Multivariate Behavioral Research, *35* (1), 21-50

Investigating Spearman's Hypothesis by Means of Multi-Group Confirmatory Factor Analysis

Conor V. Dolan
University of Amsterdam

Differences between blacks and whites on cognitive ability tests have been attributed to a fundamental difference between these groups in general intelligence (or *g*, as it is denoted). The hypothesized difference in *g* gives rise to Spearman's hypothesis, which states that the differences in the means of the tests are related to the tests' factor loadings on *g*. Jensen has investigated this hypothesis by correlating differences in means and tests' *g* loadings. The aim of the present article is to investigate B-W differences using multi-group confirmatory factor analysis. The advantages of multi-group confirmatory factor analysis over Jensen's test of Spearman's hypothesis are discussed.

A published data set is analyzed. Strict factorial invariance is tested and judged to be tenable. Various models are tested, which do and do not incorporate *g*. It is observed that it is difficult to distinguish between several hypotheses, including and excluding g, concerning group differences. The inability to distinguish between competing models using multi-group confirmatory factor analysis makes it difficult to draw clear conclusions about the exact nature of black-white differences in cognitive abilities. The implications of the results for Jensen's test of Spearman's hypothesis are discussed.

Introduction

Jensen has proposed that the differences between blacks (B) and whites (W) in psychometric intelligence test scores are attributable mainly to a difference in general intelligence, or g, as it is denoted (Jensen, 1985; 1992; 1998). One way in which Jensen has investigated this proposition is by formulating and testing Spearman's hypothesis. There are two versions of this hypothesis, a weak version and a strong version. The strong version states that "variation in the size of the mean W-B difference across various tests is solely a positive function of variation in the tests' *g* loadings" (Jensen, 1998, p. 372). The weak version, which is somewhat ill-defined, states that the variation is *mainly* a positive function of variation in the tests' *g* loadings. The rationale is that a test that is

---

The research of Conor Dolan was made possible by a fellowship of the Royal Netherlands Academy of the Arts and Sciences. The author may be reached at University of Amsterdam, Developmental Psychology, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands or e-mail to op_dolan@macmail.psy.uva.nl.

I thank Gitta Lubke, 2 anonymous reviewers and the editor, Roger Millsap, for their constructive criticism of this article.

strongly influenced by *g* (i.e., characterized by a large factor loading on *g*) will display a large B-W difference in mean. Support for Spearman's hypothesis has been presented in the form of correlations (designated "Spearman correlations") between the vector of differences in means and factor loadings. In calculating these correlations, Jensen (1985; 1992) advocates a procedure consisting of the following steps: (a) Exploratory factor analyses of psychometric data are carried out in representative samples of blacks and whites, separately, to extract factor loadings of the tests on *g* (Jensen suggests that this can be done in a number of ways, see Jensen & Weng, 1994); (b) Factorial invariance over groups is established by calculating measures of factorial congruence of the factor loadings in the white and the black samples; (c) Differences in means are standardized by dividing by the pooled standard deviations; (d) The standardized differences in means are correlated with the *g* factor loadings. A mean (Spearman's rho) correlation of .59 (sd .12), obtained using this procedure, has been reported on the basis of 11 studies (Jensen, 1985). For other applications of Jensen's test, see Naglieri and Jensen (1987), Lynn and Owen (1994), te Nijenhuis and van der Flier (1997), and Rushton (1999).

A number of commentators and researchers, including the users of Jensen's test, view the reported correlations as strongly supportive of the weak version of the hypothesis. Schönemann (1997a, pp. 666-667) provides a brief survey. Others, however, have criticized Jensen's test. Notably Schönemann (1997a, 1997b), in a special issue of Cahiers de Psychologie Cognitive (CPC), claims to have proven that Jensen's test is fundamentally flawed, because the correlations between the factor loadings and the differences in means are necessarily positive. A detailed discussion is beyond the scope of the present article, and much has been written on the subject by the contributors to the special CPC issue. One recommendation that emerges repeatedly in the comments on Schönemann's critique is that a model fitting approach, based on the theory of factorial invariance, should be favoured to Jensen's test in investigating B-W differences in IQ data (Horn, 1997; Millsap, 1997a; Gustafsson, 1992; Dolan, 1997). Confirmatory factor analysis has been applied to B-W differences in measures of cognitive abilities (law students' exam scores in the case of Rock, Werts, & Flaughter, 1978; children's WISC scores in the case of Gustafsson, 1992). However, these analyses have been limited to just one specific factor model.

