Journal of Research in Personality 99 (2022) 104242

Contents lists available at ScienceDirect

Journal of Research in Personality

journal homepage: www.elsevier.com/locate/jrp

Testing the personality differentiation by intelligence hypothesis in a representative sample of Swedish hexagenerians

Anton Andersson[a], Petri Kajonius[c,d], Valgeir Thorvaldsson[a,b,*]

[a] Department of Psychology, University of Gothenburg, Sweden
[b] Center for Ageing and Health (AgeCap), University of Gothenburg, Sweden
[c] Department of Behavioral Science, University West, Sweden
[d] Department of Psychology, Lund University, Sweden

ARTICLE INFO

Keywords:
Personality Differentiation by Intelligence
MG-CFA
Big Five
Intelligence

ABSTRACT

The Personality Differentiation by Intelligence Hypothesis (PDIH) predicts larger trait-variances, and smaller across-trait covariances for individuals with higher intelligence. We tested these predictions using multiple-group confirmatory factor analyses (MG-CFA), while controlling for the potential confound of systematic method variance related to reversed items using a correlated trait, correlated method (CTCM) approach. Participants between the ages of 62 and 68 completed measures of personality (Mini-IPIP: Donnellan et al., 2006) and intelligence (Raven APM-12: Arthur & Day, 1994). After establishing strict measurement invariance (MI), we found no support for larger variances, and only minor support for lower trait covariances as related to higher intelligence. Overall, the findings provide scant support for the PDIH when controlling for systematic method variance.

1. Introduction

1.1. Personality differentiation by intelligence

The personality differentiation by intelligence hypothesis (PDIH; Brand, Egan, & Deary, 1994) suggests that higher intelligence is associated with greater differentiation in personality. This is similar to the finding that higher general intelligence (*g*) leads to greater differentiation in abilities related to intelligence, with lower intercorrelations for different facets of intelligence with higher *g* (Abad, Colom, Juan-Espinosa & García, 2003; Blum & Holling, 2017; Deary et al., 1996; Spearman, 1927: p. 217-219). More specifically, PDIH predicts greater trait-variance and lower covariance across personality traits with increasing intelligence (Austin, Deary, & Gibson, 1997; Austin, Hofer, Deary, & Eber, 2000). Both predictions can be explained by individual differences in intelligence potentially influencing how scale items of personality questionnaires are perceived and understood (Austin et al., 2000).

Greater trait-variance could be the result of those with higher intelligence viewing the scale items of personality questionnaires as more meaningful self-descriptors, which would lead to a more extensive use of scale-ranges and more extreme scores, and thus, greater variance in the upper parts of the intelligence distribution (Austin et al., 2000). This greater variance and higher prevalence of extreme scores could also be regarded as reflecting a higher degree of traitedness (LaHuis, Barnes, Hakoyama, Blackmore, & Hartman, 2017), which would comply with the assertion of higher intelligence individuals having "more" personality, in terms of being more distinguishable from one another on self-reports (Brand et al., 1994). Moreover, scale reliability would also be expected to be higher with higher intelligence as "…extreme scores can only be obtained by consistent responding" (Austin et al., 2000: p. 407).

Smaller trait-covariation by higher intelligence could reflect those of higher intelligence being more sensitive to distinctions between items belonging to different factors. It would then be expected that the number of personality dimensions increase as intelligence increases (Austin et al., 1997). In this study, we tested these two predictions as derived from the PDIH using a confirmatory factor analytical framework fitted to a large, representative sample of Swedish hexagenerians.

1.2. Previous research

Findings concerning PDIH have been inconclusive. For example, Austin, Deary and Gibson (1997) found larger individual differences among individuals with higher intelligence but found no support for a difference in the magnitude of correlations between personality factors depending on ability level. However, Austin et al. (2002) found

* Corresponding author at: Department of Psychology, University of Gothenburg, Box 500, SE-40530 Gothenburg, Sweden.
E-mail address: Valgeir.Thorvaldsson@psy.gu.se (V. Thorvaldsson).

https://doi.org/10.1016/j.jrp.2022.104242
Received 1 December 2021; Received in revised form 9 May 2022; Accepted 13 May 2022
Available online 18 May 2022

moderating effects of intelligence on the correlation between *psychoticism* and *neuroticism*, suggesting that the correlation between these two traits decreased with increasing intelligence. Harris, Vernon, and Jang (2005) found higher variability on 15 out of 20 personality factors relating to higher intelligence, while De Fruyt, Aluja, Garcia, Rolland and Jung (2006) reported larger trait variances as depending on higher intelligence for three out of five factors. The variance differences were relatively small, however, and there were only minor differences in the correlations between factors as depending on intelligence. Furthermore, Möttus, Allik, and Pullman (2007) found some support for lower Big Five intercorrelations as depending on intelligence. More recently, Schermer, Bratko, and Bojic (2020) found support for the PDIH in a young, Croatian sample with higher variance, higher scale reliabilities, and greater ranges in the upper tertile intelligence-group compared to the lower tertile group.

Using Item Response Theory (IRT), Waiyawutti, Deary and Johnson (2012) found no evidence for differential item functioning between groups differing in intelligence. Similarly, De Fruyt et al. (2006) found personality trait invariance across groups of differing intelligence, using multiple group confirmatory factor analysis (MG-CFA). The results of McLarnon and Carswell (2013), however, contradict these findings by demonstrating non-invariance for personality factors across ability groups using a similar analytical approach. Also, Schermer, Krammer, Goffin and Biderman (2020) found lack of metric and scalar invariance (evincing non-invariance at the measurement level) in measures of 16PF based on analysis of a median split of participants' scores on an extracted *g*-factor, yet concluding little support for the PDIH based on the inconclusive results of the expected patterns of intercorrelations.

1.3. Potential confounds of PDIH

Austin et al. (1997) noted higher reliability of measurements for high-intelligence groups, which may account for the PDIH, given that extreme scores and higher scale variability might be a consequence of more precise measurements, rather than reflecting "true" personality differentiation. Thus, differential item comprehension or greater response consistency, as reflected by higher reliability coefficients among more able respondents (e.g., Allik et al., 2004; Schermer, Bratko, & Bojic, 2020), may fully or partly explain differences in variance among groups differing in intelligence (Austin et al., 2000, 2002). Findings concerning reliability differences by intelligence have been mixed, however. Möttus et al. (2007) found few statistically significant differences in reliability between high- and low-intelligence groups, although this might have been attributable in part to the study's small sample size, as the high-intelligence group displayed higher reliability for all facets of personality except three. Others have suggested that the higher response consistency and measurement reliability indices may in and of themselves be reflective of a higher degree of "traitedness", with high-intelligence individuals perceiving traits with more clarity and as being more salient for them (LaHuis et al., 2017; Navarro-Gonzalez, Ferrando, & Vigil-Colet, 2018). In an attempt to account for the potential confounding effects of reliability, Escorial, Navarro-Gonzalez, Ferrando, and Vigil-Colet (2019) found that when controlling for persons ́ reliability parameters, the finding of more multi-dimensional factor structure derived for higher intelligence participants (indicating higher differentiation) disappeared. The recommended number of factors were the same regardless of ability, lending some support to the notion of higher personality differentiation among higher intelligence being the result of confounding effects of differential reliability. Furthermore, Murray, Booth and Molenaar (2016), using a moderated factor model to investigate the PDIH with moderated residuals to account for differential reliability, found limited support for the overall hypothesis.

A related issue is that of confounds of systematic method biases. Rammstedt, Goldberg and Borg (2010) found that the Big Five trait structure did not emerge as clearly in low-educated portions of their samples. Yet, when controlling for biases related to acquiescent responding, the factor structure emerged clearly in both high- and low-educated groups. This illustrates the potential confounding effects of sources of systematic method biases when drawing conclusions about invariances between groups. Weijters, Baumgartner and Schillewaert (2013) showed that biases related to response inconsistency accounted for up to nine percent of the total variance in the observed measures. If not accounted for, this could lead to substantial reduction in model fit, due to the systematic nature of the variance introduced by the presence of reversed item bias.

