Multi-group covariance and mean structure modeling of the relationship between the WAIS-III common factors and sex and educational attainment in Spain

Conor V. Dolan [a],*, Roberto Colom [b], Francisco J. Abad [b], Jelte M. Wicherts [a], David J. Hessen [a], Sophie van de Sluis [a]

[a] Department of Psychology, FMG, University of Amsterdam, Roetersstraat 15, 1018WB Amsterdam, The Netherlands
[b] Facultad de Psicologia, Universidad Autonoma de Madrid, 28049 Madrid, Spain

Abstract

We investigated sex effects and the effects of educational attainment (EA) on the covariance structure of the WAIS-III in a subsample of the Spanish standardization data. We fitted both first order common factor models and second order common factor models. The latter include general intelligence (g) as a second order common factor. The results indicate that sex differences in means are due mainly to the WAIS-III factors Working Memory (WM) and Perceptual Organization (PO), rather than to the second order g factor.

In treating EA as a predictor of the first common factors, we found that EA explained a significant, but clearly varying, amount of variance in the 4 first order common factors. Differences in explained variance over sex were very small. Treating EA as a predictor in the second order common factor model, we found that EA predicted g and first order factor Verbal Comprehension (VC). The relationships between PO, WM, and Perceptual Speed (PS) and EA are mediated by g. In treating EA as a dependent variable, we found that VC and PS were the only significant predictors of EA in the first order common factor model. In the second order common factor model, the same result was obtained. While VC and PS were significant predictors of EA, the second order factor g, and PO and WM were not.

The results are of theoretical interest in the light of the Ankney–Rushton paradox, and the role of g in educational attainment.
© 2005 Elsevier Inc. All rights reserved.

1. Introduction

The aim of the present paper is to examine the effects of sex and level of educational attainment on the factor structure of the Wechsler Adult Intelligence Scale (WAIS-III; TEA, 1998) in the 1998 Spanish standardization sample. Previous analyses of these data addressed the relationship between educational attainment, sex, and general intelligence (g; Jensen, 1998) in separate analyses (Colom, Abad, Garcia, & Juan-Espinosa, 2002; Colom, Garcia, Juan-Espinosa, & Abad, 2002). These analyses involved the method of correlated vectors, or the factor analysis of the WAIS subtest correlation matrix augmented with the sex–subtest point bi-serial correlations (Jensen, 1998; Nyborg, 2003). The present analyses are based on the method of covariance and mean structure analysis (CMSA; Bollen, 1989; Jöreskog & Sörbom, 1999). As discussed

* Corresponding author. Tel.: +31 20 5256775.
E-mail address: c.v.dolan@uva.nl (C.V. Dolan).

below, this methodological approach to the multi-group analysis of IQ test scores has a number of advantages.

The issue of sex effects on psychometric intelligence test scores is interesting in the light of the so-called Ankney–Rushton paradox (Lynn, 1994, 1999). This paradox concerns the relationship between sex, brain volume, and IQ test scores. It has been established that brain volume and g correlate about .4 (Rushton & Ankney, 1996; Vernon, Wickett, Bazana, & Stelmack, 2000), and that males have larger brain volumes (Ankney, 1992; Rushton, 1992). However, according to the received view, sex differences in g are absent (Brody, 1992; Jensen, 1998; Loehlin, 2000). The apparent paradox is this: if brain size is related to g, and males and females differ in brain volume, why do they not differ in g?[1] One solution was presented by Lynn (1994, 1999), who proposed that, contrary to the received view, there are sex differences in general intelligence (see also Nyborg, 2004). According to Lynn, the received view is incorrect due to the failure to take into account developmental differences in the maturation between boys and girls (see also Jensen, 1998). Differences between boys and girls may be difficult to discern in samples in the age range from, say, 10 to 16 years, because sex differences in general intelligence and differences in maturation rates are confounded. This is an important point, for which Colom and Lynn (2004) offered support. Another source of contention concerns the operationalization of g. For instance, sex differences are absent on Raven Progressive matrices (RPM), which measures g well according to Jensen (1998). Lynn views the sum of primary or group factors (e.g., verbal comprehension, reasoning, and spatial abilities) as a more appropriate definition of g. It is with respect to this sum score that Lynn had established a sex difference, which is consistent with the sex difference in brain size. Lynn (1999) cautioned that the absence of a sex effect in the Wechsler full-scale IQ test scores is due to the under-representation of abilities that favor males in the Wechsler tests. Ankney (1992, 1995; see also Jensen, 1998) suggested that sex differences in intelligence subtest scores are due to sex differences in primary or group factors, rather than g. These sex differences concern a well-established male advantage in quantitative (i.e., mathematical, spatial) reasoning (Voyer, Voyer, & Bryden, 1995).

Colom, Garcia, et al. (2002) investigated the role of g in sex differences in the WAIS-III standardization data using the method of correlated vectors, and found that g was not the source of sex difference in WAIS-III subtest scores. Colom, Juan-Espinosa, Abad, and Garcia (2000) reported a similar result based on other intelligence tests, including the Primary Mental Ability subscales. Aluja-Fabregat, Colom, Abad, and Juan-Espinosa (2000) also investigated the role of sex differences in g using a variety of measures pertaining to cognitive abilities and academic performance. Factor analyses of the correlation matrix augmented with the point-biserial sex–subtest correlation suggested that a sex difference in g is absent.

Educational attainment (EA) is uncontroversial as there is no doubt that IQ test scores and level of EA are correlated (as high as .4 to .7; Brody, 1992; Jensen, 1998; MacKintosh, 1998). As Jensen (1998) states: "If there is any unquestioned fact in applied psychometrics, it is that IQ tests have a high degree of predictive validity for many educational criteria, such as grades, school drop out, number of years of schooling, probability of entering college, and after entering, probability of receiving a bachelor's degree" (p. 277). Although the association between IQ and EA is beyond doubt, there is some disagreement concerning the causal nature of this association (Ceci & Williams, 1997; Herrnstein & Murray, 1994; Wahlsten, 1997; Winship & Korenman, 1997). In addition, as with sex differences, demonstrating an association between EA and IQ subtest scores or full-scale IQ is one thing, but establishing the role of g in the association is another. Ackerman and Lohman (2003, p. 289) have noted the paucity of results explicitly addressing the relationship between g, rather than IQ test scores, and educational outcomes.

Colom, Abad, et al. (2002) investigated the relationship between g and EA using the method of correlated vectors. They concluded that there is no significant association between g and mean differences between levels of EA on the WAIS-III subtest scores. This suggests that, although mean subtest differences between the levels of education are substantial, g is not an important source of these differences.

In the present paper we address the issues of sex effects and EA effects using multi-group covariance and mean structure analyses (MG-CMSA). The outline of the present paper is as follows. First we explain briefly why we use MG-CMSA, rather than other, simpler methods. We present details concerning the sample, the measures, and the results of preliminary analyses. We then present the MG-CMSA models that we apply to investigate sex differences, and explain

---

[1] We note that from a purely statistical point of view, the correlations themselves do not necessarily pose a paradox. If X is positively correlated with Y (say, .4), and Y is positively correlated with Z (say, .5), X and Z may be uncorrelated (Langford, Schwertman, & Owens, 2001).

how we evaluated goodness of fit. Subsequently, we present the results of MG-CMSA of sex differences. We then present the regression models used to investigate the role of EA in psychometric IQ. We treat EA both as a dependent variable and an independent variable in relation to the IQ test scores. This strategy is not meant to unveil the direction of causality, which is not possible in cross-sectional data. Rather it is meant to gain a detailed insight into the relationship between *g* and EA. We conclude this paper with a discussion.

2. Multi-group covariance and mean structure analysis (MG-CMSA)

MG-CMSA is a well-established method to study group differences in means and covariance structures within the confirmatory common factor model (e.g., Byrne, Shavelson, & Muthén, 1989; Jöreskog, 1971; Little, 1997; Marsh & Grayson, 1990; Millsap & Everson, 1991; Ployhart & Oswald, 2004; Rock, Werts, & Flaugher, 1978; Sörbom, 1974; Vandenberg & Lance, 2000; Widaman & Reise, 1997). Specifically in the study of group differences in IQ test scores, MG-CMSA was advocated by Gustafsson (1992), Horn (1997), Millsap (1997), and Dolan (2000). Gustafsson (1992), Dolan (2000), Dolan and Hamaker (2001), Lubke, Dolan, Kelderman, and Mellenbergh (2003), and Dolan, Roorda, and Wicherts (2004) applied MG-CMSA to study ethnic differences in psychometric intelligence. Wicherts et al. applied this method to study the Flynn effect (Wicherts et al., 2004), and stereotype threat (Wicherts, Dolan, & Hessen, in press). van der Sluis et al. (in press) used this method to study sex differences in the Dutch WAIS-III-R.

Compared to other methods that are used to investigate group differences in IQ, MG-CMSA has the following advantages. First, it allows one to formulate and test measurement invariance in a rigorous manner (Lubke et al., 2003; Mellenbergh, 1989; Meredith, 1993; Widaman & Reise, 1997). This is an essential prerequisite for the comparison of groups with respect to latent variables (e.g., *g*). Often measurement invariance is studied by calculating congruence measures, which express the collinearity of factor loadings. However, a formal statistical test of this measure is lacking. Also, it is unclear what value of this measure is viewed as acceptable, which compounds the problems of statistical testing. Finally, and most importantly, the congruence measure is simply *not* sufficient to establish measurement invariance (Meredith, 1993). A second advantage of MG-CMSA is that it offers the means to

compare competing hypotheses using rigorous statistical (goodness of fit) criteria. These include the hypothesis that *g* is the only or the main source of group differences, and the hypothesis that primary group factors are the source of group differences. Clearly, to demonstrate convincingly the predominance of *g* as the source of group differences, it is important to establish that the *g* model fits the data acceptably, and that it fits the data better than competing models.