The present article has several related objectives. First, we present a variety of multi-group confirmatory factor models that are suitable to investigate B-W difference in cognitive ability test scores. The models are specific instances of the linear common factor model[1] and, as such, are not new.

---

[1] At present this is the preferred approach to the modeling of psychometric data relating to cognitive abilities (e.g., Gustafsson, 1984, 1988).

The models vary in the manner in which the test scores are modeled. Some, but not all, of these models incorporate *g*. Second, we discuss the relationship between Spearman's hypothesis and the hypothesis of factorial invariance, and the advantages of multi-group confirmatory factor analysis over Jensen's test in investigating B-W differences in psychometric data relating to cognitive abilities. Millsap (1997a) has already noted that Spearman's hypothesis addresses a single aspect of the more comprehensive hypothesis of factorial invariance. Third, we apply the models presented to a published data set (Jensen & Reynolds, 1982). Jensen's test of Spearman's hypothesis is evaluated in the light of the results of fitting the various models. It is important to consider a variety of models in investigating the role of *g* in B-W differences for two related reasons. First, given results for a variety of models, we may evaluate the goodness of fit of a subset of models including *g* by considering the goodness of fit of alternative models. If, following such an evaluation, the models incorporating *g* are found to fit well, this may be taken as support in favor of *g*, and its role in B-W differences. In Jensen's procedure, competing models are not considered. Second, once we have identified models that fit relatively well, we can better evaluate the reported Spearman correlations. Ideally, these models should include *g* as an important factor in B-W differences. However, if these models do not incorporate *g*, or if models with and without *g* cannot be distinguished, the Spearman correlation cannot be safely interpreted as indicative of the importance of *g* in B-W differences. Our results suggest that the Spearman correlations reported for the data set that we analyze (Jensen & Reynolds, 1982) cannot safely be interpreted as support for the central role of *g* in B-W differences.

In the next section, we present various multi-group confirmatory factor models in general form. Subsequently we introduce constraints that are necessary to achieve a meaningful comparison of the groups, and constraints that are required to render the parameters in the models identified. We discuss the relationship between these models and Spearman's hypothesis. We fit the models to a data set published in Jensen and Reynolds (1982), and we discuss the results. The article concludes with a general discussion.

Multi-Group Confirmatory Factor Models

Multi-group confirmatory factor analysis (MGCFA) is a well established technique to investigate group differences in means and covariances within the common factor model (e.g., Jöreskog, 1971; Sörbom, 1974; Rock, Werts, & Flaugher, 1978; Byrne, Shavelson & Muthén, 1989; Marsh & Grayson, 1990; Millsap & Everson, 1991; Dolan &

Molenaar, 1994; Little, 1997). Standard software can be used to carry out MGCFA, such as EQS (Bentler, 1992), Mx (Neale, 1997), or LISREL (Jöreskog & Sörbom, 1993). Here we consider three general models, which we call the 1st order multi-group confirmatory factor model (MGCFM), the 2nd order MGCFM, and the 1st order alternative MGCFM (aMGCFM). We first present the models in a general form and subsequently introduce identifying and substantive constraints.

First Order Multi-Group Confirmatory Factor Model

Let $\mathbf{y}_{ij}$ denote the observed p-dimensional random column vector of subject j in population i. The following factor model is assumed to hold for the observations $\mathbf{y}_{ij}$

(1) $$\mathbf{y}_{ij} = \boldsymbol{\nu}_i + \boldsymbol{\Lambda}_i \boldsymbol{\eta}_{ij} + \boldsymbol{\varepsilon}_{ij},$$