In latent-factor models, latent factors represent systematic covariance among the indicators for the construct. This systematic covariance will also include common (systematic) method covariance, to the extent that it exists, which might be confounded with trait "true" variance (Podsakoff, MacKenzie, & Podsakoff, 2012). From the perspective that all items, *ceteris paribus*, should serve as equally good manifestations of the latent construct, differences in responses solely related to the keying of items would constitute a common method effect (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). This could lead to common method variance biasing inferences about group-level variance, implying heterogeneity in the trait of interest even if the trait "true" variance is homogeneous. Considering that response inconsistency and method effects related to reverse-keyed items (e.g. DiStefano & Motl, 2006) seem to affect respondents of low intelligence more severely (Sorjonen, Hemmingson & Melin, 2020), these biases may explain inadequate model fit for low intelligence groups. If these sources of method variance, extraneous to the true variance of the substantive latent factor of interest, contribute to group differences, the invariance testing may be falsely rejected (i.e., if common method factors are not taken into account).

1.4. Present study

In the present study, we tested the predictions of the PDIH as relating to the differences in latent trait-variances and trait-covariances within a MG-CFA analytical framework. We propose the following: If sufficient measurement invariance (MI) can be established, allowing for unbiased group comparisons, the PDIH predicts that latent variances and covariances based on a constrained MI model will differ between the two groups. The latent trait-variances are expected to be higher for the high-intelligence group and latent trait-covariances higher for the low-intelligence group. Furthermore, we will attempt to control for systematic method variance relating to reversed items, as ruling out one explanation of PDIH.

2. Method

2.1. Participants and procedure

We used data from the third wave of the longitudinal research project Health, Aging and Retirement Transitions in Sweden (HEARTS: see Lindwall et al., 2017). The study participants were initially recruited from the Swedish state personal address register (*Statens personadressregister*: SPAR). The sample contained a nationally representative sample of 14 990 individuals, born between 1949 and 1955. Of those initially recruited, 5913 (39.4%) responded and chose to participate in the first wave. For the present study we used the third wave sample of 2017, which was the first wave to include measures of personality and consisted of 4320 individuals (28.8% of those initially recruited). 1315 were excluded for the present study due to insufficient completion of relevant measures of personality or intelligence. All measures were self-administered and gathered online using Qualtrics. In terms of highest achieved educational background, of the remaining sample ($N = 3005$, 52% women), 46% had at least some university-level education and 41% secondary school/trade school/community college. Consequently, the sample consists of a higher proportion of individuals with a university background than what would be expected from a representative Swedish sample. Ethical approval was granted by the University of Gothenburg's

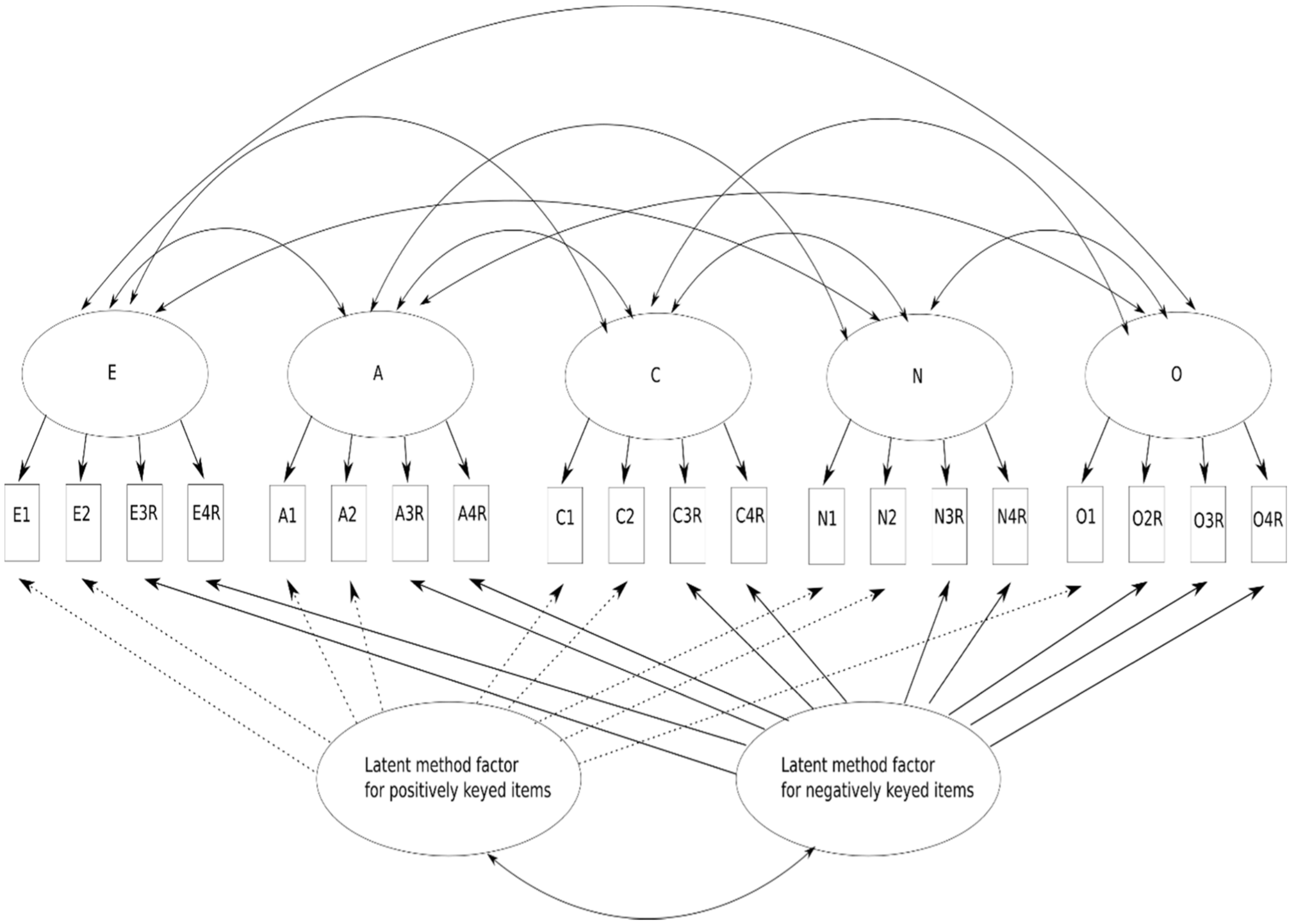

Fig. 1. Path Diagram Representation of the Big Five Model Including Two Method Factors. *Notes.* E = Extraversion. A = Agreeableness. C = Conscientiousness. N = Neuroticism. O = Openness. _R = negatively keyed item. Ovals represent latent factors and rectangles represent manifest indicators. Double-headed arrows signify covariance and one-headed arrows regression coefficient. Residuals (error terms) are omitted for clarity. The dotted arrows of positively keyed items is solely for clarity of interpretation in the graphical representation.

ethical approval board.

2.2. Measures

2.2.1. Personality

To measure personality, we used the Mini-IPIP (Donnelan, Oswald, Baird, & Lucas, 2006). The Mini-IPIP consists of 20 items, measured with a 5-point Likert scale, in which participants indicate degree of agreement with various statements. The scale consists of four items to measure each of the Big Five factors (Extraversion, Conscientiousness, Agreeableness, Openness, and Neuroticism). The scale has been found to have acceptable psychometric properties in terms of internal consistencies as well as discriminant, convergent and criterion validity (Donnellan et al., 2006). In the present sample, coefficient alphas for the five traits were 0.75, 0.55, 0.62, 0.62 and 0.59 for Extraversion, Conscientiousness, Agreeableness, Openness and Neuroticism, respectively, which were lower than the values obtained by Donnellan et al. (2006) of 0.77, 0.69, 0.70, 0.65 and 0.68.