The method of correlated vectors was devised by Jensen to investigate the role of *g* in the relationship between IQ subtest scores and other variables such as sex, race, scholastic achievement or educational attainment (Jensen, 1998). The relationship between the method of correlated vectors and MG-CMSA is discussed at length in Millsap (1997), Dolan (2000), Dolan and Hamaker (2001), and Lubke et al. (2003). Although this method has largely met with explicit or implicit approval and is viewed as viable (e.g., Gray & Thompson, 2004), very little is known about its validity or robustness. Lubke, Dolan, and Kelderman (2001), in the only known simulation study of this method, demonstrated that the method of correlated vectors is insensitive to model violations. Dolan et al. (2004) demonstrated this in analyses of real data. In addition, this method produces a single correlation, but little is known about the values that this correlation should assume for one to reject the role of *g* in group differences. Values ranging from about .3 to .8 have been interpreted in support of a major role of *g* in group differences. We note that Nyborg (2003) warns that the method of correlated vectors is not suitable to investigate small sex differences. This raises the problem of how large group differences should be for this method to work. In view of the many applications of the method of correlated vectors, these problems appear to be poorly understood (see also Ashton & Lee, in press).

Nyborg favors the Schmid–Leiman factor analysis of the correlation matrix of the subtests and sex (using the point bi-serial correlation coefficient to express the relationship between subtest scores and sex). This exploratory procedure does not involve goodness of fit testing or the comparison of competing models. In addition, because this statistical procedure has yet to be validated, nothing is known about how robust it is given model violations.

Finally, Lynn (1999) favors comparing sexes with respect to a measure of *g* derived by adding primary common factors scores. Assuming the second order *g* factor model is tenable, the simple summation of first order factor scores may yield a reasonable approxima-

Table 1
Amount of education associated with different levels of educational attainment in the Spanish standardization of the WAIS III

Level 1		Level 2						Level 3				Level 4
Primary and secondary school								High school				College
1°	2°	3°	4°	5°	6°	7°	8°	1°	2°	3°	4°	
6y	7y	8y	9y	10y	11y	12y	13y	14y	15y	16y	17y	18y+

The age is shown in the 4th row, grades are shown in the 3rd row.

tion to the scores on the *g* factor within a given population. However, in the investigation of between group differences, this procedure can produce misleading results. The sum of first order mean differences is a number that says nothing about the *structure* of these mean differences. MG-CMSA however, was designed specifically to investigate covariance and mean structures.

3. Sample and measures

The WAIS-III was standardized in Spain in 1998 (TEA, 1998). The 14 subtests of the WAIS-III are denoted vocabulary (VOC), similarities (SIM), arithmetic (ARIT), digit span (DS), information (INFO), comprehension (COM), letter–number series (L–N), picture completion (PC), digit–symbol coding (COD), block design (BD), matrix reasoning (MA), picture arrangement (PA), symbol search (SS), and object assembly (OA). The total sample comprises 1369 cases. In addition to the 14 WAIS-III subtest scores, the data set included a measure of educational attainment (4 levels), an indicator of age group (16–19Y, 20–24Y, 25–34Y, 35–54Y, 55–69Y, and 70+Y), and a geographical location indicator (north, east, west, and south Spain). The 4 levels of educational attainment are explained in Table 1.

The analyses of sex differences in IQ were limited to the first 3 age groups, i.e., ages 16 to 34Y. These groups were pooled to form a male and a female sample, comprising 307 and 281 cases, respectively. The decision to limit the analyses to this age range is based on the fact that an analysis of variance (ANOVA) revealed a significant effect of age group on educational attainment ($F(5, 1363) = 112.6$, $p < 0.001$). Post hoc analyses revealed that the first three age groups (16–34Y) did not differ with respect to mean educational level. The means (std) are 2.75 (.45), 2.84 (.62), 2.82 (.74), in the age group 16–19, 20–24, and 25–34, respectively. Age groups 4 to 6 differed both mutually and from the first three age groups. The means (std) equal 2.28 (.90), 1.74 (.84), 1.42 (.63) in the age groups 35–54, 55–69, and 70+, respectively. In addition, ANOVAs revealed a

main effect of age group on all 14 subtests. As with educational level, we found that this main effect was due to mean differences between the first three age groups (16–34Y) versus age groups 4, 5, and 6. Differences in means of the subtest in the first three groups were absent. Males and females in the age range 16–34Y are comparable with respect to educational attainment ($F(1, 586) = 1.17$, ns) and age ($F(1, 586) = .042$, ns), and also with respect to geographical region ($\chi^2(3) = .978$, ns). In conclusion, limiting the MG-CMSA of sex differences to the first three groups has the important advantage that these age groups are homogeneous with respect to the educational attainment and to the WAIS-III subtest scores. Were one to pool all age groups, the covariance and mean structure would obviously be distorted by age group differences.

In the analyses of the relationship between WAIS-III scores and educational attainment (EA), we face the problem that age in the range (16 to 34Y) and level of EA are partially confounded. Specifically, because enrollment in college (level 4) generally takes place after 18 in Spain, it is unknown whether a given 16-year-old will attain level 4. We therefore investigated the relationship between the WAIS-III scores and educational attainment in the 20 to 34-year-olds (223 males, 202 females).

4. Preliminary analyses

The sample, limited to age range 16 to 34Y, comprised 307 males and 281 females. Table 2 contains the means and standard deviations of the males and females

Table 2
Univariate means, standard deviations, effect sizes

Subtest	Males			Females			δ	*t*-test
	N	Mean	Std.	N	Mean	Std		
VOC	307	42.1	9.96	281	43.3	8.88	−0.127	.146
SIM	307	20.2	5.77	281	20.3	5.24	−0.018	.916
ARIT	307	14.6	3.39	281	12.9	3.41	0.500	<.001
DS	307	17.4	4.29	281	16.6	3.75	0.198	.016
INFO	307	18.8	4.87	281	17.1	4.99	0.345	<.001
COM	307	19.8	5.39	281	19.4	5.40	0.074	.363
L–N	307	11.6	2.85	281	11.2	2.46	0.150	.032
PC	306	20.2	3.06	281	20.5	2.55	−0.106	.220
COD	307	76.6	17.00	281	78.8	16.96	−0.129	.144
BD	306	47.7	10.77	281	44.7	10.53	0.281	<.001
MA	307	19.5	4.20	281	19.0	3.95	0.122	.063
PA	307	15.7	4.23	281	15.1	3.92	0.146	.070
SS	305	35.8	8.37	281	35.0	7.87	0.098	.179
OA	306	35.0	8.93	281	34.5	7.67	0.060	.388

The last column contains the *p* values of the independent *t*-test. Delta (δ) represents the males − females mean difference, standardized by the pooled standard deviation.

with respect to the 14 WAIS-III subtest scores. Means and covariance matrices are available upon request. Missing data were limited to 5 subtest scores in the male sample.[2] A multivariate ANOVA revealed a main effect of sex ($F(14, 568) = 7.944$, $p < .001$). Table 2 also contains the mean differences ($\mu_m - \mu_f$) divided by the pooled standard deviation (Cohen's effect size, denoted δ), and the p-values associated with the t-test for each subtest. Adopting a liberal type I error rate (α) of .05, we find that 5 of the 14 subtests display a significant mean difference. The males display a significant mean advantage on the subtests arithmetic, digit span, information, letter–number series, and block design.

Colom, Abad, et al. (2002), Colom, Garcia, et al. (2002) used the method of correlated vectors to investigate the hypothesis that g was the main source of sex differences in the complete sample (full age range). They concluded that this hypothesis was not supported. Using the same method in the present age range, we obtained similar results based on the Schmid–Leiman exploratory factor analysis (Schmid & Leiman, 1957). The congruence measure between the g loadings equaled .993. The correlation between the vector of standardized mean differences and g loadings equals .31 (Pearson product moment correlation) and .22 (Spearman's rho). Although it is presently unknown how large this statistic should be to conclude that g is important, we assume that this value is too low to conclude that g is the main source of group differences. Colom, Abad, et al. (2002), Colom, Garcia, et al. (2002) reported much lower values.

5. First and second order MG-CMSA

We present the first and second order factor models in general terms and discuss some identifying constraints. We then discuss the constraints associated with measurement invariance.

Let $\mathbf{y}_{ij}$ denote the observed p-dimensional (px1) random column vector of subject j in population i (i.e., males or females). In the present case, $\mathbf{y}_{ij}$ contains WAIS-III subtest scores. The following common factor model is assumed to hold for the observation $\mathbf{y}_{ij}$

$$\mathbf{y}_{ij} = \mathbf{v}_{yi} + \mathbf{\Lambda}_i \, \mathbf{\eta}_{ij} + \mathbf{\epsilon}_{ij}, \tag{1}$$

---

[2] MG-CMSA was based on the mean vectors and covariance matrices. In the male sample these were calculated using raw data likelihood estimation in LISREL 8, in view of the 5 missing subtest scores (see Shafer & Graham, 2002). Raw data likelihood estimation is optimal, but given the very small number of missing data, other methods, e.g. pairwise deletion, are likely to be as good.

where $\mathbf{\eta}_{ij}$ is a q-dimensional random vector of common factor scores ($q < p$), and $\mathbf{\epsilon}_{ij}$ is a p-dimensional vector of residuals (Bollen, 1989). The residuals in $\mathbf{\epsilon}_{ij}$ contain both random error and a systematic component, uncorrelated with $\mathbf{\eta}_{ij}$. The (pxq) matrix $\mathbf{\Lambda}_i$ contains factor loadings, which may be viewed as regression coefficients in the regression of subtest scores $\mathbf{y}_{ij}$ on the common factors $\mathbf{\eta}_{ij}$. The (px1) vector $\mathbf{v}_{yi}$ contains intercepts.