where $\boldsymbol{\eta}_{ij}$ is a q-dimensional random vector of correlated common factor scores ($q < p$), and $\boldsymbol{\varepsilon}_{ij}$ is a p-dimensional vector of specific scores unique to each observed variable. These specific scores include random measurement error. The ($p \times q$) matrix $\boldsymbol{\Lambda}_i$ contains factor loadings, and the ($p \times 1$) vector $\boldsymbol{\nu}_i$ contains intercepts. In confirmatory factor analysis, one usually has a good idea about the relationship between the observed variables and the common factors so that many elements in $\boldsymbol{\Lambda}$ will be fixed to equal zero. We assume that $\boldsymbol{\varepsilon}_{ij} \sim N_p(0, \boldsymbol{\Theta}_i)$ and $\boldsymbol{\eta}_{ij} \sim N_q(\boldsymbol{\alpha}_i, \boldsymbol{\Psi}_i)$, where the ($p \times p$) diagonal matrix $\boldsymbol{\Theta}_i$ is positive, and the ($p \times q$) covariance matrix $\boldsymbol{\Psi}_i$ is positive definite. As the specific scores include both random measurement error and unique systematic effects, the specification that $\boldsymbol{\varepsilon}_{ij} \sim N_p(0,\boldsymbol{\Theta}_i)$ represents a strong assumption (Meredith, 1993; Sörbom, 1975). We return to this assumption in the discussion. The observed variables are distributed $\mathbf{y}_{ij} \sim N_p(\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)$, where, assuming $E[(\boldsymbol{\eta}_{ij} - \boldsymbol{\alpha}_i)\boldsymbol{\varepsilon}_{ij}^t] = 0$,

(2) $$\boldsymbol{\mu}_i = \boldsymbol{\nu}_i + \boldsymbol{\Lambda}_i \boldsymbol{\alpha}_i$$

(3) $$\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i \boldsymbol{\Psi}_i \boldsymbol{\Lambda}_i^t + \boldsymbol{\Theta}_i.$$

We call this the *1st order MGCFM* (Sörbom, 1974).

Second Order Multi-Group Confirmatory Factor Model

In Jensen's tests of Spearman's hypothesis, the common factor *g* features as a second order common factor (e.g., see Jensen & Reynolds, 1982). To accommodate this conceptualization of *g* within the common factor model we extend the model as follows. Let $\boldsymbol{\eta}_{ij} = (\boldsymbol{\Gamma}_i \boldsymbol{\xi}_{ij} + \boldsymbol{\zeta}_{ij})$, where $\boldsymbol{\Gamma}_i$ is a $(q \times r)$ matrix of loadings of the q first order factor scores, $\boldsymbol{\eta}_{ij}$ on the r second order factor scores, $\boldsymbol{\xi}_{ij}$, and $\boldsymbol{\zeta}_{ij}$ is a q-dimensional vector of random (first order) residual terms. The model for the observations is now:

$$\mathbf{y}_{ij} = \boldsymbol{\nu}_i + \boldsymbol{\Lambda}_i (\boldsymbol{\Gamma}_i \boldsymbol{\xi}_{ij} + \boldsymbol{\zeta}_{ij}) + \boldsymbol{\varepsilon}_{ij}, \tag{4}$$

We assume that $\boldsymbol{\zeta}_{ij} \sim N_q(\boldsymbol{\alpha}_i, \boldsymbol{\Psi}^*_i)$, where the asterisk indicates that the covariance matrix is diagonal, and $\boldsymbol{\xi}_{ij} \sim N_r(\boldsymbol{\kappa}_i, \boldsymbol{\Phi}_i)$. Furthermore, we assume that $E[(\boldsymbol{\zeta}_{ij} - \boldsymbol{\alpha}_i)\boldsymbol{\varepsilon}_{ij}^t] = E[(\boldsymbol{\xi}_{ij} - \boldsymbol{\kappa}_i)\boldsymbol{\varepsilon}_{ij}^t] = E[(\boldsymbol{\xi}_{ij} - \boldsymbol{\kappa}_i)(\boldsymbol{\zeta}_{ij} - \boldsymbol{\alpha}_i)^t] = 0$. So we have $\mathbf{y}_{ij} \sim N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, where

$$\boldsymbol{\mu}_i = \boldsymbol{\nu}_i + \boldsymbol{\Lambda}_i \boldsymbol{\alpha}_i + \boldsymbol{\Lambda}_i \boldsymbol{\Gamma}_i \boldsymbol{\kappa}_i \tag{5}$$

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i(\boldsymbol{\Gamma}_i \boldsymbol{\Phi}_i \boldsymbol{\Gamma}_i^t + \boldsymbol{\Psi}^*_i)\boldsymbol{\Lambda}_i^t + \boldsymbol{\Theta}_i \ . \tag{6}$$

We call this the *2nd order MGCFM*. The Schmid-Leiman hierarchical factor analysis (Schmid & Leiman, 1957) that Jensen uses to extract *g* as a second order factor (e.g., Naglieri & Jensen, 1987) can be viewed as the exploratory version of the model in Equation 6.