2.2.2. Intelligence

For the measurement of intelligence, a 12-item version (Arthur & Day, 1994) of Raven Advanced Progressive Matrices (APM) was administered, which assessed intelligence by way of a series of progressively more difficult matrices. Participants were presented with a series of patterns and were then asked to choose the option which best completes it. There was a 3-minute time limit for completing all 12 items. The coefficient alpha for the present sample was 0.63, which was slightly lower than the internal consistency estimates obtained by Arthur and Day (1994) for the short form of $\alpha = 0.65$-0.69. The median composite score was 6 points and the mean score was 5.4 points (SD = 1.8), indicating a slight negative skewness (-0.28) of the sample data distribution. There were relatively fewer participants with high scores (only 35 participants scored 10 or higher) than with low scores (213 scored 2 or lower). This was likely in part attributable to the fact that the APM was designed to improve discrimination in the upper part of the ability distribution, therefore making it more suitable for samples of higher intelligence (Raven, 2000).

Table 1
Descriptive Statistics and Internal Reliabilities of Observed Personality Trait Scores without Controlling for Method-Factors.

Personality trait	E		A		C		N		O	
Intelligence	Low	High	Low	High	Low	High	Low	High	Low	High
Mean	3.27	3.16	3.91	3.89	4.07	3.96	2.32	2.18	3.35	3.61
SD	0.87	0.88	0.68	0.68	0.67	0.67	0.79	0.77	0.80	0.81
Reliability[a]	.73	.78	.59	.67	.51	.60	.55	.63	.59	.66

Note. Low (N = 1591), High (N = 940). The scores are based on ratings on a 1–5 Likert scale.
[a] Quantified by Cronbach's alpha.

Table 2
Measurement- and Construct-level Invariance Testing of the Big Five Model Including Two Latent Method Constructs.

Model	χ^2	*Df*	*CFI*	*RMSEA*	*90% CI RMSEA*	*Comparison*	*ΔCFI*	*ΔRMSEA*
Configural	720.773*	278	0.958	0.037	0.034-0.040	–	–	–
Metric	779.715*	293	0.953	0.038	0.035-0.041	vs. configural	-0.005	+0.001
Strong	800.601*	308	0.953	0.037	0.034-0.040	vs. metric	+0.001	- 0.001
Strict	874.430*	328	0.948	0.038	0.034-0.040	vs. strong	-0.005	+0.001
Latent Variances	880.338*	333	0.948	0.038	0.035-0.041	vs. strict	+/-0	+/-0
Latent Covariances	909.572*	342	0.946	0.038	0.035-0.041	vs. strict	-0.002	+0.0

Note. ΔCFI and ΔRMSEA refer to absolute change relative to a less constrained model. Latent covariances test was based on 9 degrees of freedom as one covariance (Agreeableness by Openness) was an outlier contributing to most of the misfit, and in the opposite direction of the tested hypothesis.
* $p < .001$.

Table 3
Latent Personality Trait Variances and Covariances/Correlations Based on Estimates from the Strict Invariance Model.

		E	A	C	N	O
E	High [a]	**0.662**	0.45	0.17	-0.17	0.26
	Low [b]	**0.683**	0.50	0.24	-0.28	0.30
	Δ	*0.021*	*0.05**	*0.07**	*0.11***	*0.04*
A	High	0.155	**0.181**	0.34	-0.17	0.55
	Low	0.181	**0.189**	0.37	-0.24	0.34
	Δ	*0.026**	*0.008*	*0.03*	*0.07*	*0.21****
C	High	0.097	0.102	**0.504**	-0.31	0.06
	Low	0.139	0.113	**0.487**	-0.38	0.01
	Δ	*0.042**	*0.011*	*0.017*	*0.07*	*0.05*
N	High	−0.062	−0.034	−0.099	**0.200**	-0.10
	Low	−0.111	−0.050	−0.129	**0.230**	-0.12
	Δ	*0.049***	*0.016*	*0.030*	*0.030*	*0.02*
O	High	0.145	0.161	0.029	−0.031	**0.469**
	Low	0.166	0.099	0.004	−0.039	**0.449**
	Δ	*0.021*	*0.062****	*0.025*	*0.008*	*0.020*

Note. E = Extraversion. A = Agreeableness. C = Conscientiousness. N = Neuroticism. O = Openness. Values on diagonals refer to latent variance (bolded), below diagonals to covariances and above diagonals to correlations. Δ refers to the absolute parameter difference between the two groups. Significance testing of group differences of each pair of latent trait-covariances was based on χ^2-difference tests with one degree of freedom between the strict invariance model and a model with equality constraints applied to the covariances between the two latent factors.
* $p < .05$ ** $p < .01$ *** $p < .001$.
[a] High-intelligence group ($n = 883$).
[b] Low-intelligence group ($n = 1433$).

2.2.3. Intelligence (high and low groups)

We split the participants into two groups based on their scores on the cognitive measure. Creating two groups for comparison has the benefit of making analyses more easily interpretable and enables efficient application of statistical techniques such as MG-CFA. Participants whose composite score on the cognitive measure was 6 (corresponding to the sample mode and median) were excluded from all analyses ($n = 689$, 22.9% of full sample), leaving one group of low ability (M = 4.00; SD = 1.11; $n = 1433$, 47.7% of initial total sample) and one of high ability (M = 7.57; SD = 0.80; $n = 883$, 29.3% of initial total sample), which served as the basis for intergroup comparisons in the analyses.

2.3. Statistical analyses

All statistical analyses were performed using the statistical software JASP (JASP Team, 2021). The modelling of confirmatory factor analyses was conducted through the SEM-module using lavaan syntax for model specification. We first fitted two separate CFA models to the total sample (i.e., both intelligence groups) data using different model specifications to assess whether the model with method factors, implemented to control for systematic method variance related to reversed items, improved the fit. We then conducted a multiple group confirmatory factor analysis (MG-CFA) to first test for incremental stages of measurement-level invariance between the high- and low-intelligence groups, followed by invariance testing of construct-level relations (latent variances and covariances)[1]. We obtained parameter estimates of latent variances and covariances based on the constrained, strict measurement-level model (i.e., the model with factor loadings, intercepts and residuals constrained to equality across groups). We used diagonally weighted least squares (DWLS) estimation for all CFA-models, rather than the more commonly used maximum likelihood (ML) method of estimation. This was due to the superiority of DWLS when estimating data based on ordered categorical data (such as measures based on Likert-scales), which deviate from assumptions of symmetry and multivariate normality associated with ML (Li, 2016). For testing of differences between groups in terms of personality trait covariances, we used χ^2-difference testing with one degree of freedom per parameter to assess statistical significance of applying equality constraints for each pair of personality traits (i.e., we constrained all possible inter-trait correlations to equality across groups one at a time, that is a total of 10 parameters).

2.3.1. Model specification

All five personality factors were modelled as latent (unmeasured) constructs, each with four loadings on their respective four indicators, i.e. the four items in the inventory intended to measure the factor (see Fig. 1).

One statistical remedy for problems related to inflated common method variance is to add a method factor "*…whose only measures are the indicators of the theoretical constructs of interest that share a common*

[1] See Little (1997) for a more in-depth discussion of the relation between measurement-level invariance testing and construct-level invariance testing.

method" (Podsakoff et al., 2012, p. 553). In an attempt to control for systematic method variance, we added two latent method factors to the five-factor personality model. One method factor was specified to have cross-loadings on all eleven reversed items (two for each personality factor except for Openness, which consisted of three reversed items), and the other method factor was specified to have cross-loadings on all nine positively keyed (non-reversed) items. This approach is commonly referred to in the literature as correlated trait, correlated method (CTCM) and has been shown to yield substantial improvements in model fit with other measures in the past (Quilty, Oakman, & Risko, 2007). Consequently, the full model (see Fig. 1) consisted of seven latent factors: five substantive (personality) factors, each with their own four indicators, and two method factors, one for reversed items and the other for non-reversed items. The five substantive factors were allowed to covary freely and the two method factors to covary with one another. The covariation between the method factors and the substantive factors were fixed to zero. Given that the method factors did not have their own unique indicators associated with them, they correspond to what is commonly referred to in the literature as unmeasured latent method constructs (ULMC), to separate this method from other statistical techniques intended for partialling out method variance (e.g. Johnson, Rosen & Djurdjevic, 2011; Richardson, Simmering & Sturman, 2009; Williams & O'Boyle, 2015).