In Jensen's applications of the method of correlated vectors, the common factor g features as a second order common factor (Jensen, 1998), which is assumed to account for the covariance among the first order factors. In MG-CMSA, we introduce g as a second order factor upon which the first order factors are regressed (e.g., Gustafsson, 1984, 1988; Rindskopf and Rose, 1988): $\mathbf{\eta}_{ij} = (\mathbf{\Gamma}_i g_{ij} + \mathbf{\omega}_{ij})$, where $\mathbf{\Gamma}_i$ is a (qx1) matrix of loadings of the q first order factors $\mathbf{\eta}_{ij}$ on the single second order factor, g_{ij}, and $\mathbf{\omega}_{ij}$ is a (qx1) vector of random (first order) uncorrelated residual terms. g_{ij} and $\mathbf{\omega}_{ij}$ are assumed to be uncorrelated. The model of the observations is now:

$$\mathbf{y}_{ij} = \mathbf{v}_{yi} + \mathbf{\Lambda}_i \big(\mathbf{\Gamma}_i g_{ij} + \mathbf{\omega}_{ij}\big) + \mathbf{\epsilon}_{ij}. \tag{2}$$

Assuming that $\mathbf{\eta}_{ij}$ and $\mathbf{\epsilon}_{ij}$ in Eq. (1) are uncorrelated, the expected (model) mean vector ($\mathbf{\mu}_{yi}$) and covariance matrix ($\mathbf{\Sigma}_{yi}$) in population according to this model are:

$$\mathbf{\mu}_{yi} = \mathbf{v}_{yi} + \mathbf{\Lambda}_i \, \mathbf{\mu}_{\eta i} \tag{3}$$

$$\mathbf{\Sigma}_{yi} = \mathbf{\Lambda}_i \mathbf{\Sigma}_{\mathbf{g}i} \, \mathbf{\Lambda}_i^t + \mathbf{\Sigma}_{\epsilon i}, \tag{4}$$

where superscript t denotes transposition, $\mathbf{\Sigma}_{\eta i}$ represents the (qxq) covariance matrix of the first order common factors ($\mathbf{\eta}_{ij}$), $\mathbf{\Sigma}_{\epsilon i}$ represents the (pxp) covariance matrix of the residuals ($\mathbf{\epsilon}_{ij}$), and $\mathbf{\mu}_{\eta i}$ denotes the (qx1) vector of factor means, $E[\mathbf{\eta}_{ij}] = \mathbf{\mu}_{\eta i}$. In the second order factor model (Eq. (2)), $\mathbf{\mu}_{\eta i}$ and $\mathbf{\Sigma}_{\eta i}$ are modeled as follows:

$$\mathbf{\mu}_{\eta i} = \mathbf{\Gamma}_i \mathbf{\mu}_{\mathbf{g}i} + \mathbf{\mu}_{\omega i} \tag{5}$$

$$\mathbf{\Sigma}_{\eta i} = \mathbf{\Gamma}_i \mathbf{\sigma}_{\mathbf{g}i}^2 \, \mathbf{\Gamma}_i^t + \mathbf{\Sigma}_{\omega i}, \tag{6}$$

where $\mathbf{\sigma}_{\mathbf{g}i}^2$ is the variance of g, $\mathbf{\Sigma}_{\omega i}$ is the diagonal (qxq) covariance matrix of the first order residuals, $\mathbf{\mu}_{\mathbf{g}i}$ is the mean of g, and $\mathbf{\mu}_{\omega i}$ is the (qx1) vector of the means of the first order factor residuals (i.e., $\mathbf{\omega}_{ij}$). The second order factor model is thus:

$$\mathbf{\mu}_{yi} = \mathbf{\nu}_{yi} + \mathbf{\Lambda}_i \Big[\mathbf{\Gamma}_i \mathbf{\mu}_{gi} + \mathbf{\mu}_{\omega i}\Big] \tag{7}$$

$$\mathbf{\Sigma}_{yi} = \mathbf{\Lambda}_i \Big[\mathbf{\Gamma}_i \mathbf{\sigma}_{\mathbf{g}i}^2 \mathbf{\Gamma}_i^t + \mathbf{\Sigma}_{\omega i}\Big] \mathbf{\Lambda}_i^t + \mathbf{\Sigma}_{\epsilon i}. \tag{8}$$

So far we have presented general multi-group covariance and mean structure models. In fitting these models to the WAIS-III data we introduce a number of identifying constraints (Bollen, 1989). First, we specify a configuration of factor loadings, which corresponds to the expected configuration of the WAIS-III (WAIS-III, WMS-III Technical Manual, Psychological Corporation, 1997). This configuration includes sufficient fixed zero loadings to avoid rotational indeterminacy (see Table 3). Second, to be able to estimate the (co)variances of latent variables, we fix certain elements in $\mathbf{\Lambda}_i$ and $\mathbf{\Gamma}_i$ to equal 1. This is a standard scaling constraint in confirmatory factor analysis (Bollen, 1989). Third, where factor means are included in the models, we estimate factor mean differences, relative to an arbitrary reference group (Bollen, 1989; Sörbom, 1974). In the present case, the males will act as reference group, i.e., the means of the females will be estimated relative to the to zero constrained means of the males. In addition to these identifying constraints, we introduce substantive constraints to investigate measurement invariance over the groups (Meredith, 1993). We outline these below.

6. Sex differences: factorial invariance models

We fit the following sequence of first order factor models in the male (subscript m) and female (subscript f) samples (see Horn & McArdle, 1992; Meredith, 1993; Widaman & Reise, 1997). First we fit the following model, noted *model A1*:

$$\mathbf{\Sigma_{yi}} = \mathbf{\Lambda}_i \mathbf{\Sigma_{\eta i}} \mathbf{\Lambda}_i^t + \mathbf{\Sigma_{\varepsilon i}} \text{ and } \mathbf{\mu_{yi}} = \mathbf{\nu}_{yi}, \qquad (9)$$

where the group subscript i again denotes males or females. The matrices $\mathbf{\Lambda}_m$ and $\mathbf{\Lambda}_f$ have an identical

Table 3
Expected WAIS-III factor loading matrix

Subtest	VC	PO	WM	PS
VOC	λ	0	0	0
SIM	λ	0	0	0
ARIT	0	λ	λ	0
DS	0	0	λ	0
INFO	λ	0	0	0
COM	λ	0	0	0
L–N	0	0	λ	0
PC	0	λ	0	0
COD	0	0	0	λ
BD	0	λ	0	0
MA	0	λ	0	0
PA	λ*	λ	0	0
SS	0	0	0	λ
OA	0	λ	0	0

* Note: Loading of PA on VC was added on the basis of modification indices.

configuration of factor loadings, but equality constraints over sex are absent. The configuration of estimated and zero factor loadings is based on the expected WAIS-III factor structure. Note that the means are included, even though they are unconstrained. This is important to maintain comparability of certain fit indices with indices of models, in which the means are constrained (Wicherts & Dolan, 2004). Model A1 serves to establish that the expected WAIS-III common factor model is acceptable in females and males. Subsequently, we introduce the constraint that the factor loadings are equal, $\mathbf{\Lambda}_m = \mathbf{\Lambda}_f = \mathbf{\Lambda}$. We denote this *model A2*:

$$\mathbf{\Sigma_{yi}} = \mathbf{\Lambda} \mathbf{\Sigma_{\eta i}} \mathbf{\Lambda}^t + \mathbf{\Sigma_{\varepsilon i}} \text{ and } \mathbf{\mu_{yi}} = \mathbf{\nu}_{yi}. \qquad (10)$$

Model A2 implies that the regressions of the observed variables on the common factors are identical over the groups. This is a necessary condition for the comparison of the groups with respect to the common factors. Next we constrain the residual covariance matrices $\mathbf{\Sigma_{\varepsilon m}} = \mathbf{\Sigma_{\varepsilon f}} = \mathbf{\Sigma_{\varepsilon}}$ (*model A3*):

$$\mathbf{\Sigma_{yi}} = \mathbf{\Lambda} \mathbf{\Sigma_{\eta i}} \mathbf{\Lambda}^t + \mathbf{\Sigma_{\varepsilon}} \text{ and } \mathbf{\mu_{yi}} = \mathbf{\nu}_{yi}. \qquad (11)$$

Finally, we add the means structure (*model A4*):

$$\mathbf{\Sigma_{ym}} = \mathbf{\Lambda} \mathbf{\Sigma_{\eta i}} \mathbf{\Lambda}^t + \mathbf{\Sigma_{\varepsilon}} \text{ and } \mathbf{\mu_{yi}} = \mathbf{\nu}_y + \mathbf{\Lambda} \mathbf{\mu_{\eta i}}. \qquad (12)$$

As it stands, model A4 is not identified, because it is not possible to estimate the factor means $\mathbf{\mu_{\eta m}}$ and $\mathbf{\mu_{\eta f}}$ simultaneously. As mentioned above, we estimate factor mean differences by fixing $\mathbf{\mu_{\eta m}}$ to zero and estimating $\mathbf{\mu_{\eta f}}$ relative to $\mathbf{\mu_{\eta m}}$:

$$\mathbf{\mu_{ym}} = \mathbf{\nu}_y \qquad (13)$$

$$\mathbf{\mu_{yf}} = \mathbf{\nu}_y + \mathbf{\Lambda}\left(\mathbf{\mu_{\eta f}} - \mathbf{\mu_{\eta m}}\right) = \mathbf{\nu}_y + \mathbf{\Lambda} \mathbf{\mu_{\eta\Delta}}, \qquad (14)$$

where $\mathbf{\mu_{\eta\Delta}} = (\mathbf{\mu_{\eta f}} - \mathbf{\mu_{\eta m}})$. Model A4 represents the hypothesis of strict factorial invariance (Meredith, 1993). In this model, the observed group differences are modeled explicitly as a function of latent group differences, i.e., differences with respect to the variables that we purport to measure (see Lubke et al., 2003, for a detailed discussion). Depending on the details of the results in fitting models A1 to A4, we may, where necessary, introduce modifications (see below).

Next, retaining the constraints of model A3 to ease presentation, we introduce the 2nd order factor model (*model B1*):

$$\mathbf{\Sigma_{yi}} = \mathbf{\Lambda} \left(\mathbf{\Gamma}\sigma_{gi}^2\mathbf{\Gamma}^t + \mathbf{\Sigma_{\omega i}}\right) \mathbf{\Lambda}^t + \mathbf{\Sigma_{\varepsilon}} \text{ and } \mathbf{\mu_{yi}} = \mathbf{\nu}_{yi}.$$

$$(15)$$

Note that we fit the model with unstructured means so that we can evaluate the fit of the *g* model on the covariance structure first. We subsequently introduce the model for the means (*model B2*):

$$\mu_{ym} = v_y \tag{16}$$

$$\mu_{yf} = v_y + \Lambda\mu_{\omega\Delta} + \Lambda\Gamma\mu_{g\Delta}, \tag{17}$$

where $\mu_{g\Delta} = (\mu_{gm} - \mu_{gf})$ and $\mu_{\omega\Delta} = (\mu_{\omega m} - \mu_{\omega f})$, i.e., as in model A4, we model latent mean differences. It is not possible to estimate $\mu_{g\Delta}$ and all elements in $\mu_{\omega\Delta}$ simultaneously. For reasons of identification, one element in $\mu_{\omega\Delta}$ or the parameter $\mu_{g\Delta}$ has to be fixed to zero.