Identifying Constraints

We introduce various constraints into these models, which are required to ensure identification of the parameters, or necessary to ensure the meaningful comparison of the groups. In all models we specify that the factor loadings ($\boldsymbol{\Lambda}$ and, in the second order factor model, $\boldsymbol{\Gamma}$) and the specific variances ($\boldsymbol{\Theta}$) are equal over the groups. These restrictions are amenable to statistical testing. Throughout we retain the restriction that $\boldsymbol{\Theta}$ is diagonal, and we introduce sufficient fixed zeroes in the matrix $\boldsymbol{\Lambda}$ to avoid rotational indeterminacy given correlated common factors. To determine the variance of the common factors, we fix certain elements in $\boldsymbol{\Lambda}$ to equal 1 (see Table 2). This allows us to estimate latent (co)variances. In the second order factor model, likewise we fix an element in $\boldsymbol{\Gamma}$ to equal 1, so that we can estimate the variances of *g* in the two groups.

As explained by Sörbom (1974), we are bound to model latent differences in means instead of latent means for reasons of identification. We therefore equate the observed means with the vector of intercepts $\boldsymbol{\nu}$ in the white group and estimate the latent differences in means, $\boldsymbol{\alpha}_b - \boldsymbol{\alpha}_w$, in the black group.[2] Given these restrictions, the 1st order MGCFM is:

(7) $$\boldsymbol{\mu}_w = \boldsymbol{\nu}$$

(8) $$\boldsymbol{\Sigma}_w = \boldsymbol{\Lambda}\boldsymbol{\Psi}_w\boldsymbol{\Lambda}^t + \boldsymbol{\Theta}$$

(9) $$\boldsymbol{\mu}_b = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\delta}$$

(10) $$\boldsymbol{\Sigma}_b = \boldsymbol{\Lambda}\boldsymbol{\Psi}_b\boldsymbol{\Lambda}^t + \boldsymbol{\Theta},$$

where the vector of differences in latent means, $\boldsymbol{\delta}$, equals $\boldsymbol{\alpha}_b - \boldsymbol{\alpha}_w$. The choice to equate the observed means in the white group with $\boldsymbol{\nu}$ is arbitrary; equating $\boldsymbol{\nu}$ and the observed means in the black group does not substantively change the results. In this model, all common factors contribute to the observed differences in means (i.e., all components of $\boldsymbol{\delta}$ are estimated). As mentioned, the introduction of fixed 1s in $\boldsymbol{\Lambda}$, enables us to specify the elements in $\boldsymbol{\Psi}_w$ and $\boldsymbol{\Psi}_b$ as free parameters.

One may also test the hypothesis that the groups differ with respect to a subset of the common factors. For instance, given three correlated common factors, it is possible that two groups differ only with respect to the (unobserved) causes of the first common factor, $\boldsymbol{\eta}_1$. However, because the factors are correlated, it is reasonable to assume that the causes of the $\boldsymbol{\eta}_1$ are shared partially by the causes of $\boldsymbol{\eta}_2$ and $\boldsymbol{\eta}_3$. Due to the overlap between the causes of the factors, the differences with respect to the causes of the first factor will also be evident in the second and third common factor, $\boldsymbol{\eta}_2$ and $\boldsymbol{\eta}_3$. Dolan and Molenaar (1994) present a model that takes into account the secondary effects on other common factors of primary differences with respect to a subset of common factors. Their model is based on the assumption that the differences between the groups can be modeled as the result of selection effects at the level of a subset of the common factors. We assume that the first order factor model holds, as shown in Equations 7 and 8, in (say) the white population. The black population is supposed to differ systematically from the white population with respect to some of the common factors. Differences between the two groups are modeled through

[2] The parameterization based on latent differences in means gives rise to a model for the observed means which comprises a consistent set of linear equations with a unique solution. In modeling latent means, rather than differences in means, this is not the case.