2.3.2. Measurement invariance

Measurement invariance (MI) testing in a MG-CFA framework may work by starting with an unconstrained model and then step-by-step imposing constraints in the model, and assessing the change in model fit for each step. A sizeable decrease in model fit when imposing new constraints indicates that MI does not hold. If the population parameter values are identical for the two groups, constraining the parameter value to equality should not lead to any substantial degradation in model fit. The larger the discrepancy in the values of the parameter, the worse the constrained model will fit the data (i.e., the poorer the model-implied data matrix will match the observed data matrix). Stages and sequences involved in the testing of MI may differ across studies, both in terms of preferred nomenclature and the number of recommended steps and constraints. In the next paragraph, we will describe the steps (models) that we used in the present study.

We first assessed *configural*, or *form*, invariance to examine whether the overall factor structure was equivalent across groups. We then tested for *metric*, or *weak* (e.g., Morin, Madore, Morizot, Boudrias & Tremblay, 2009) invariance (sometimes also called factorial invariance, e.g., Cheung & Rensvold, 2000), to examine whether factor loadings between items and their latent constructs were equal across groups. Metric invariance differs from configural invariance in the sense that configural variance tests whether items are associated with the same latent factors across groups, while metric invariance tests whether *the strength* of the association between items and their latent factors (i.e., the factor loadings) are equal. Establishing metric invariance (equivalence of factor loadings) is necessary for unbiased comparisons of factor covariances and variances at the construct-level between groups (Liu, Millsap, West, Tein, & Tanaka, 2017), which is the primary purpose of this study. With the inclusion of latent method constructs, the practice of applying omnibus constraints to factor loadings, i.e., constraining all factor loadings, would have been problematic. This is because the model then presupposes that the invariance in factor loadings for the method factors between the two groups ought to be invariant as well, while our rationale for including them was that they might be an important source of confounding non-invariance between the two groups. Consequently, we only applied equality constraints to the factor loadings of the substantive personality factors while allowing the factor loadings of the method factors to vary freely between the groups. We then assessed *scalar* invariance, in which both factor loadings as well as intercepts were constrained to equality, to examine whether the item-level intercepts were invariant across groups, or whether the raw score on items that equal a certain value on the latent construct are the same for participants in both groups (Cheung & Rensvold, 2000). As the final stage of invariance testing at the measurement-level, we applied further constraints to residual variances and covariances to examine *strict* invariance (Van de Schoot, Lugtig & Hox, 2012), or *residual* invariance (Chen, 2007). This step serves to measure whether items function as equally good indicators of the latent construct across groups (Cheung & Rensvold, 2002). Residual variance refers to all item-level variance that is not shared with other items belonging to the same hypothesised latent factor, such as variance due to unsystematic measurement error.[2] Therefore, strict invariance holds if scale items measure the latent construct with the same degree of measurement error across groups (Cheung & Rensvold, 2002). After conducting the first four principal stages of MI-testing, we then investigated group invariance at the construct-level by applying further constraints to the latent variances followed by the latent covariances across the personality factors.

2.3.3. Assessing model fit

For the purposes of examining model fit, we report χ^2, Comparative Fit Index (CFI: Bentler, 1990), and Root Mean Square Error of Approximation (RMSEA: Browne & Cudeck, 1992). The χ^2 test gives the probability of the data given the null hypothesis, stating that the model-implied covariance matrix and the observed covariance matrix are identical, with larger values thus indicating poor model fit. This test is highly sensitive to sample size, leading to trivial discrepancies between the hypothesised and observed matrices resulting in significance and thereby model rejection, given a large enough sample (Cheung & Rensvold, 2002). The Comparative Fit Index is an incremental fit index which assesses the superiority of the fitted model over a null model in terms of more closely mirroring the observed variance–covariance matrix and ranges from 0 to 1, with higher values indicating better model fit. RMSEA is an absolute fit index measuring the degree of deviation of the model-implied covariance matrix from the observed covariance matrix per degree of freedom with values closer to 0 being optimal (Chen, 2007). Both CFI and RMSEA are based upon the χ^2-test but do not share its sensitivity to sample size (Cheung & Rensvold, 2002).

For assessing violations of invariance, we used the guidelines of Cheung and Rensvold (2002) of change in CFI (ΔCFI) of more than 0.01, which is also recommended by Chen (2007) as long as sample sizes are large and not too unequal, supplemented by examining changes in RMSEA, using the more stringent criteria recommended by Chen (2007) of ΔRMSEA greater than 0.010. For assessing construct-level invariance (latent variances and covariances) we also used χ^2-difference tests.

3. Results

Means, standard deviations and internal consistency estimates based on composite observed scores are presented in Table 1. The coefficient alphas for the five factors were consistently larger for the high-ability group, in line with most previous findings, with the biggest difference in Conscientiousness (difference of 0.09) and the smallest for Extraversion (difference of 0.05). Differences in alpha between the two groups were all statistically significant: Extraversion ($\chi^2(1) = 7.28$, $p < .01$), Agreeableness ($\chi^2(1) = 8.17$, $p < .01$), Conscientiousness ($\chi^2(1) = 7.15$, $p < .01$), Neuroticism ($\chi^2(1) = 6.66$, $p < .01$), and Openness ($\chi^2(1) =$

[2] This communality conception treats all unsystematic residual variance as measurement error, even though personality items (nuances) contain specific variance, with its own stability (t-t reliability but not split-half reliability, i.e. coefficient alpha) and predictive validity, which is likely inappropriately classified as measurement error (see McCrae, 2015; Möttus, 2016).

6.09, $p < .05$).

To test whether the inclusion of the method factors improved the Big Five model fit, we compared two CFA models. One with only the five personality factors and one with the two method factors added (see Fig. 1). Both CFAs were run using both the low and high-ability groups together. The model without the method factors displayed a relatively poorer fit, χ^2 (160) = 1784.244, $p < .001$; CFI = 0.883; RMSEA = 0.058, 90% CI[0.56, 0.061], in comparison to the model including the method factors, χ^2 (139) = 831.415, $p < .001$; CFI = 0.950; RMSEA = 0.041, 90% CI[0.038, 0.043][3], indicating that the inclusion of method factors led to a model that more closely resembled the observed variance/covariance matrix, χ^2_{diff} (21) = 952.83, $p < .001$.[4]

The results of the measurement invariance testing based on the model with method factors included are presented in Table 2. The configural model displayed good fit, χ^2 (278) = 720.773, $p < .001$; CFI = 0.958; RMSEA = 0.037, 90% CI[0.34, 0.40], indicating that the overall factor structure was similar in the two groups. The metric model with equality constraints applied to the factor loadings of the substantive personality factors showed minor deterioration of model fit compared to the configural model, but well below the cut off-criteria for rejecting invariance, in terms of changes in both CFI (Δ-0.005) as well as RMSEA (Δ + 0.001). If we applied equality constraints to the factor loadings of the method factors as well in the process of testing for metric invariance, the deterioration in model fit (ΔCFI = -0.011) would have led to the rejection of metric invariance based on the adopted cut-off criterium. This suggested that the factor loadings of the method factors were a larger source of non-invariance between groups than the loadings of the personality factors. Both the model testing of strong invariance, with constraints added to the intercepts (χ^2_{diff} (15) = 20.89, $p = .14$), and the model testing of strict invariance (χ^2_{diff} (20) = 73.83, $p < .001$)[5], with further constraints added to the residuals, exhibited no (for the strong invariance model) or minor (for the strict invariance model) deterioration of model fit, based on CFI and RMSEA. In terms of establishing the measurement invariance of personality by intelligence in the present sample, invariance seems to hold overall, even at the level of strict invariance.