7. Assessment of model fit

Below we first fit models A1, A2, A3, A4, and B1. To assess overall goodness of fit, we considered a variety of fit indices as recommended by Bollen and Long (1993). See Schermelleh-Engel, Moosbrugger, and Müller (2003) for a comprehensive account of goodness of fit indices. We report the consistent Akaike's Information Criterion (CAIC), the expected cross-validation index (ECVI), the root mean square error of approximation (RMSEA), and the χ^2 with the associated degrees of freedom. The lower the value of the CAIC, the better the fit of the model. The ECVI provides an indication of the discrepancy between the fitted covariance matrices in the analyzed samples and the expected covariance matrices that one would obtain in a second sample of the same size. ECVI produces the same rank order of models as the Akaike Information Criterion (AIC). Compared to the ECVI, and thus AIC, CAIC favors more parsimonious models. As with the CAIC, models with low values of the ECVI are preferable to models with large values. The RMSEA is a measure of the error of approximation of the specified model covariance and mean structures to the covariance and mean structures in the population(s). As a rule of thumb, Browne and Cudeck (1993) suggest that a RMSEA of about 0.05 or less is indicative of a good approximation. The χ^2 is treated as a measure of (badness of) fit rather than as a formal test statistic (Jöreskog, 1993). We assess local misfit by inspecting the modification indices (MI). These are calculated for each parameter that is subjected to a constraint (e.g., fixed to a given value, or subjected to an equality constraint), and represent the decrease in χ^2 that is expected were one to relax the constraint. A MI of 3.84 (6.63) is associated with a probability of .05 (.01). As the use of MIs may readily result in chance capitalization, we use the MIs with moderation. All models were fitted

Table 4
Fit indices of MG-CMSA

Models	df	χ^2	RMSEA	ECVI	CAIC
1st order					
A1	140	271.9	.056	.789	994.3
A1r	136	212.2	.042	.700	958.0
A2	148	239.8	.044	.703	895.9
A3	163	253.1	.041	.672	797.2
A4	173	356.6	.059	.823	831.8
A4r	169	270.6	.043	.678	768.2
2nd order					
B1	170	266.9	.042	.671	758.9
B2	176	284.4	.043	.677	729.7
B2r	178	284.5	.043	.670	715.1

using normal theory maximum likelihood estimation in LISREL 8 (Jöreskog & Sörbom, 1999).[3]

8. Sex differences: results

In fitting multi-group confirmatory factor models, we adhere to the theoretical WAIS-III factor structure as closely as possible (TEA, 1998; Saklofske, Hildebrand, & Gorsuch, 2000; Ward, Ryan, & Axelrod, 2000). The WAIS-III is designed to measure the following first order, or primary factors: verbal comprehension (VC), perceptual organization (PO), working memory (WM), and perceptual speed (PS). Table 3 contains the theoretical configuration of factor loadings.

We first fitted model A1 (Eq. (9)), i.e., a confirmatory factor model with four correlated factors (VC, PO, WM, PS), factor loading configuration (Λ_i) as shown in Table 3, and diagonal residual variance matrices ($\Sigma_{\varepsilon i}$). The fit indices for this model are shown in Table 4.

The $\chi^2(140)$ equals 271.9, the RMSEA indicates that the approximate model fit is acceptable (.056). The largest modification indices concerned the covariance between the residuals of BD and OA (22.1 and 16.1, in male and female samples, respectively), and the PA subtest loading on the factor VC (6.9 and 19.0 in the female and male samples, respectively). Although the value 6.9 in the male sample is not particularly large, it is the largest MI in the matrix Λ_m. We introduced these four parameters (covariance between the residuals of BD and OA, and factor loadings of PA on VC) to form model A1r (A1 revised).[4] The fit indices for this

---

[3] All LISREL input files and summary statistics are available upon request.

[4] The factor loading of PA on VC is also observed in exploratory factor analyses of the WAIS III reported in the WAIS III 1997 Technical Manual. The covariance between the residuals of BD and OA may be attributable to the fact that both involve a motor hand-coordination component.

model are shown in Table 4. The $\chi^2(136)$ equals 212.2, the RMSEA indicates that the approximate model fit is acceptable (.042), and both CAIC (958.0) and ECVI (.700) indicate that model A1r fits better than A1. The difference in $\chi^2(4)$ is considerable: 59.7 ($p < .001$). Fig. 1 contains the path diagram of the fitted factor model.

Retaining the modifications of model A1r, we introduced equal factor loadings $\Lambda_m = \Lambda_f$ (model A2; Eq. (10)). These equality constraints are acceptable. The $\chi^2(148)$ equals 239.8, the RMSEA equals .044, and the CAIC of 895.9 indicates that the more parsimonious model A2 fits better than A1r. The ECVI shows negligible deterioration (.703 vs. .700). Next we fitted model A3 (Eq. (11)), in which the residual covariance matrices ($\Sigma_{\epsilon m} = \Sigma_{\epsilon f}$) and factor loadings ($\Lambda_m = \Lambda_f$) are equal over sex. The $\chi^2(163)$ equals 253.1, the RMSEA equals .041, and both CAIC (797.2) and ECVI (.672) indicate that model A3 fits relatively well.

We proceeded by fitting model A4, which includes the model for the means (Eq. (12)). The $\chi^2(173)$ equals 356.6, the RMSEA equals .059, and both CAIC (831.8) and ECVI (.823) indicate that model A4 fits poorly compared to model A3. These results suggest that the sex related subtest mean differences could not be modeled adequately by means of sex related common factor mean differences. The modification indices identify the subtests VOC (MI=21.8), ARIT (MI=24.8), INFO (MI=42.6), and PC (MI=14.9) as the variables causing this misfit. In the female sample the difference between the observed and the fitted means of these variables are .10 (VC), −.15 (ARIT), −.16 (INFO), and −.11 (PC) in female std. units. Thus, female mean on VC is overestimated in this model, while the means of

ARIT, INFO, and PC are underestimated. We removed the intercept equalities on these variables and refit the model, which we denote model A4r. Relaxing the intercept equalities in effect removes these four variables from the model of the means. The $\chi^2(169)$ of model A4r equals 270.6, the RMSEA equals .043, and both CAIC (768.2) and ECVI (.678) indicate that model A4r fits relatively well.

As we judge model A4r to be acceptable, we consider the results of this model in detail. The latent mean differences are −0.027 (VC; s.e. .091), −0.232 (PO; s.e. .090), −0.211 (WM; s.e. .086), and 0.002 (PS, s.e., .102). Expressed in standard deviation units in the female sample, these are −0.029 (VC), −0.280 (PO), −0.255 (WM), and 0.002 (PS). Judging by the standard errors, the mean differences are limited to the PO and the WM factors. Indeed, refitting the model with the mean differences of VC and PS fixed to zero, did not result in any appreciable deterioration in fit ($\chi^2(171)=270.7$, RMSEA=.042, ECVI=.671, CAIC=753.5). As shown in Table 5, the standard deviations of the common factors are slightly smaller in the female sample, as are the intercorrelations. Bearing in mind that the subtest VOC, ARIT, INFO, and PC have been effectively removed from the structural model for the means (and thus play no role in the common factor mean differences), we thus find that the sex differences with respect to the factor means (favoring males) are limited to the common factors PO and WM. Table 6 contains the observed and fitted mean, standardized residuals of the means and the MI of the intercepts.

The fact that the equality constraints on the intercepts of the subtests VOC, ARIT, INFO, and PC were unten-
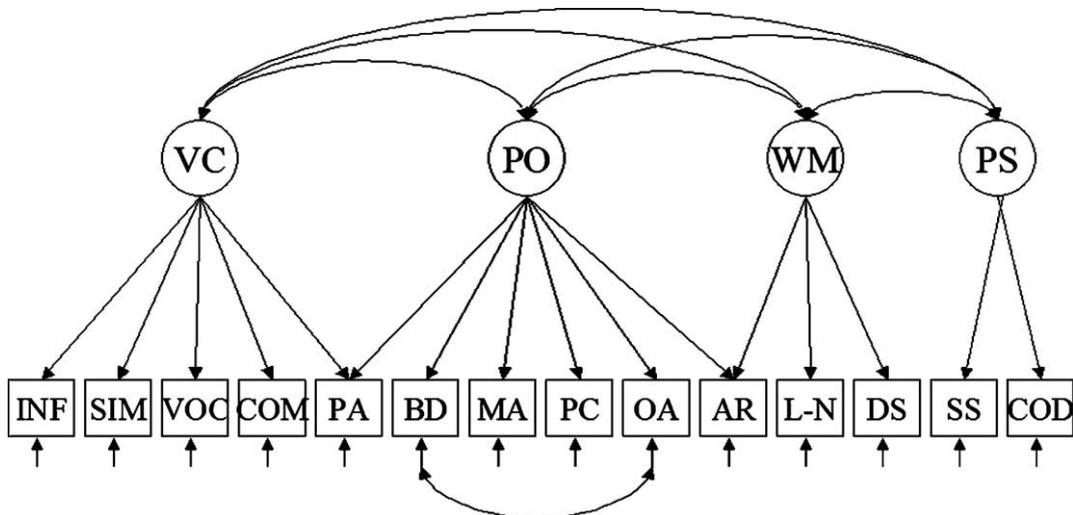

Fig. 1. WAIS-III confirmatory common factor structure.

Table 5
Covariance and mean differences of the common factors in model A4r, first order oblique common factor model with structured means

	VC	PO	WM	PS
Males				
VC	1.000*			
PO	0.792	1.000*		
WM	0.616	0.537	1.000*	
PS	0.624	0.711	0.587	1.000*
sd	1.00*	1.00*	1.00*	1.00*
Mean	0.00*	0.00*	0.00*	0.00*
Females				
VC	1.00*			
PO	0.714	1.00*		
WM	0.500	0.510	1.00*	
PS	0.417	0.556	0.280	1.00*
sd	0.919	0.826	0.826	0.947
Mean diff.	0.00*	− .232	− .211	0.00*
s.e	−.−	0.090	0.086	−.−

Parameters with an asterisk are fixed. Standard errors are reported for the differences in common factor means, which are estimated in the female sample.
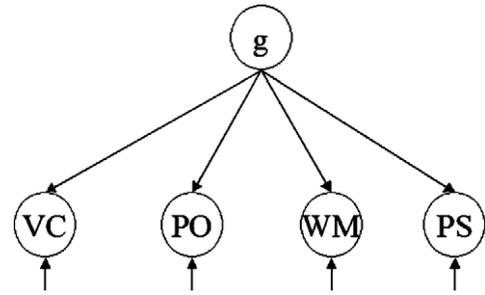

Fig. 2. WAIS-III second order common factor structure.

order factor model fits the covariance structure adequately. Model B1 resembles model A3 in all details, except for the introduction of the second order factor g, as shown in Fig. 2.