a undefined process of selection on the basis of the factor scores. Factor scores are not actually calculated, but the weight matrix used to calculate factor scores does feature in the derivation of the model. There are a number of ways to calculate such a weight matrix (Saris, de Pijper, & Mulder, 1978; Lawley & Maxwell, 1971). Here we limit our attention to the factor scores regression matrix ($p \times q$), which is calculated as follows: $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}\mathbf{\Lambda}^t\mathbf{\Psi}_w$. If the groups differ with respect to a number of the components of the vector of factor scores, $\mathbf{\eta} = \mathbf{\Omega}^t\mathbf{y}$, the covariance matrix and mean vector in the black group may be modeled as follows (see Dolan & Molenaar, 1994; Equations 13 and 14):

$$\boldsymbol{\mu}_b = \boldsymbol{\nu} + \mathbf{\Sigma}_w \mathbf{\Omega}\mathbf{\Psi}_w\boldsymbol{\delta} = \boldsymbol{\nu} + \mathbf{\Lambda}\mathbf{\Psi}_w\boldsymbol{\pi} \tag{11}$$

$$\mathbf{\Sigma}_b = \mathbf{\Sigma}_w + \mathbf{\Sigma}_w(\mathbf{\Omega}\mathbf{\Delta}\mathbf{\Omega}^t)\mathbf{\Sigma}_w = \mathbf{\Lambda}(\mathbf{\Psi}_w \mathbf{\Delta}\mathbf{\Psi}_w^t + \mathbf{\Psi}_w)\mathbf{\Lambda}^t + \mathbf{\Theta}, \tag{12}$$

where ($\mathbf{\Psi}_w\boldsymbol{\pi}$) represents the vector of latent differences in means, and ($\mathbf{\Psi}_w\mathbf{\Delta}\,\mathbf{\Psi}_w^t + \mathbf{\Psi}_w$), the covariance structure of the common factors in the black group. We call this model the 1st order aMGCFM. The decision to model the data of the white group using Equations 7 and 8, and the data of the black group using Equations 11 and 12 is arbitrary. Modeling the data of the whites using Equations 11 and 12, and the data of the blacks using Equations 7 and 8 (with the appropriate substitution of subscripts *w* and *b*), does not substantively change the results.

The ($q \times 1$) vector $\boldsymbol{\pi}$ and the ($q \times q$) symmetric matrix $\mathbf{\Delta}$ are a function of the differences in means and the (co)variances of the latent variable(s) with respect to which the white and the black groups differ (Dolan & Molenaar, 1994; Eqs. 9 and 10). Because we are considering differences with respect to a subset of common factors, both $\boldsymbol{\pi}$ and $\mathbf{\Delta}$ will contain zero elements. For instance, let us suppose that we specify that the groups differ primarily with respect to the second of three common factors. As mentioned above, primary differences in one common factor will introduce secondary differences in the other common factors, because the common factors, being correlated, share causes. The primary effect is modeled by parameters in $\boldsymbol{\pi}$ and $\mathbf{\Delta}$. For instance, in the present case, we have

$$\mathbf{\Delta} = \begin{matrix} 0 & 0 & 0 \\ 0 & \mathbf{\Delta}_{22} & 0 \\ 0 & 0 & 0 \end{matrix} \tag{13}$$

and

(14) $$\boldsymbol{\pi}^t = (0 \quad \boldsymbol{\pi}_2 \quad 0).$$

Should we wish to test the hypothesis that the groups differ primarily with respect to the two first order factors, we would specify:

(15) $$\boldsymbol{\Delta} = \begin{matrix} \boldsymbol{\Delta}_{11} & \boldsymbol{\Delta}_{12} & 0 \\ \boldsymbol{\Delta}_{12} & \boldsymbol{\Delta}_{22} & 0 \\ 0 & 0 & 0 \end{matrix}$$

and

(16) $$\boldsymbol{\pi}^t = (\boldsymbol{\pi}_1 \quad \boldsymbol{\pi}_2 \quad 0).$$