The addition of equality constraints to the latent variances did not lead to any marked deterioration in model fit based on CFI or RMSEA, indicating that the latent variances did not differ noticeably between the groups. This was further corroborated by a χ^2-difference test for the model with latent variances constrained resulting in non-significance, χ^2_{diff} (5) = 5.91, $p = .32$. Adding equality constraints to the factor covariances resulted in a worsening of model fit. Particularly, it was the Agreeableness by Openness covariance that differed across groups, and in the opposite direction hypothesised (see Table 3). Constraining the remaining nine covariances across the personality factors to group equality resulted in only minor change in fit, as quantified by CFI or RMSEA, the χ^2-difference test was however significant (χ^2_{diff}(9) = 29.23, $p = .001$).

For examination of covariances and variances of the substantive latent factors at a group-level, the estimated latent variances and covariances of the personality factors are presented in Table 3. As expected, based on the lack of change in model fit indices when equality constraints were placed on the latent variances, group differences in latent variances were small overall. Furthermore, for three out of the five personality factors, the low-intelligence group were the one to display higher variance, in opposition to the hypothesised direction. Consequently, the hypothesised difference in trait-variance given differences in intelligence was rejected.

In terms of the latent covariances, differences between groups were more pronounced, as hinted at by the relatively larger deterioration in model fit when applying constraints to the covariances. Moreover, most of the differences in covariances were in the hypothesised direction of the high-intelligence group displaying lower trait covariances, with the two exceptions of the covariances of Openness with Agreeableness and Conscientiousness, where the covariances were higher for the high-intelligence group. However, while constraining all latent covariances to equality resulted in significant differences between groups, only four out of ten group differences in covariances based on χ^2-difference tests with one degree of freedom were significant. We further conducted sensitivity analyses in order to examine the gradient of effects when retaining the middle group in the analysis (see Appendix I), in terms of the latent factor variances and covariances. This did not result in appreciable differences compared to the two-group analysis. Taken together, these findings indicate minor support at best for the hypothesised difference in covariance across personality traits with differing levels of intelligence.

4. Discussion

In the present study, we set out to test the PDIH predictions concerning hypothesised differences in variances and covariances of personality traits depending on levels of intelligence. The results of the invariance testing suggested comparable measurement models across groups, thereby allowing valid inferences of the personality trait-variance and covariance group differences. We also added two ULMCs to the five-factor solution, one with loadings on negatively and one on positively keyed items. As expected, given previous examinations of such approaches (e.g., Quilty et al., 2006), this led to substantial improvement in model fit, compared to the model consisting only of the substantive personality factors.

Austin et al. (2000) noted that increased scale reliability with higher intelligence would be an expected result in line with the PDIH, and several studies have found this effect (e.g., Austin et al., 1997; Allik et al., 2004; Möttus et al., 2007; Schermer, Bratko, & Bojic, 2020). We also found this effect in the HEARTS data. On the other hand, Sorjonen et al. (2020) found that low intelligence was associated with particular difficulty in handling reversed items, which may act as a confounding factor contributing to lower reliability for lower intelligence individuals when responding to inventories containing reversed items. Seeing the magnitude of the factor loadings as a measure of the reliability of the latent factors (with an item's squared factor loading equivalent to "true" variance; Shevlin, Miles, Davis, & Walker, 2000), our results indicate that when controlling for systematic method variance, related to reversed items, the two ability groups demonstrated approximately similar levels of factor score reliability (as indicated and tested by the small change in model fit when applying constraints).

Establishment of strict invariance meant that our findings were aligned with those of Waiyavutti et al. (2012) and De Fruyt et al. (2006), while standing in contradiction to those of McLarnon and Carswell (2013), who found that all personality factors except one failed to demonstrate either strong or strict invariance by intelligence. Possible reasons for these differences are many. For example, McLarnon and Carswell (2013) used a different measure of personality, the Six Factor Personality Questionnaire, which is derived from a similar model to the

[3] See Appendix II for factor loading matrices.

[4] We used ML-estimation solely to derive estimates of AIC and BIC for the two models (all other indices are based on DWLS). The first model had AIC = 172139.025 and the second AIC = 171009.090 (Δ1129.935). For BIC, the first model had BIC = 172439.427 and the second 171435.660 (Δ1003.767). The lower information criteria values of the second model (with method factors) indicate a better trade-off between model complexity and model fit (Van de Schoot et al., 2012).

[5] See Appendix II for factor loading matrices for the strict invariance model.

Big Five factor structure but splits the Conscientiousness factor into two factors (methodicalness and industriousness). It is difficult to rule out the possible influences of such differing inventories which, despite purportedly measuring similar personality structures, might display differential measurement invariance by intelligence. Another difference was our inclusion of method factors to control for systematic method variance, which may have contributed to non-invariance in personality in earlier studies. Notably, had we instead applied omnibus constraints to the factor loadings in the process of testing for metric invariance (i.e., added equality constraints to the loadings of the method factors as well) the drop in CFI compared to the configural model would have increased from 0.005 to 0.011. This has two implications. First, it would have led us to (falsely) reject metric-level invariance. Second, it suggests that group differences in the loadings of the method factors were a larger source of non-invariance than those of the personality factors.

We found no support for the PDIH's prediction of larger trait-variances in personality factors associated with higher intelligence, unlike many previous studies (e.g., Austin et al., 1997; De Fruyt et al., 2006; Harris et al., 2005; Schermer, Bratko, & Bojic, 2020). Not only were the overall differences in variance between the groups of little magnitude, as reflected by the lack of change in fit indices when applying equality constraints to the latent variances, but for three out of the five factors the low-intelligence group had higher variance in personality, i.e., the opposite direction as expected by the hypothesis.

We found some support for the PDIH's prediction of lower covariance between traits for high-intelligence participants, as reflecting a higher degree of differentiation. This was hinted at by worsening of model fit when equality constraints were applied to the latent covariances as well as the highly significant χ^2-difference test. Moreover, all except two of the trait covariances were in the hypothesised direction (see Table 2). On the other hand, only four out of the ten differences in covariances between the two groups were statistically significant when tested pair for pair, despite the large sample size. We also note that the largest differences (i.e., across Agreeableness and Openness) was in the opposite direction from the PDIH prediction. Furthermore, while applying constraints to the latent covariances did result in some decrease in model fit indices, these did not exceed traditional cut-off criteria for rejecting invariance. As for the lower trait covariance among high-intelligence individuals being emblematic of a more multi-factorial structure with increasing intelligence (e.g. Austin et al., 1997; 2000), any major differences in the basic factor structure of personality between the two groups would likely have been reflected by poor fit for the unconstrained, configural model of invariance testing.

5. Limitations

The sample in our study consisted solely of individuals aged between 62 and 68, which should engender a degree of caution when attempting to generalise findings to other age-segments of the population. While some mean changes in personality over lifetime are well documented (e.g., Graham et al., 2020; Soto, John, Gosling, & Potter, 2011), as personality differences tend to be generalisable across the life-span (e.g., Anusic & Schimmack, 2016; Gnambs, 2014), we do not believe this should threaten the validity of our general findings. Furthermore, our study used a relatively short personality inventory, the mini-IPIP (Donnelan et al., 2006), which consists of only 4 items per factor, and a relatively short cognitive measure (Arthur & Day, 1994) with a total of 12 items. Both measures exhibited low internal consistency estimates, which might be expected in part given their brevity. The internal consistency estimates in our study were slightly lower than those obtained by Donnellan et al. (2006) and Arthur and Day (1994). However, the use of a confirmatory factor analytical framework should mitigate these issues by yielding valid and unbiased estimates of construct interrelations despite low internal consistency[6](Little et al., 1999). This is especially true with the large sample size, resulting in relatively smaller standard errors and estimates of variance and covariance components that are more precise. Moreover, the use of a short personality inventory enabled us to model and perform invariance testing on the full personality structure, i.e., all five traits and their respective indicators simultaneously, without having to limit the invariance testing to one isolated personality factor at a time, or to resort to parcelling items, which is generally not recommended (Marsh, Lüdtke, Nagengast, Morin, & Von Davier, 2013; Meade & Kroustalis, 2006).