Comparing the fit indices of model B1 to those of A3, we find that the model fits relatively well. The $\chi^2(170)$ equals 266.9, the RMSEA equals .042, and both CAIC (758.9) and ECVI (.671) indicate that the fit of model B1 is adequate. The female standard deviation of g is about .8 times the male standard deviation of g (6.74 vs. 5.43). The g factor explains 78% (VC), 80% (PO), 43% (WM), and 53% (PS) of the variance of the first order factors in the male sample, and 62% (VC), 80% (PO), 38% (WM), and 36% (PS) in the female sample.

We next fitted the model with structured means (model B2, Eqs. (16) and (17)). As in model A4r, we do not constrain the means of the variables VOC, ARIT, INFO, and PC, i.e., the intercepts of these variables are again free to vary over sex. From fitting model A4r, we know that sex differences are limited to the first order common factors PO and WM. In model B2, we limited the latent mean difference to the second order common factor (g), and to the primary factors

able does not detract too much from the present results. Specifically, removing these variables from the structural model of the means, leaves three indicators (SIM, COM, PA) of VC, four indicators of PO, and two indicators of WM. These remaining indicators are sufficient to identify and test the common factor mean differences.

We proceeded by examining the second order factor model. In fitting model B1 (Eq. (15)), we removed the model for the means (these are now unconstrained as in model A3), and introduced the second order common factor (g). We removed the restrictions on the means, because we first want to establish whether the second

Table 6
Results of model A4r: fitted and observed means, standardized residuals (st.res) of the means, and modification indices (MI) associated with the equal intercepts

Subtest	Observed males	Fitted males	st.res males	Observed females	Fitted males	st.res females	MI
VOC	42.1	42.0	0.26	43.3	43.4	− 0.29	–
SIM	20.2	20.2	− 0.20	20.3	20.2	0.25	.84
ARIT	14.6	14.5	0.21	12.9	12.9	− 0.24	–
DS	17.4	17.3	0.48	16.6	16.6	− 0.55	.23
INFO	18.8	18.8	0.26	17.1	17.1	− 0.29	–
COM	19.8	19.6	0.73	19.4	19.6	− 0.79	1.15
L–N	11.6	11.6	− 0.06	11.2	11.1	0.16	.23
PC	20.2	20.2	0.21	20.5	20.5	− 0.23	–
COD	76.6	77.6	− 1.55	78.8	77.6	1.67	6.03
BD	47.7	47.0	1.86	44.7	45.4	− 2.10	6.90
MA	19.5	19.6	− 0.55	19.0	18.9	0.75	1.52
PA	15.7	15.6	0.61	15.1	15.2	− 0.65	0.40
SS	35.8	35.4	1.21	35.0	35.4	− 1.22	4.26
OA	35.0	35.3	− 1.07	34.5	34.1	1.23	4.03

The intercepts of the subtests VOC, ARIT, INFO, and COD are free to vary over sex. These MIs are thus zero.

Table 7
Covariance and mean differences of the common factors in model B2

Common factor covariance matrix (correlation in parentheses)

	VC	PO	WM	PS	*g*
Males					
VC	4.627				
PO	3.914	5.224			
	(.796)				
WM	1.645	1.775	1.746		
	(.579)	(.588)			
PS	2.131	2.300	0.966	2.252	
	(.660)	(.671)	(.487)		
g	1.904	2.055	0.864	1.119	1.000*
	(0.885)	(0.899)	(0.654)	(0.746)	
% explained by g					
	0.784	0.809	0.427	0.556	
Means					
	0.00*	0.00*	0.00*	0.00*	0.00*
Females					
VC	3.809				
PO	2.536	3.398			
	(.705)				
WM	1.065	1.150	1.271		
	(.484)	(.553)			
PS	1.381	1.490	0.626	2.214	
	(0.475)	(0.543)	(0.373)		
g	1.234	1.331	0.559	0.725	0.648
	(0.785)	(0.897)	(0.616)	(0.605)	
% var explained					
	0.617	0.805	0.380	0.366	
Mean differences (standard errors in parentheses)					
	0.0*	−0.501	−0.270	0.0*	0.0*
		(0.184)	(0.104)		

Asterisk indicates a fixed parameter. Factor mean differences are limited to the first order common factors PO and WM.

VC, PO, and WM. We fixed the latent mean difference associated with PS to zero for reasons of identification. We chose PS as here mean sex differences were absent (see above). The $\chi^2(176)$ of model B2 equals 284.4, the RMSEA equals .043, ECVI equals .677, and CAIC equals 729.7. As such this model fits as well as model A3r. However, the estimate of the mean difference in *g* equals .032 (s.e. .953), which is not significant judging by its standard error. The estimates of the first order factor residual mean differences are − .252 (VC; s.e. .989), − 1.785 (PO; s.e. .921); − .713 (WM, s.e. .447). Judging by the standard errors the latent mean difference associated with VC is not significant. Fixing the latent mean differences associated with *g* and VC to equal zero, and refitting the model, we have $\chi^2(178)=284.5$, the RMSEA equals .043, both ECVI equals .670, and CAIC equals 715.1. We refer to this

model as model B2r in Table 4. This model closely resembles model A4r. The only difference is that A4r is an oblique first order common factor model, while B2r is a second order common factor model. The model for the means is identical to A4r as it identifies the common factors PO and WM as the source of the sex differences. Table 7 contains the common factor covariance matrices and mean vectors. The observed and fitted means, standardized residuals of the means, and the modification indices associated with the intercepts closely resemble those of model A4r (see Table 6).

9. Educational attainment

The second aim of the present paper is to investigate the relationship between educational attainment (EA) and the common factors of the WAIS-III. The objective is to investigate the role of *g* in the relationship between the common factors VC, PO, WM, and PS and EA. As mentioned above, Colom, Abad, et al. (2002), Colom, Garcia, et al. (2002) found that the relationship between EA and the WAIS-III subtest scores was not due to *g*. As the exact causal model linking EA and psychometric intelligence is unknown, we treat EA both as a dependent and as an independent variable in relation to the common factors. In adding EA to the factor model, we built on previous results (models A3 and B1). Having established the nature of sex differences, we relinquish the means model, as there are no sex related mean differences in EA. We do retain the two group approach, because there may be sex differences in the strength of association between educational attainment and the common factors, including *g*. Note that here the issue of measurement invariance remains important. For instance, any gender difference in the variance of EA explained by the WAIS common factor can only be interpreted unambiguously if measurement invariance of the WAIS with respect to gender has been established.

Table 8 contains the distribution of EA over the age groups and sex. Because subjects younger than 18

Table 8
Educational level by age and sex

	Age				
	Level	16–19	20–24	25–34	20–34
Male	1	0	4	8	12
	2	20	16	37	53
	3	64	54	76	129
	4	0	6	22	28
Female	1	1	1	3	4
	2	19	13	35	48
	3	59	51	69	120
	4	0	8	22	30

cannot achieve level 4 (college), we limited the subsequent analyses to the age groups 20 to 34. These data comprise 223 males and 202 females. Summary statistics for this subset of the data are available on request. There is no difference in the distribution of EA between the 20–24 and the 25–34 year olds (males: $\chi^2(3)=5.13$, ns; females: $\chi^2(3)=5.19$, ns).

10. Educational attainment as a predictor

We specify various regression model at the level of the common factors, while retaining the measurement constraints of model A3. Treating educational attainment as a predictor of the common factors in model A3, we specify the following regression

$$\eta_{ij} = \beta_1 \, EA_{ij} + \zeta_{\eta ij}, \tag{18}$$

where the subscript i again denotes sex, and the subject j denotes case. A schematic representation is shown in Fig. 3 (bottom left).

The (4×1) vector β_1 contains the regression coefficients. Assuming EA and $\zeta_{\eta ij}$ are uncorrelated, the composition of common factor variance is as follows:

$$\Sigma_{\eta i} = \beta_1 \, \sigma^2_{EA i} \, \beta^t_1 + \Sigma_{\eta \zeta i}, \tag{19}$$

where $\sigma^2_{EA i}$ is the variances of EA. The (4×4) matrix $\Sigma_{\eta \zeta i}$ represents the covariance matrices of the common

factor residuals. The only differences with model A3 is that the covariance matrix of the common factors (e.g., in the males $\Sigma_{\eta m}$; see Eq. (11)) is now decomposed into a component due to EA ($\beta_1 \sigma^2_{EAm} \beta^t_1$) and a residual covariance matrix ($\Sigma_{\zeta\eta m}$). As it is unreasonable to suppose that EA can completely account for the common factor covariances, the residual covariance matrices include freely estimated covariances.

