0.053; 0.074]	0.83	0.83***	-0.01	0.043 [0.038; 0.049]	0.94	0.19**	0.069 [0.063; 0.076]	0.89
C	0.81***	-0.32***	0.072 [0.062; 0.082]	0.80	0.88***	-0.04	0.057 [0.053; 0.062]	0.90	0.13*	0.075 [0.068; 0.082]	0.86
I	0.88***	-0.23**	0.063 [0.052; 0.073]	0.84	0.94***	0.00	0.067 [0.062; 0.072]	0.88	-0.04	0.060 [0.053; 0.067]	0.92

NOTE: N = 209 (LBC1921); N = 750 (LBC1936); N = 1,373 (cross-cohort comparison, i.e. LB1921 vs LBC1936). ES = Emotional Stability; E = Extraversion; A = Agreeableness; C = Conscientiousness; I = Intellect; r = correlation between latent factors scores from the first and second testing occasion; d = mean-level difference (Cohen's d) between the first and the second testing occasion (or between the first testing occasions in cross-cohort comparisons), positive values indicating older people having higher scores; RMSEA = root mean square error of approximation; CIs = 90% confidence intervals; CFI = Comparative fit index. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Appendix I. The pairs of the IPIP items that were allowed to have correlated residuals in all models (indicated by identical letters).

Domain / Item	Domain / Item	Domain / Item
Emotional Stability	***Agreeableness***	***Intellect***
I get stressed out easily	I feel little concern for others	I have a rich vocabulary (m)
I am relaxed most of the time	I am interested in people	I have difficulty understanding abstract ideas (j)
I worry about things	I insult people	I have a vivid imagination (l)
I seldom feel blue (b)	I sympathise with others' feelings	I am not interested in abstract ideas (j)
I am easily disturbed	I am not interested in other people's problems (h)	I have excellent ideas (k)
I get upset easily	I have a soft heart	I do not have a good imagination (l)
I change my mood a lot (a)	I am not really interested in others (h)	I am quick to understand things
I have frequent mood swings (a) (c)	I take time out for others	I use difficult words (m)
I get irritated easily	I feel others' emotions	I spend time reflecting on things
I often feel blue (b) (c)	I make people feel at ease	I am full of ideas (k)
Extraversion	***Conscientiousness***	
I am the life of the party (f)	I am always prepared	
I don't talk a lot (e)	I leave my belongings around (i)	
I feel comfortable around people (d)	I pay attention to details	
I keep in the background (g)	I make a mess of things	
I start conversations	I get chores done right away	
I have little to say (e)	I often forget to put things back in their proper place (i)	
I talk to a lot of different people at parties (d) (f)	I like order	
I don't like to draw attention to myself (g)	I shirk my duties	
I don't mind being the centre of attention	I follow a schedule	
I am quiet around strangers	I am exacting in my work	