Furthermore, we split the sample into two groups based on the cognitive measure, even though dichotomising continuous variables typically leads to a loss of power (Cohen, 1983); however, the large sample size should mitigate this issue. To avoid the similarity of groups that ensues from splitting a normally distributed variable in the middle (resulting in a large portion of both groups being similar on the variable of interest), we created more extreme groups for comparison by removing all participants whose composite score on the cognitive measure corresponded to the sample mode and median. This resulted in groups that were more distinct from one another on the variable of interest (intelligence), at the cost of a smaller sample. Moreover, given the slight negative skew in the data with more participants clustered on the lower side of the distribution, this led to the low-intelligence group being larger than the high-intelligence group. This may have affected the results as non-invariance is more likely to be undetected with unequal sample sizes (Chen, 2007). However, the uneven sample size ratio in our study (1.62:1) was smaller than those in Chen's (2007) simulations of uneven sample sizes (2:1 or 3:1) and the total sample size (both groups) in our study ($n = 2316$) was larger than the largest simulated sample sizes of Chen (2007) of a total of 1000. Moreover, additional sensitivity analyses to test the robustness of effects, based on a three-group split with the middle-group retained (see Appendix I), did not show differences of a kind that would suggest an alternative interpretation of the data.

It is also worth considering whether different types of intelligence-measures contribute to differences between our findings and those of others. McLarnon and Carswell (2013) highlighted Raven's Progressive Matrices, and that this may have influenced findings of invariance of personality by intelligence, pointing out that responding to a self-report questionnaire involves mostly verbal skills. Other studies which found support for differences in personality measures by intelligence mainly used a combined aggregated measure of multiple aspects of intelligence (e.g., Austin et al., 1997; Harris et al., 2005; McLarnon & Carswell, 2013). Our study, like that of De Fruyt et al. (2006), only used a measure of fluid intelligence, which may assess a facet of intelligence which impacts personality measures less than one using a verbal subscale. Our results may thus not be generalisable to studies using different measures of intelligence. Future research may want to examine to what degree different facets of intelligence, e.g., verbal and fluid, affect the interaction between personality and intelligence.

5.1. Conclusion

We found scant support for the PDIH in our sample, regarding differences in variance and covariance structure of personality with different levels of intelligence. The method factors included to control for systematic method variance were differentially associated with

[6] At least if the latent variables are considered reflective, as in our analyses, rather than formative (for further discussion of reflective and formative conceptions of latent variables, see Borsboom, Mellenbergh, & van Heerden, 2003; Edwards & Bagozzi, 2000).

intelligence, suggesting an effect of intelligence on the use and understanding of reversed items. Taken together, this suggests that controlling for common method effects is of importance when making judgements regarding differences in personality structure between high- and low-intelligence groups.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The analyses were based on data from the research project HEARTS at the University of Gothenburg, financed by the Swedish Research Council for Health, Working Life and Welfare (FORTE; Dnr: 2013-2291). Work on this paper was supported by grants from FORTE to the Centre of Ageing and Health (AGECAP, 2013-2300), Riksbankens Jubileumsfond (RJ P21-0112), and from the Swedish Research Council (VR) for the National E-Infrastructure for Ageing Research (NEAR).

Author contributions

The first author drafted the manuscript. The first and third author were responsible for the study design, concept, development and data analysis. The second and third author revised the manuscript. All authors contributed to the final version of the manuscript.

Appendix

Appendix I

Estimates from sensitivity analyses of latent factor variances and covariances when retaining the middle-group and fitting a three-group unconstrained solution.

Table A1.

Table A1

		E	A	C	N	O
E	Low[a]	*0.693*				
	Mid[b]	*0.606*				
	High[c]	*0.665*				
A	Low	0.193	*0.205*			
	Mid	0.172	*0.182*			
	High	0.149	*0.181*			
C	Low	0.146	0.126	*0.510*		
	Mid	0.098	0.073	*0.362*		
	High	0.089	0.097	*0.449*		
N	Low	−0.097	−0.046	−0.111	*0.162*	
	Mid	−0.135	−0.065	−0.132	*0.316*	
	High	−0.080	−0.048	−0.127	*0.319*	
O	Low	0.134	0.084	−0.011	−0.026	*0.280*
	Mid	0.179	0.141	0.010	−0.085	*0.401*
	High	0.155	0.174	0.034	−0.075	*0.696*

Note. Latent variances in diagonal (italicised).
[a] Low-intelligence group (n = 1433).
[b] Mid-intelligence group (n = 689).
[c] High-intelligence group (n = 883).

Appendix II

Factor loading estimates from an unconstrained model, χ^2 (139) = 831.415, p <.001; CFI = 0.950; RMSEA = 0.041, 90% CI[0.38, 0.43], with method factors, fitted to the data from the full sample (see Tables B1 and B2). Below (see Tables B3 and B4) are also estimates derived from a model with strict group invariance constraint, χ^2 (328) = 874.430, p <.001; CFI = 0.948; RMSEA = 0.038, 90% CI[0.35, 0.41].

Table B1
Parameter Estimates of Indicators Factor Loadings and Variances/Residuals from an Unconstrained Model with Method Factors Fitted to the Total Sample (N = 3005).

Parameter	Factor loadings		Variances/Residuals	
	Est.	SE	Est.	SE
E1	1.00	n/a	0.53***	0.04
E2R	0.75***	0.03	1.04***	0.03
E3	1.12***	0.04	0.46***	0.05
E4R	0.78***	0.03	0.84***	0.04
A1	1.00	n/a	0.39***	0.02
A2R	0.99***	0.05	0.95***	0.03
A3	1.38***	0.06	0.58***	0.03
A4R	1.28***	0.06	0.70***	0.04
C1	1.00	n/a	0.80***	0.04
C2R	0.70***	0.04	0.93***	0.04
C3	0.75***	0.04	0.53***	0.03
C4R	0.52***	0.04	0.75***	0.04
N1	1.00	n/a	0.58***	0.05
N2R	1.53***	0.12	0.77***	0.05
N3	1.07***	0.07	0.52***	0.05
N4R	1.48***	0.12	1.50***	0.05
O1R	1.00	n/a	0.77***	0.04
O2	1.10***	0.07	0.98***	0.06
O3R	0.98***	0.05	0.99***	0.04
O4R	1.14***	0.06	0.68***	0.04
ME2R	1.00	n/a		
ME4R	2.20***	0.58		
MA2R	1.98***	0.53		
MA4R	2.72***	0.71		
MC2R	2.99***	0.79		
MC4R	1.58***	0.44		
MN2R	0.65*	0.31		
MN4R	0.38	0.33		
MO1R	3.55***	0.92		
M03R	3.04***	0.80		
M04R	0.31	0.26		
ME1	1.00	n/a		
ME3	0.20	0.15		
MA1	0.42	0.23		
MA3	5.04***	1.20		
MC1	4.20***	0.95		
MC3	0.43	0.23		
MN1	1.36***	0.34		
MN3	1.24***	0.31		
M02	5.22***	1.24		

Note. SE = Standard Error.
* p <.05 ** p <.01 *** p <.001.

Table B2
Parameter Estimates of Latent Variables (Covariances and Variances) from an Unconstrained Model with Method Factors Fitted to the Total Sample (N = 3005).