The 1st order aMGCFM coincides with the 1st order MGCFM in Equations 7 and 10, if all elements of the symmetric matrix $\boldsymbol{\Delta}$ and the vector $\boldsymbol{\pi}$ are estimated. In this case ($\boldsymbol{\Psi}_w\boldsymbol{\pi}$) equals $\boldsymbol{\delta}$, in Equation 9, and ($\boldsymbol{\Psi}_w\boldsymbol{\Delta}\boldsymbol{\Psi}_w^{\,t}$ + $\boldsymbol{\Psi}_w$) equals $\boldsymbol{\Psi}_b$ in Equation 10. This implies that the 1st order aMGCFM is nested under the 1st order MGCFM (Equations 7 to 10). Note that $\boldsymbol{\Delta}$ and $\boldsymbol{\pi}$ will contain zero elements, but the latent mean vector ($\boldsymbol{\Psi}_w\boldsymbol{\pi}$) and covariance matrix ($\boldsymbol{\Psi}_w\boldsymbol{\Delta}\boldsymbol{\Psi}_w^{\,t}$ + $\boldsymbol{\Psi}_w$) will not, because the covariance matrix $\boldsymbol{\Psi}_w$ will not (generally) contain zero elements.

In considering the constraints in the 2nd order MGCFM, we specify a single second order common factor, that is, $r = 1$. The model for the strong version of Spearman's hypothesis is as follows:

(17) $$\boldsymbol{\mu}_w = \boldsymbol{\nu}$$

(18) $$\boldsymbol{\Sigma}_w = \boldsymbol{\Lambda}(\boldsymbol{\Gamma}\boldsymbol{\Phi}_w\boldsymbol{\Gamma}^t + \boldsymbol{\Psi}^*)\boldsymbol{\Lambda}^t + \boldsymbol{\Theta}$$

(19) $$\boldsymbol{\mu}_b = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\tau}$$

(20) $$\boldsymbol{\Sigma}_b = \boldsymbol{\Lambda}(\boldsymbol{\Gamma}\boldsymbol{\Phi}_b\boldsymbol{\Gamma}^t + \boldsymbol{\Psi}^*)\boldsymbol{\Lambda}^t + \boldsymbol{\Theta}.$$

The matrix (1×1) $\boldsymbol{\tau}$ equals $\boldsymbol{\kappa}_b$ - $\boldsymbol{\kappa}_w$, the B-W difference in the mean of g. The (1×1) covariance matrix of the second order factor, g, is allowed to differ across the groups (i.e., $\boldsymbol{\Phi}_b \neq \boldsymbol{\Phi}_w$). The second order factor loadings in $\boldsymbol{\Gamma}$, like those in $\boldsymbol{\Lambda}$, are constrained to be equal over the groups. This constraint is a necessary condition for the equivalence of g in the two populations. A single element in $\boldsymbol{\Gamma}$ is fixed to equal 1, so that we may estimated the variances of g in the two groups, $\boldsymbol{\Phi}_b$ and $\boldsymbol{\Phi}_w$. The strong version of Spearman's hypothesis implies that the observed differences in mean are due only to latent difference in g. The weak version of Spearman's hypothesis is accommodated by introducing the parameter vector ($q \times 1$) $\boldsymbol{\delta}$. The model for the weak version is:

$$\boldsymbol{\mu}_w = \boldsymbol{\nu} \tag{21}$$

$$\boldsymbol{\Sigma}_w = \boldsymbol{\Lambda}(\boldsymbol{\Gamma}\boldsymbol{\Phi}_w\boldsymbol{\Gamma}^t + \boldsymbol{\Psi}_w{}^*)\boldsymbol{\Lambda}^t + \boldsymbol{\Theta}. \tag{22}$$

$$\boldsymbol{\mu}_b = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\delta} + \boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\tau} \tag{23}$$

$$\boldsymbol{\Sigma}_b = \boldsymbol{\Lambda}(\boldsymbol{\Gamma}\boldsymbol{\Phi}_b\boldsymbol{\Gamma}^t + \boldsymbol{\Psi}_b{}^*)\boldsymbol{\Lambda}^t + \boldsymbol{\Theta} \tag{24}$$

We have added the ($q \times 1$) vector $\boldsymbol{\delta}$, which accounts for latent differences in the means of the first order residuals (i.e., the components of the vector $\boldsymbol{\zeta}$). These differences are independent of those attributable to g. It is not possible to estimate all components of $\boldsymbol{\delta}$ besides the parameter in $\boldsymbol{\tau}$; specifically, q - 1 of the q components of $\boldsymbol{\delta}$ are identified. Note that the introduction of non-zero components in $\boldsymbol{\delta}$ is accompanied by possible group differences in the variance of relevant first order residuals, $\boldsymbol{\zeta}$, that is, $\boldsymbol{\Psi}_w{}^* \neq \boldsymbol{\Psi}_b{}^*$. We consistently specify that a difference in mean of a latent variable is accompanied by a difference in the variance of the latent variable.