We also introduced EA as predictor in model B1, i.e., the hierarchical common factor model with VO, PO, WM, and PS as first order common factors, and g as the second order common factor. We specify the following regressions on g on EA and the first order common factor residuals on EA:

$$g_{ij} = \beta_2 \, EA_{ij} + \zeta_{gij}, \tag{20}$$

$$\omega_{ij} = \beta_3 \, EA_{ij} + \zeta_{\omega ij}, \tag{21}$$

where β_2 is a single regression coefficient, the (4×1) vector β_3 contains 4 regression coefficients, and ζ_{gij} and $\zeta_{\omega ij}$ are residuals. See Fig. 3 (bottom right) for a schematic representation. The decomposition of variance is as follows:

$$\sigma^2_{gi} = \beta^2_2 \sigma^2_{EA i} + \sigma^2_{\zeta gi} \tag{22}$$

$$\Sigma_{\omega i} = \beta_3 \, \sigma^2_{EA i} \, \beta^t_3 + \Sigma_{\omega \zeta i}. \tag{23}$$
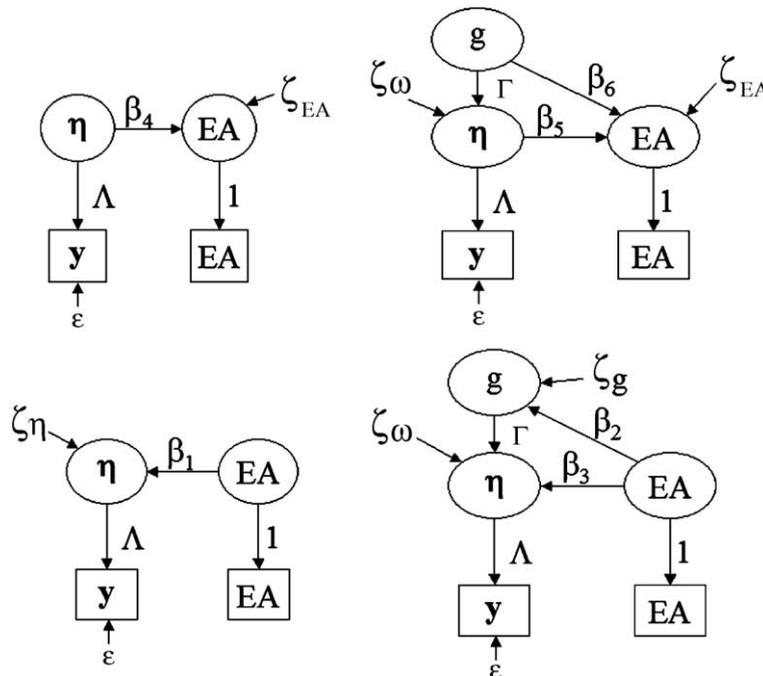

Fig. 3. EA as predicted (top) and predictor (bottom) in the first (left) and second order (right) common factor model.

Table 9
Covariance matrices among common factors and educational attainment (EA), and percentage of variance explained by EA as predictor of VC, PO, WM, and PS

	VO	PO	WM	PS	EA
Males (aged 20–34)					
VO	1.000				
PO	0.768	1.000			
WM	0.618	0.502	1.000		
PS	0.589	0.723	0.624	1.000	
EA	0.526	0.415	0.316	0.418	1.000
% expl	27.7	17.2	10.0	17.5	
Females (aged 20–34)					
VO	1.000				
PO	0.768	1.000			
WM	0.570	0.628	1.000		
PS	0.451	0.642	0.384	1.000	
EA	0.539	0.453	0.364	0.429	1.000
% expl	29.1	20.6	13.3	18.4	

Results are based on the first order common factor model.

In terms of the implied decomposition of the first order factors we have:

$$\boldsymbol{\Sigma}_{\eta i} = \boldsymbol{\Gamma} \left\{ \beta_2^2 \sigma_{EAi}^2 + \sigma_{\zeta gi}^2 \right\} \boldsymbol{\Gamma}^t$$
$$+ \left\{ \beta_3 \sigma_{EAi}^2 \beta_3^t + \boldsymbol{\Sigma}_{\omega \zeta i} \right\}. \quad (24)$$

Limiting ourselves to the males, we thus decompose the total variance of g into a part that is due to EA ($\beta_2^2 \sigma_{EAm}^2$) and a residual ($\sigma_{\zeta gm}^2$). In addition, the first order common factor variance is decomposed into a part due to EA ($\beta_3 \sigma_{EAm}^2 \beta_3^t$) and a residual (diagonal covariance matrix $\boldsymbol{\Sigma}_{\omega \zeta m}$). It is not possible to estimate all four regression coefficients in β_3 and the regression coefficient β_2. One regression coefficient has to be fixed for reasons of identification.

11. Educational attainment as a predictor: results

We added EA to model A3, and regressed the common factors VC, PO, WM, and PS on this variable (Eq. (18)). The fit indices indicate that the model fitted well ($\chi^2(1817)=273.6$, RMSEA=.043, EVCI=1.013, CAIC=847.4). The correlations between educational attainment and the common factors are shown in Table 9. The regression coefficients are all highly significant ($p<.001$).

As shown in Table 9, educational attainment explained about 28% (VC), 17% (PO), 10% (WM) and 17% (PS) of the variance of the common factors in the males sample, and about 29% (VC), 21% (PO), 13% (WM) and 18% (PS) in the female sample. The sex differences in explained variances are thus quite

minor. The associated path diagram is shown in Fig. 4 (top).

Adding EA as a predictor to model B2, we face the problem that we cannot estimate the regression coefficients of the g and the four common factor residuals simultaneously. For reasons of identification, we have to fix a single regression coefficient to zero. However, in terms of goodness of fit it makes no difference which regression coefficient is fixed to zero. Model selection therefore has to be based on standard errors of the parameter estimates, on comparisons with more parsimonious models, and on ease of interpretation of the results. As shown in Table 10, we fitted all identified models (denote model A to J in Table 10).

On the basis of parsimony and the approximate significance of the parameters (underlined in Table 10), models A to E are rejected. Models G to J involve g and one first order factor. Of these models, G provides the best fit and is interpretable. Model H fits as well, but includes a negative regression coefficient, which is hard to interpret. Comparing model F (g only) and model G (g and VC), we find that model G fits better in terms of the fit indices ($308.5 - 299.9 = 8.6$, $df=1$, $p=.003$). Furthermore compared to models A to E, model G fits well. The likelihood ratio is not significant ($299.9 - 296.7 = 3.2$, $df=2$, ns).

Based on these results, we favor the model, in which EA predicts g and VC directly. The prediction of PO,
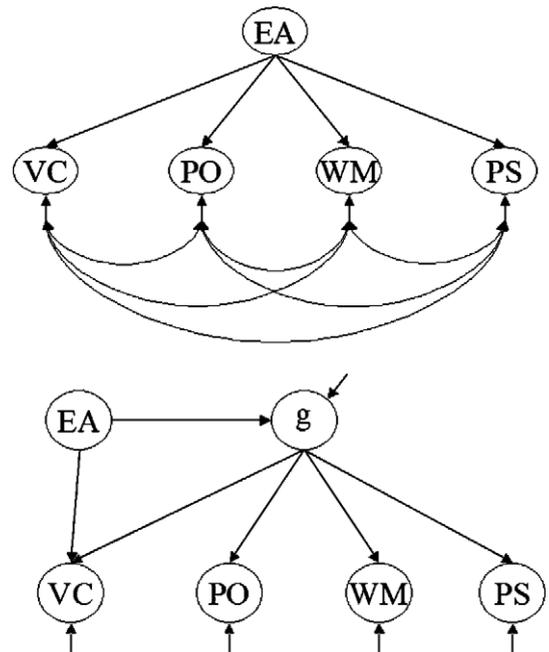

Fig. 4. Educational attainment as a predictor. Top: Educational attainment predicts WAIS-III common factors. Bottom: Educational attainment predicts second order factor g and first order residual VC.

Table 10
Goodness of fit indices, raw regression coefficients and standard errors, in models A to J, in which one or more common factors (g, VC, PO, WM, PS) are regressed on EA

Model	Goodness of fit indices					Raw regression coefficients and standard errors				
	χ^2	df	RMSEA	ECVI	CAIC	g	VC	PO	WM	PS
A	296.7	194	0.047	1.032	820.3	*6.05*	*.500*	*−1.73*	*−.45*	*0**
						1.15	1.03	1.07	.41	−.−
B						*4.66*	*1.81*	*−.35*	*0**	*1.48*
						.93	.84	.87	−.−	1.22
C						*4.32*	*2.14*	*0**	*−.11*	*1.86*
						.55	.61	−.−	.27	1.02
D						*6.58*	*0**	*−2.26*	*−.62*	*−.568*
						.91	−.−	.81	.33	1.22
E						*0**	*6.21*	*4.31*	*1.52*	*6.49*
						−.−	.55	.56	.25	.98
F	308.5	197	.048	1.044	810.3	*4.88*	*0**	*0**	*0**	*0**
						.52	−.−	−.−	−.−	−.−
G	299.9	196	.047	1.030	809.4	*4.52*	*1.80*	*0**	*0**	*0**
						.53	.57	−.−	−.−	−.−
H	300.5	196	.047	1.030	811.4	*6.13*	*0**	*−1.83*	*0**	*0**
						.80	−.−	.68	−.−	−.−
I	306.7	196	.048	1.040	814.9	*4.94*	*0**	*0**	*−.37*	*0**
						.54	−.−	−.−	.28	−.−
J	307.6	196	.049	1.050	817.4	*4.83*	*0**	*0**	*0**	*.97*
						.53	−.−	−.−	−.−	1.03

As explained in the text, model G is the preferred model. An asterisk denotes a fixed parameter (invariably fixed to zero). Note that the fit is identical for model A to E.

WM, and PS runs via g (see Fig. 4 bottom). Table 11 contains the correlation matrices of the common factors, g, and EA obtained in model G. Table 11 also contains the percentages of variance in the first order factors explained by EA, and by g, and EA combined. For instance in the males, we find that EA explained 22.8% of the variance of g. This is a little higher in the females: 26.9%. The percentages of total variance explained in the first order factors are high, ranging from about 43.2% (WM in the males) to 86.1% (PO in the females). The variance explained in VC (75% and 68%) and PO (79% and 86%) is much higher than the variance explained in WM (43% and 47%) and PS (58% and 45%). The percentage of variance of the first order factors explained by EA independently of g is relatively smaller (e.g., 10% WM in males, 23.2% PO in females), but here again we find that the prediction of VC and PO is much stronger.

12. Educational attainment as a predicted variable

In treating EA as a dependent variable in relation to the first order common factors in model A3 involves the following model:

$$EA_{ij} = \beta_4 \, \eta_{ij} + \zeta_{EAij}, \tag{25}$$

and, assuming η_{ij} and ζ_{EAij} are uncorrelated, the following decomposition of variance of EA:

$$\sigma_{EAi}^2 = \beta_4 \, \Sigma_{\eta i} \, \beta_4^t + \sigma_{\zeta EAi}^2, \tag{26}$$

where the (4×1) vector β_4 contains the regression coefficients, and $\sigma_{\zeta EAi}^2$ represents the residual variance of EA (see Fig. 3, top right). The matrices $\Sigma_{\eta i}$ is the covariance matrices of the common factors. Thus we decompose EA variance into a part that is due to the common factors VO, PO, WM, and PS (i.e., in the females, $\beta_4 \Sigma_{\eta f} \beta_4^t$), and an unexplained, residual component ($\sigma_{\zeta EAf}^2$).