		Covariances (variances in diagonal)						
		E	A	C	N	O	MR	M
E	Est.	*0.69****						
	SE	0.03						
A	Est.	0.18***	*0.20****					
	SE	0.008	0.01					
C	Est.	0.12***	0.10***	*0.46****				
	SE	0.009	0.007	0.03				
N	Est.	−0.09***	−0.04***	−0.11***	*0.20****			
	SE	0.008	0.005	0.01	0.02			
O	Est.	0.14***	0.11***	−0.01	−0.04***	*0.36****		
	SE	0.009	0.007	0.008	0.007	0.03		
MR	Est.	–	–	–	–	–	*0.01**	
	SE	–	–	–	–	–	0.007	
M	Est.	–	–	–	–	–	−0.006**	*0.01**
	SE	–	–	–	–	–	0.002	0.006

Note. MR = Method factor for reverse-keyed items. M = Method factor for positively keyed items, SE = Standard Error. Covariances between method factors and personality factors were fixed to zero.
* *p* <.05 ** *p* <.01 *** *p* <.001.

Table B3
Parameter Estimates of Constrained Intercepts, Factor Loadings and Variances/Residuals from a Strict Invariance Model with Method Factors for the High and Low Ability Groups (N = 2316).

Parameter	Factor loadings		Intercepts		Variances/Residuals	
	Est.	SE	Est.	SE	Est.	SE
E1	1.00	n/a	0.00	n/a	0.56***	0.04
E2R	0.77***	0.03	1.44***	0.10	1.03***	0.04
E3	1.08***	0.04	0.33**	0.12	0.47***	0.05
E4R	0.82***	0.04	1.12***	0.10	0.83***	0.04
A1	1.00	n/a	0.00	n/a	0.40***	0.02
A2R	1.11***	0.06	−0.74**	0.25	0.96***	0.04
A3	1.38***	0.07	−2.23***	0.27	0.59***	0.04
A4R	1.38***	0.07	−1.66***	0.29	0.71***	0.04
C1	1.00	n/a	0.00	n/a	0.75***	0.05
C2R	0.67***	0.05	1.81***	0.18	0.87***	0.05
C3	0.66***	0.05	1.83***	0.16	0.55***	0.03
C4R	0.45***	0.04	2.93***	0.13	0.73***	0.04
N1	1.00	n/a	0.00	n/a	0.61***	0.05
N2R	1.45***	0.13	−0.50	0.25	0.79***	0.06
N3	1.11***	0.08	−0.02	0.16	0.56***	0.05
N4R	1.39***	0.12	−0.04	0.24	1.48***	0.06
O1R	1.00	n/a	0.00	n/a	0.75***	0.05
O2	0.83***	0.06	−0.28	0.22	1.01***	0.06
O3R	0.99***	0.05	−0.12	0.20	0.97***	0.05
O4R	0.94***	0.05	0.48*	0.20	0.74***	0.04

Method Factor Loadings	High[a]		Low[b]	
	Est.	SE	Est.	SE
ME2R	1.00	n/a	1.00	n/a
ME4R	−3.26	2.51	2.75**	0.97
MA2R	−1.94	1.66	1.88**	0.69
MA4R	−1.99	1.61	1.67**	0.64
MC2R	−2.53	1.99	2.49**	0.88
MC4R	−4.66	3.50	3.95**	1.36
MN2R	−4.08	3.05	2.08**	0.75
MN4R	0.13	0.85	1.40*	0.59
MO1R	2.83	2.30	1.93*	0.77
M03R	−1.17	1.17	2.24**	0.82
M04R	1.99	1.58	0.54	0.36
ME1	1.00	n/a	1.00	n/a
ME3	−0.58	0.57	0.31**	0.11
MA1	−0.17	0.67	0.64***	0.18
MA3	6.82	4.30	2.64***	0.46
MC1	7.90	4.89	2.53***	0.43
MC3	−1.50	1.20	0.93***	0.20
MN1	0.12	0.55	1.27***	0.23
MN3	0.38	0.55	0.62***	0.15
M02	6.15	3.89	2.69***	0.47

Note. SE = Standard Error. Group specific values are only given for parameters unconstrained across groups (i.e., method factor loadings).
* *p* <.05 ** *p* <.01 *** *p* <.001.
[a] High-intelligence group (*n* = 883). [b]Low-intelligence group (*n* = 1433).

Table B4
Parameter Estimates of Latent Variables (Covariances and Variances) from a Constrained, Strict Invariance Model with Method Factors for the High and Low Ability Group, Respectively.

		Covariances (variances in diagonal)						
		E	A	C	N	O	MR	M
E	High[a]	*0.66*** (0.04)*						
	Low[b]	*0.69*** (0.04)*						
A	High	0.15*** (0.01)	*0.18*** (0.01)*					
	Low	0.18*** (0.01)	*0.19*** (0.01)*					
C	High	0.10*** (0.02)	0.10*** (0.01)	*0.51*** (0.05)*				
	Low	0.14*** (0.01)	0.11*** (0.01)	*0.50*** (0.05)*				
N	High	−0.06*** (0.01)	−0.03*** (0.008)	−0.10*** (0.02)	*0.19*** (0.03)*			
	Low	−0.11*** (0.01)	−0.05*** (0.007)	−0.13*** (0.01)	*0.22*** (0.03)*			
O	High	0.14*** (0.01)	0.16*** (0.01)	0.02 (0.008)	−0.03* (0.01)	*0.47*** (0.04)*		
	Low	0.17*** (0.01)	0.10*** (0.01)	0.002 (0.014)	−0.04*** (0.01)	*0.45*** (0.04)*		
MR	High	–	–	–	–	–	*0.003 (0.005)*	
	Low	–	–	–	–	–	*0.01 (0.01)*	
M	High	–	–	–	–	–	0.005 (0.005)	*0.005 (0.006)*
	Low	–	–	–	–	–	−0.01** (0.005)	*0.05** (0.01)*

Note. Standard Errors given in parentheses. MR = Method factor for reverse-keyed items. M = Method factor for positively keyed items. Covariances between method factors and personality factors were fixed to zero.
* $p < .05$ ** $p < .01$ *** $p < .001$.
[a] High-intelligence group ($n = 883$).
[b] Low-intelligence group ($n = 1433$).

References

Abad, F. J., Colom, R., Juan-Espinosa, M., & Garcıa, L. F. (2003). Intelligence differentiation in adult samples. *Intelligence, 31*(2), 157–166. https://doi.org/10.1016/S0160-2896(02)00141-1

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*(6), 445–462. https://doi.org/10.1002/per.524

Anusic, I., & Schimmack, U. (2016). Stability and change of personality traits, self-esteem, and well-being: Introducing the meta-analytic stability and change model of retest correlations. *Journal of Personality and Social Psychology, 110*(5), 766–781.

Arthur, W., Jr, & Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices Test. *Educational and Psychological measurement, 54*(2), 394–403.

Austin, E. J., Deary, I. J., & Gibson, G. J. (1997). Relationships between ability and personality: Three hypotheses tested. *Intelligence, 25*(1), 49–70. https://doi.org/10.1016/S0160-2896(97)90007-6

Austin, E. J., Hofer, S. M., Deary, I. J., & Eber, H. W. (2000). Interactions between intelligence and personality: Results from two large samples. *Personality and Individual Differences, 29*(3), 405–427. https://doi.org/10.1016/S0191-8869(99)00202-0

Austin, E. J., Deary, I. J., Whiteman, M. C., Fowkes, F. G. R., Pedersen, N. L., Rabbitt, P., … McInnes, L. (2002). Relationships between ability and personality: Does intelligence contribute positively to personal and social adjustment? *Personality and Individual Differences, 32*(8), 1391–1411. https://doi.org/10.1016/S0191-8869(01)00129-5

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Blum, D., & Holling, H. (2017). Spearman's law of diminishing returns. A meta-analysis. *Intelligence, 65*, 60–66. https://doi.org/10.1016/j.intell.2017.07.004

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*(2), 203–219. https://doi.org/10.1037/0033-295X.110.2.203

Brand, C., Egan, V., & Deary, I. J. (1994). *Intelligence, personality, and society: Constructivist versus essentialist possibilities*. In D. K. Detterman (Vol. Ed.), Current topics in human intelligence: 4, (pp. 29–42). Norwood, NJ: Ablex Theories of Intelligence.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*(2), 230-258. https://doi.org/10.1177%2F0049124192021002005.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology, 31*(2), 187–212. https://doi.org/10.1177%2F0022022100031002003.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*(3), 249–253. https://doi.org/10.1177%2F014662168300700301.

Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence, 23*(2), 105–132. https://doi.org/10.1016/S0160-2896(96)90008-2

De Fruyt, F., Aluja, A., García, L. F., Rolland, J. P., & Jung, S. C. (2006). Positive presentation management and intelligence and the personality differentiation by intelligence hypothesis in job applicants. *International Journal of Selection and Assessment, 14*(2), 101–112. https://doi.org/10.1111/j.1468-2389.2006.00337.x

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling: A Multidisciplinary Journal, 13*(3), 440–464. https://doi.org/10.1207/s15328007sem1303_6

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological assessment, 18*(2), 192–203. https://psycnet.apa.org/doi/10.1037/1040-3590.18.2.192.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods, 5*(2), 155–174. https://psycnet.apa.org/doi/10.1037/1082-989X.5.2.155.

Escorial, S., Navarro-González, D., Ferrando, P. J., & Vigil-Colet, A. (2019). Is individual reliability responsible for the differences in personality differentiation across ability levels? *Personality and Individual Differences, 139*, 331–336. https://doi.org/10.1016/j.paid.2018.12.004

Gnambs, T. (2014). A meta-analysis of dependability coefficients (test–retest reliabilities) for measures of the Big Five. *Journal of Research in Personality, 52*, 20–28.

Graham, E. K., Weston, S. J., Gerstorf, D., Yoneda, T. B., Booth, T., Beam, C. R., … Mroczek, D. K. (2020). Trajectories of Big Five Personality Traits: A Coordinated Analysis of 16 Longitudinal Samples. *European Journal of Personality, 34*(3), 301–321. https://doi.org/10.1002/per.2259

Harris, J. A., Vernon, P. A., & Jang, K. L. (2005). Testing the differentiation of personality by intelligence hypothesis. *Personality and Individual Differences, 38*(2), 277–286. https://doi.org/10.1016/j.paid.2004.04.007

JASP Team (2021). JASP (Version 0.16.1)[Computer software].

Johnson, R. E., Rosen, C. C., & Djurdjevic, E. (2011). Assessing the impact of common method variance on higher order multidimensional constructs. *Journal of Applied Psychology, 96*(4), 744–761. https://doi.org/10.1037/a0021504

LaHuis, D. M., Barnes, T., Hakoyama, S., Blackmore, C., & Hartman, M. J. (2017). Measuring traitedness with person reliabilities parameters. *Personality and Individual Differences, 109*, 111–116. https://doi.org/10.1016/j.paid.2016.12.034

Li, C. H. (2016). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods, 21*(3), 369–387. https://doi.org/10.1037/met0000093

Lindwall, M., Berg, A. I., Bjälkebring, P., Buratti, S., Hansson, I., Hassing, L., … Johansson, B. (2017). Psychological health in the retirement transition: Rationale and first findings in the Health, Ageing and Retirement Transitions in Sweden (HEARTS) study. *Frontiers in Psychology, 8*, 1634. https://doi.org/10.3389/fpsyg.2017.01634

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*(1), 53–76. https://doi.org/10.1207/s15327906mbr3201_3

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When" good" indicators are bad and" bad" indicators are good. *Psychological Methods, 4*(2), 192–211. https://doi.org/10.1037/1082-989X.4.2.192

Liu, Y., Millsap, R. E., West, S. G., Tein, J. Y., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods, 22*(3), 486–506. https://doi.org/10.1037/met0000075

Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J., & Von Davier, M. (2013). Why item parcels are (almost) never appropriate: Two wrongs do not make a right—Camouflaging misspecification with item parcels in CFA models. *Psychological Methods, 18*(3), 257–284. https://doi.org/10.1037/a0032773

McCrae, R. R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review, 19*(2), 97–112. https://doi.org/10.1177%2F1088868314541857.

McLarnon, M. J., & Carswell, J. J. (2013). The personality differentiation by intelligence hypothesis: A measurement invariance investigation. *Personality and Individual Differences, 54*(5), 557–561. https://doi.org/10.1016/j.paid.2012.10.029

Meade, A. W., & Kroustalis, C. M. (2006). Problems with item parceling for confirmatory factor analytic tests of measurement invariance. *Organizational Research Methods, 9*(3), 369–403. https://doi.org/10.1177%2F1094428105283384.

Morin, A. J., Madore, I., Morizot, J., Boudrias, J. S., & Tremblay, M. (2009). Multiple targets of workplace affective commitment: Factor structure and measurement invariance of the Workplace Affective Commitment Multidimensional Questionnaire. *Advances in Psychology Research, 59*(1), 45–75.

Murray, A. L., Booth, T., & Molenaar, D. (2016). Personality differentiation by cognitive ability: An application of the moderated factor model. *Personality and Individual Differences, 100*, 73–78.

Mõttus, R. (2016). Towards more rigorous personality trait–outcome research. *European Journal of Personality, 30*(4), 292–303. https://doi.org/10.1002%2Fper.2041.

Mõttus, R., Allik, J., & Pullmann, H. (2007). Does personality vary across ability levels? A study using self and other ratings. *Journal of Research in Personality, 41*(1), 155–170. https://doi.org/10.1016/j.jrp.2006.03.007

Navarro-González, D., Ferrando, P. J., & Vigil-Colet, A. (2018). Is general intelligence responsible for differences in individual reliability in personality measures? *Personality and Individual Differences, 130*, 1–5. https://doi.org/10.1016/j.paid.2018.03.034

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

Quilty, L. C., Oakman, J. M., & Risko, E. (2006). Correlates of the Rosenberg self-esteem scale method effects. *Structural Equation Modeling: A Multidisciplinary Journal, 13*(1), 99–117. https://doi.org/10.1207/s15328007sem1301_5

Rammstedt, B., Goldberg, L. R., & Borg, I. (2010). The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality, 44*(1), 53–61. https://doi.org/10.1016/j.jrp.2009.10.005

Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology, 41*(1), 1–48. https://doi.org/10.1006/cogp.1999.0735

Richardson, H. A., Simmering, M. J., & Sturman, M. C. (2009). A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational Research Methods, 12*(4), 762–800. https://doi.org/10.1177%2F1094428109332834.

Schermer, J. A., Bratko, D., & Bojić, J. M. (2020). A test of the differentiation of personality by intelligence hypothesis using the Big Five personality factors. *Personality and Individual Differences, 156*, Article 109764. https://doi.org/10.1016/j.paid.2019.109764

Schermer, J. A., Krammer, G., Goffin, R, D., & Biderman, M. D. (2020). Using the 16PF to Test the Differentiation of Personality by Intelligence Hypothesis. *Journal of Intelligence*, *8*(12). https://doi.org/10.3390/jintelligence8010012.

Shevlin, M., Miles, J. N. V., Davies, M. N. O., & Walker, S. (2000). Coefficient alpha: A useful indicator of reliability? *Personality and Individual Differences, 28*(2), 229–237. https://doi.org/10.1016/S0191-8869(99)00093-8

Sorjonen, K., Hemmingsson, T., & Melin, B. (2020). Intelligence, consistency, and Emerson's dilemma. *Personality and Individual Differences, 160*, Article 109943. https://doi.org/10.1016/j.paid.2020.109943

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of personality and social psychology, 100*(2), 330.

Spearman, C. (1927). *The abilities of man*. London: MacMillan.

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740

Waiyavutti, C., Johnson, W., & Deary, I. J. (2012). Do personality scale items function differently in people with high and low IQ? *Psychological Assessment, 24*(3), 545–555. https://doi.org/10.1037/a0026266

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods, 18*(3), 320–334. https://doi.org/10.1037/a0032121

Williams, L. J., & O'Boyle, E. H. (2015). Ideal, nonideal, and no-marker variables: The confirmatory factor analysis (CFA) marker technique works when it matters. *Journal of Applied Psychology, 100*(5), 1579–1602. https://doi.org/10.1037/a0038855