Like the 1st order aMGCFM (Equations 7, 8, 11, 12), the 2nd order MGCFM's (Equations 17 to 20, and Equations 21 to 24) are nested under the 1st order MGCFM (Equations 7 to 10). The 2nd order MGCFM for the strong version of Spearman's hypothesis (Equations 17 to 20) is nested under the 2nd order MGCFM for the weak version (Equations 21 to 24). These nestings enable us to compare loglikelihood ratios in assessing competing models.

In comparing the factor structure in various groups using MGCFA, the equality constraint on the covariance matrices of the specific variance terms is sometimes dropped ($\boldsymbol{\Theta}_w \neq \boldsymbol{\Theta}_b$). One reason for this is that these terms contain both random measurement error and specific terms (Meredith, 1993). As groups may differ with respect to these specific terms, which are substantive latent variables like the common factors, Little (1997; see also Muthén and Lehman, 1985) advocates estimating separate $\boldsymbol{\Theta}$ matrices in each group in order to reduce the possible effects of unequal specific variance terms on the other parameters in the model. As demonstrated below, the hypothesis $\boldsymbol{\Theta}_w = \boldsymbol{\Theta}_b$ can be tested easily.

Factorial Invariance and Spearman's Hypothesis

The models presented above share the property that the differences between the groups in observed means and covariance structure are due to differences in means and covariance structure of the common factors. In these models, the factor loadings ($\mathbf{\Lambda}$ and, in the 2nd order factor model, $\mathbf{\Gamma}$), the specific variances ($\mathbf{\Theta}$), and intercepts ($\boldsymbol{\nu}$) are equal over the groups. Together these constraints comprise the hypothesis of *strict factorial invariance* (henceforth SFI; Meredith, 1993). Millsap (1997b) points out that, under SFI, "two individuals who have identical common factor scores would be expected to have identical observed scores regardless of their group membership" (p. 250). This implies that we are measuring the same constructs in both groups. Relaxation of any of the constraints mentioned would complicate this identical construct interpretation of the common factors in the model. The finding that factorial invariance is tenable suggests that the tests are unbiased with respect to group (Meredith, 1993; Millsap, 1997a, 1997b; Muthén & Lehman, 1985; Oort, 1996).

Millsap (1997a) has discussed the relationship between Jensen's test of Spearman's hypothesis and factorial invariance. He showed that the collinearity of differences in expected means and factor loadings, which forms the crux of Jensen's test, will hold if the groups differ mainly with respect to a single dominant first order common factor in the 1st order MGCFM. In this case the vectors of expected means equals $\boldsymbol{\nu}$ (whites) and $\boldsymbol{\nu} + \mathbf{\Lambda}\boldsymbol{\delta}$ (blacks), where $\boldsymbol{\delta}$ has a single non-zero component. This component equals the difference between the groups in the mean of the dominant factor. The factor loadings of the dominant factor (a column vector in $\mathbf{\Lambda}$) and the expected differences in means, $\mathbf{\Lambda}\boldsymbol{\delta}$, are necessarily collinear. The model discussed by Millsap is reminiscent of the model used by Gustafsson (1992) to analyze the Jensen and Reynold data (Jensen & Reynolds, 1982). This model, which is a special case of the 1st MGCFM, is called the *bifactor model* (Harman, 1976). As employed by Gustaffson (1992), this model includes 4 uncorrelated 1st order factors: 1 general factor representing *g*, and 3 group factors, representing specific cognitive abilities (memory, verbal and performance), unrelated to *g*. We do not consider the bifactor model in this article. We prefer to model *g* as a second order common factor, as this conceptualisation of *g* is more in line with current differential theory concerning general intelligence (Jensen & Weng, 1994; Jensen, 1998).