Treating EA as a dependent variable in model B1, we regress EA on g and on the first order common factor residuals in the following regression model:

$$EA_{ij} = \beta_5 \, \omega_{ij} + \beta_6 \, g_{ij} + \zeta_{EAij}, \tag{27}$$

where β_6 is a single regression coefficient and β_5 is a vector containing 4 regression coefficients (Fig. 3 bottom right). Assuming ω_{ij}, g_{ij}, and ζ_{EAij} are uncorrelated, we have the following decomposition of variance:

$$\sigma_{EAi}^2 = \beta_5 \Sigma_{\omega i} \beta_5^t + \beta_6^2 \, \sigma_{gi}^2 + \sigma_{\zeta EAi}^2, \tag{28}$$

where $\Sigma_{\omega i}$ is the diagonal covariance matrix of the first order common factor residuals, and $\sigma_{\zeta EAi}^2$ is the variance of the residual. The decomposition of the variance

Table 11
Correlation matrices among first order common factors, g, and educational attainment (EA) in model G (EA predicts VC and g directly, see Table 10)

	VC	PO	WM	PS	g	EA
Male correlation matrix						
VC	1.000					
PO	0.763	1.000				
WM	0.562	0.586	1.000			
PS	0.653	0.681	0.502	1.000		
g	0.856	0.892	0.657	0.763	1.000	
EA	0.527	0.426	0.314	0.365	0.478	1.000
Total explained variance by EA and g						
	75.0	79.6	43.2	58.2		
Variance explained by EA						
	27.8	18.2	10.0	13.3	22.8	
Female correlation matrix						
VC	1.000					
PO	0.757	1.000				
WM	0.560	0.637	1.000			
PS	0.553	0.629	0.465	1.000		
g	0.816	0.928	0.686	0.678	1.000	
EA	0.537	0.482	0.356	0.352	0.519	1.000
Total explained variance by EA and g						
	68.3	86.1	47.1	45.9		
Variance explained by EA						
	28.9	23.2	12.7	12.4	26.9	

Total variance explained in the first order common factors VC, PO, WM, and PS is due to both g and EA. Variance explained by EA concerns only the variance explained by EA. Note that the percentages of variance of PO, WM, and PS are explained indirectly via g, i.e., there is no direct regression of EA on these factors (see Fig. 4 bottom).

in EA is decomposed into a part due to the regression on g (i.e., in the males $\beta_6^2\sigma_{gm}^2$), a part due to the regression on the first order common factor residuals ($\beta_1\Sigma_{\omega m}\beta_1^t$), and a residual ($\sigma_{\xi xm}^2$).

13. Educational attainment as a predicted variable: results

In these analyses, we treated EA as a dependent variable, i.e., we regressed EA on the common factors (Eq. (25)). The fit of this model is acceptable: $\chi^2(187)=272.2$, RMSEA=.043, ECVI=1.013, and CAIC=847.6. The common factors explain 31.6% of the variance in EA in the male sample, and 32.5% in the female sample. Inspection of the regression coefficients reveals that the coefficients associated with the factors PO and WM are not significant. The raw estimates are .37 (VO, s.e. .070), −.078 (PO, s.e. .090), −.036 (WM,

s.e. 0.054), and .186 (PS, s.e. .070). Refitting the model with the non-significant coefficients fixed to zero did not result in any appreciable deterioration in fit: $\chi^2(189)=273.4$, RMSEA=.043, ECVI=1.00, and CAIC=835.4. The explained variance in EA remains about 30% (31.1% in males, 30.1% in females). The raw estimates of the remaining regression coefficients are significant judging by their standard errors: .317 (VO, s.e. .046) and .136 (PS, s.e. .049). The path diagram of this model is shown in Fig. 5 (top).

Finally we regressed EA on the primary factors (VO, PO, WM, PS) and on the second order factor g (Eq. (27)). The fit of this model is acceptable: $\chi^2(193)=291.3$, RMSEA=.046, ECVI=1.024, and CAIC=822.3. The estimates of the raw regression coefficients are 0.254 (VC, s.e. .072), 0.141 (PO, s.e. .157), 0.058 (WM, s.e. .071), 0.160 (PS, s.e. .072), and −.665 (g, s.e. .648). Judging by the standard errors the regression on PO, WM, and g are not significant. In addition, the sign of the regression coefficient associated with g is negative, and thus hard to interpret. We refitted the model with the regression on g fixed to zero. As expected the fit of this model is almost unchanged: $\chi^2(194)=292.5$, RMSEA=.046, ECVI=1.023, and CAIC=816.8. The estimates of the raw regression coefficients are 0.170 (VC, s.e. .037), −.022 (PO, s.e.
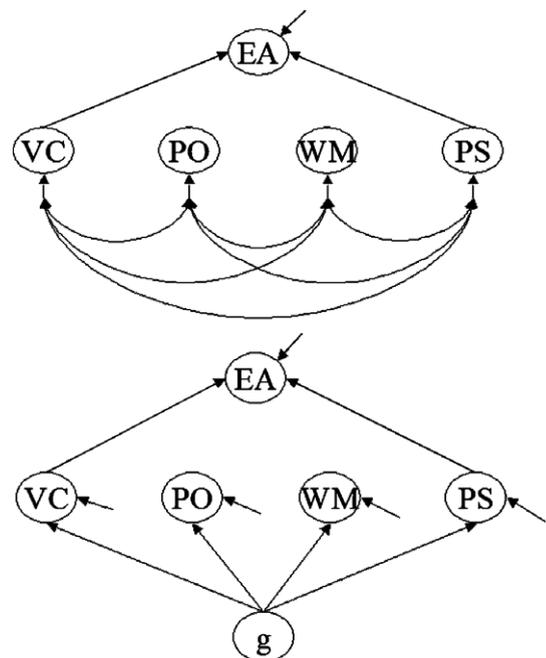

Fig. 5. Educational attainment as a dependent variable. Top: VC and PS predict education attainment in first order common factor model. Bottom: First order factors VC and PS predict education attainment in second order common factor model.

.036), −.006 (WM, s.e. .039), and 0.104 (PS, s.e. .040). In this model about 31% of the variance in EA is explained in both the male and female samples. As we again find that the regression coefficients associated with WM and PO are not significant judging by the standard errors, we refitted the model with these regression coefficients fixed to zero. The fit of this final model is about the same $\chi^2(196)=293.0$, RMSEA=.045, ECVI=1.016, and CAIC=803.5. The variance explained in EA by VC and PS remained about 31% in both samples. The path diagram associated with this model is shown in Fig. 5 (bottom).

14. Discussion

We have obtained the following results concerning the factor structure of the WAIS in the present Spanish data. Overall, the expected factor structure fitted the data well. In investigating measurement (i.e., strict factorial) invariance with respect to gender in the first order factor model (model A4), we found that four subtest mean differences could not be accounted for by the common factor mean differences. These subtests are VOC, ARIT, INFO, and PC. The local misfit is caused by a female advantage on VOC, a male advantage on INFO (both are incompatible with the zero mean difference on the common factor VC), a male advantage with respect to ARIT (too great to be accounted for by the mean difference in the common factor PO), and the absence of sex differences with respect to PC (incompatible with the difference on common factor PO). While clearly detectable, the misfit with respect to means of these subtests is not very large in terms of effect (ranging from .10 to .16 female std. units). van der Sluis et al. (in press), in similar analyses of Dutch adult WAIS-IIIR data, also found that sex related mean difference on the INFO subtest could not be accounted for by a sex mean difference on VC. The male advantage with respect to Info is consistent with the results of Lynn, Irwing, and Cammock (2001). The common factor mean differences could account for the observed mean differences on the remaining 10 subtests.

With respect to the mean structure model, we found in the first order common factor model that sex mean differences in WAIS-III subtest scores are explained by latent mean differences in the common factors Perceptual Organization (PO) and Working Memory (WM). As shown in Table 5, in the females the means of the common factors PO and WM are −.232 and −.211, respectively (.23 and .21 in male std. units, .28 and .25 in female std. units). We note also that the lower means

are associated with lower standard deviation in the female sample .826 (PO) and .826 (WM). van der Sluis et al. (in press) similarly observed a male advantage on PO and WM, and no gender difference on VC. However, in contrast to the present results, van der Sluis observed a significant female advantage in PS. This may reflect a difference between the Dutch and the Spanish samples.

We also considered sex differences in the second order common factor. In this model, we found that sex differences in the first order common factors PO and WM provided a parsimonious and well fitting account of the sex differences in the subtests. Thus here we observed similar differences in factor means as in the first order factor model.

In conclusion, the present results indicate that sex differences in WAIS-III subtest scores are *not* due to *g*. This conclusion is consistent with Colom, Garcia, et al. (2002), van der Sluis et al. (in press), and supports Ankney's hypothesis that the link between brain volume, sex differences and IQ test scores is due to differences in primary common factors (PO and WM). We found no sex-related mean difference in VC and PS. The absence of gender differences in VC is consistent with the results of the meta-analysis of Hyde and Linn (1988). The present conclusion runs counter to both Nyborg (2003) and Lynn (1994, 1999), who suggest that *g* is the source of sex differences. Presumably, this is due to the differences in statistical methodology. The present approach, comprehensive model-based modeling based on the theory of factorial invariance (Meredith, 1993), is superior to the methods used by Nyborg and Lynn.

In addition to sex differences, we investigated the relationship between the common factors of the WAIS-III and educational attainment (EA). Again we conducted analyses in both the first order common factor model and in the second order common factor model to assess the role of *g*. Because the exact causal model linking EA and IQ common factors is unknown, we treated EA both as a dependent and as an independent variable. This strategy is preferable to, say, assigning EA the role of dependent variables, as it provides detailed information on the inter-relationship of these variables.

The regression of the common factors on EA indicated that EA is a significant *predictor* of all the first order common factors. However, the variance explained by educational attainment differs considerably from factor to factors: approximately 27% (VC), 20% (PO), 10% (WM), and 15% (PS). In the second order common factor model, we found that educational attainment

predicted *g* well. A single additional path was required that directly linked the residual of VC and educational attainment. Here we thus find that *g* plays an important role in linking the effects of EA on the common factors VC, PO, WM, and PS.

In treating EA as a *dependent* variable in relationship to the common factors, we obtained the following results. In the first order factor model, we found that the common factors VC and PS predict EA. The common factors PO and WM do not contribute to this prediction. In the second order common factor model, we again found that the common factors VC and PS predict educational attainment. The second order common factor *g* is related with EA, but the relation runs via these first order factors. The direct regression of educational attainment on *g* is not significant. Overall we observed fairly minor gender differences in the variances explained in the analyses of the WAIS subtest scores and EA.

In conclusion, while there is no doubt that IQ test scores and educational attainment are related, the nature of the relationship is subtle. Specifically, the hypothesis that *g* is the dominant predictor of educational attainment is *not* supported in these data. Rather the first order factors VC and PS predict educational attainment. However, when educational attainment is treated as the independent variable, we find that it predicts the IQ test scores mainly through the *g* factor. These results demonstrate that the well-established correlation between *g* and EA (over .5 in the present data) says very little about the structure of the regression of EA on *g* and its indicators (VC, PO, WM, PS). As we have demonstrated, this structure can investigated using covariance structure modeling.

References

Ackerman, P. h. L., & Lohman, D. F. (2003). Education and g. In H. Nyborg (Ed.), *The scientific study of general intelligence. Tribute to Arthur R. Jensen* (pp. 275–292). Amsterdam: Elsevier Science Ltd.

Aluja-Fabregat, A., Colom, R., Abad, F., & Juan-Espinosa, M. (2000). Sex differences in general intelligence defined as g among young adolescents. *Personality and Individual Differences*, 28, 813–820.

Ankney, C. D. (1992). Sex differences in brain size and mental abilities: A mismeasure of women, too? *Intelligence*, 16, 329–336.

Ankney, C. D. (1995). Sex differences in brain size and mental abilities: Comment on R. Lynn and D. Kimura. *Personality and Individual Differences*, 18, 423–424.

Ashton, M. C., & Lee, K. (in press). Problems with the method of corrected vectors. *Intelligence*.

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

Bollen, K. A., & Long, J. S. (1993). Introduction. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models*. Newbury Park: Sage Publications.

Brody, N. (1992). *Intelligence*. San Diego: Academic Press.

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

Byrne, B. M., Shavelson, R. J., & Muthén, B. O. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105, 456–466.

Ceci, S. J., & Williams, W. M. (1997). Schooling, intelligence and income. *American Psychologist*, 52, 1051–1058.

Colom, R., Abad, F. J., Garcia, L. F., & Juan-Espinosa, M. (2002). Education, Wechsler's full scale IQ, and g. *Intelligence*, 30, 449–462.

Colom, R., Garcia, L. F., Juan-Espinosa, M., & Abad, F. J. (2002). Null sex differences in general intelligence: Evidence from the WAIS-III. *The Spanish Journal of Psychology*, 5(1), 29–35.

Colom, R., Juan-Espinosa, M., Abad, F., & Garcia, L. F. (2000). Negligible sex differences in general intelligence. *Intelligence*, 28(1), 57–68.

Colom, R., & Lynn, R. (2004). Testing the developmental theory of sex differences in intelligence on 12–18 year olds.

Dolan, C. V. (2000). Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis. *Multivariate Behavioral Research*, 35, 21–50.

Dolan, C. V., & Hamaker, E. L. (2001). Investigating Black–White differences in psychometric IQ: Multi-group confirmatory factor analyses of the WISC-R and K-ABC and a critique of the method of corrected factors. In F. Columbus (Ed.), *Advances of Psychological Research*, vol. 6. (pp. 31–59). Huntington: Nova Science Publishers.

Dolan, C. V., Roorda, W., & Wicherts, J. M. (2004). Two failures of Spearman's hypothesis: The GATB in Holland and the JAT in South Africa. *Intelligence*, 32, 155–173.

Gray, J. R., & Thompson, P. M. (2004). Neurobiology of intelligence: Science and ethics. *Nature Reviews Neuroscience*, 5, 471–482.

Gustafsson, J. -E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, 8, 179–203.

Gustafsson, J. -E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 35–71). Hillsdale, NJ: Erlbaum.

Gustafsson, J. E. (1992). The relevance of factor analysis for the study of group differences. *Multivariate Behavioral Research*, 27, 239–247.

Herrnstein, R. J., & Murray, C. (1994). *The Bell Curve: Intelligence and class structure in American life*. New York: Free Press.

Horn, J. L. (1997). On the mathematical relationship between factor or component coefficients and differences in means. *Cahiers de Psychologie Cognitive*, 16, 721–728.

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

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, 104, 53–69.

Jensen, A. R. (1998). *The g factor. The science of mental ability*. Westport: Praeger.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409–426.

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models*. Newbury Park: Sage Publications.

Jöreskog, K. G., & Sörbom, D. (1999). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software International.

Langford, E., Schwertman, N., & Owens, M. (2001). Is the property of being positively correlated transitive? *The American Statistician*, *55*, 322–325.

Little, T. D. (1997). Mean and covariance structures (MACS) analysis of cross-cultural data: practical and theoretical issues. *Multivariate Behavioral Research*, *32*, 53–76.

Loehlin, J. C. (2000). Group differences in intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 176–198). Cambridge: Cambridge University Press.

Lubke, G. H., Dolan, C. V., & Kelderman, H. (2001). Investigating group differences using Spearman's hypothesis: An evaluation of Jensen's method. *Multivariate Behavioral Research*, *36*, 299–324.

Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). On the relationship between sources of within- and between-group differences and measurement invariance in the common factor model. *Intelligence*, *31*(6), 543–566.

Lynn, R. (1994). Sex differences in intelligence and brain size: A paradox resolved. *Personality and Individual Differences*, *17*, 257–271.

Lynn, R. (1999). Sex differences in intelligence and brain size: A developmental theory. *Intelligence*, *27*, 1–12.

Lynn, R., Irwing, P., & Cammock, T. (2001). Sex differences in general knowledge. *Intelligence*, *30*, 27–39.

Mackintosh, N. J. (1998). *IQ and human intelligence*. Oxford, England: Oxford University Press.

Marsh, H. W., & Grayson, D. A. (1990). Public/Catholic differences in the high school and beyond data: A multi-group structural equation modeling approach to testing mean differences. *Journal of Educational Statistics*, *15*, 199–235.

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, *13*, 127–143.

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

Millsap, R. E. (1997). The investigation of Spearman's hypothesis and the failure to understand factor analysis. *Cahiers de Psychologie Cognitive*, *16*, 750–757.

Millsap, R. E., & Everson, H. (1991). Confirmatory measurement model comparisons using latent means. *Multivariate Behavioral Research*, *26*, 479–497.

Nyborg, H. (2003). Sex differences in *g*. In H. Nyborg (Ed.), *The scientific study of general intelligence* (pp. 187–222). Amsterdam: Pergamon.

Nyborg, H. (2004). Sex differences in g. In H. Nyborg (Ed.), *The scientific study of general intelligence. Tribute to Arthur R. Jensen* (pp. 187–222). Amsterdam: Elsevier Science Ltd.

Ployhart, R. E., & Oswald, F. L. (2004). Applications of Mean and Covariance Structure Analysis: Integrating correlational and experimental approaches. *Organizational Research Methods*, *7*, 27–65.

Psychological Corporation (1997). *WAIS-III WMS-III Technical manual*. San Antonio: Harcourt Brace & Co.

Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research*, *23*, 51–67.

Rock, D. A., Werts, C. E., & Flaugher, R. L. (1978). The use of analysis of covariance structures for comparing the psychometric properties of multiple variables across populations. *Multivariate Behavioral Research*, *13*, 403–418.

Rushton, J. P. (1992). Cranial capacity related to sex, rank and race in a stratified sample of 6325 military personnel. *Intelligence*, *16*, 401–413.

Rushton, J. P., & Ankney, C. D. (1996). Brain size and cognitive ability: Correlations with age, sex, social class, and race. *Psychonomic Bulletin and Review*, *3*, 21–36.

Saklofske, D. H., Hildebrand, D. K., & Gorsuch, R. L. (2000). Replication of the factor structure of the Wechsler Adult Intelligence Scale-Third edition with a Canadian sample. *Psychological Assessment*, *12*, 436–439.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research—Online*, *8*, 23–74.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*, 53–61.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, *27*, 229–239.

TEA, S. A. (1998). *Manual del WAIS III (WAIS III Manual)*. Madrid: TEA Ediciones.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–69.

van der Sluis, S., Posthuma, D., Dolan, C. V., de Geus, E. J. C., Colom, R., and Boomsma, D. I. (in press). *Gender differences on the Dutch WAIS-III*.

Vernon, P. A., Wickett, J. C., Bazana, P. G., & Stelmack, R. M. (2000). The neuropsychology and psychophysiology of human intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 245–266). Cambridge: Cambridge University Press.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta analysis and consideration of critical variables. *Psychological Bulletin*, *117*, 250–270.

Wahlsten, D. (1997). The malleability of intelligence is not constrained by heritability. In B. Devlin, S. E. Fienberg, S. P. Resnick, & K. Roeder (Eds.), *Intelligence, genes, and success. Scientists respond to The Bell Curve* (pp. 71–87). New York: Springer Verlag.

Ward, L. C., Ryan, J.J, & Axelrod, B. N. (2000). Confirmatory factor analysis of the WAIS-III standardization data. *Psychological Assessment*, *12*(3), 341–345.

Wicherts, J. M., & Dolan, C. V. (2004). A cautionary note on the use of information fit indices in covariance structure modeling with means. *Structural Equation Modeling*, *11*, 45–50.

Wicherts, J. M., Dolan, C. V., & Hessen, D. J. (in press). Stereotype threat and group differences in test performance: A question of measurement invariance. *Journal of Personality and Social Psychology*.

Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P., van Baal, G. C. M., Boomsma, D. I., et al. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence*, *32*, 509–537.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, & M. Windle (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.

Winship, C., & Korenman, S. (1997). Does staying in school make you smarter? The effect of education on IQ in the Bell Curve. In B. Devlin, S. E. Fienberg, S. P. Resnick, & K. Roeder (Eds.), *Intelligence, genes, and success. Scientists respond to The Bell Curve* (pp. 215–234). New York: Springer Verlag.