Finally, if the 1st order MGCFM comprises a single common factor, the collinearity between differences in means and factor loadings will necessarily hold. A problem with the single common factor model, however, is that it does not fit the covariance structure of multi-dimensional cognitive ability tests (e.g., the WISC). Likewise, if the 2nd order MGCFM holds, given the strict version of Spearman's hypothesis (i.e., mean differences restricted to the 2nd order factor), collinearity is observed between $\mathbf{\Lambda\Gamma}$ and the differences in means.

The 1st order aMGCFM is only applicable given multiple common factors. Generally, in the case of multiple common factors, which all contribute to the differences in means, there is no such simple collinearity relationship between differences in expected means and factor loadings.

MGCFA has several advantages over Jensen's test of Spearman's hypothesis. Using MGCFA to investigate the observed B-W differences, one can formulate and test a variety of competing hypotheses, of which Jensen's hypothesis concerning the importance of *g*, is but one. Decisions made on the basis of MGCFA concerning the nature of B-W differences are based on fit indices. The use of goodness of fit measures and the explicit comparison of competing models provide a more transparent justification than a Spearman correlation for the acceptance or rejection of the hypothesis that *g* is fundamental to B-W differences. In addition, Jensen's test focuses on a single aspect of factorial invariance, namely, the relationship between factor loadings and differences in means. Jensen does not investigate whether the factor loadings in the black and the white samples are equal using a test based on goodness of fit. Rather he does this in a separate procedure based on measures of factorial congruence. Factorial congruence, which concerns the hypothesis that $\mathbf{\Lambda}_b = \mathbf{\Lambda}_w$, is only one aspect of SFI (Meredith, 1993).

In summary, in the models that we consider, the observed B-W differences are attributed to differences with respect to the common factors. Strictly speaking, besides the 1st order MGCFM with a single common factor, it is only the second order MGCFM that incorporates the strong version of Spearman's hypothesis. Figure 1 provides an illustration of how the groups differ with respect to the common factors in the three models.

Illustration: Jensen and Reynolds (1982) Data

We analyze the data set published in Jensen and Reynolds (1982). The samples consist of 1868 whites and 305 blacks between the ages of 6 and 16.6 years. The samples were stratified on the basis of age, sex, race, SES, geographic region of residence, and the dichotomy urban-rural. The variables are the following 13 sub-scales of the WISC-R: Information (I), Similarities (S), Arithmetic (A), Vocabulary (V), Comprehension (C),

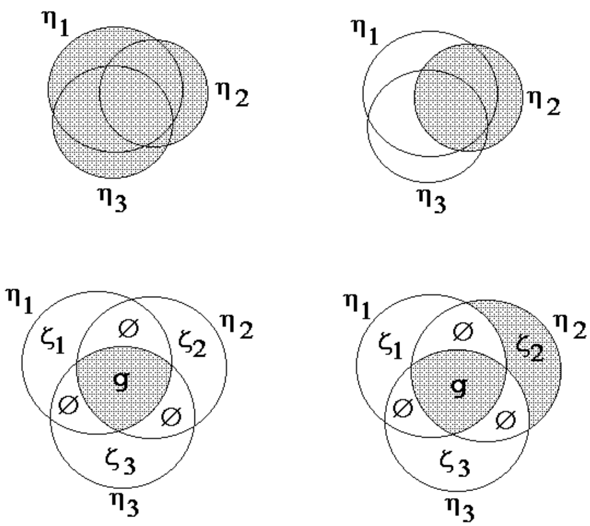

Figure 1
Diagrammatic Representation Model Group Differences with Respect to Three Common Factors 1, 2, and 3.

Grey areas denote sources of B-W differences in means and covariance structure. Top left: model A4; bottom left: strong version of Spearman's hypothesis (model B2; see Table 3); top right: model C1 (C2, C3); bottom right: weak version of Spearman's hypothesis (B3, B4, B5).

Digit Span (DS), Tapping Span (TS), Picture Completion (PC), Picture arrangement (PA), Block Design (BD), Object Assembly (OA), Coding (CO), and Mazes (MA). The effects of age on the test scores are removed by means of age-standardization. The reader is referred to Jensen and Reynolds (1982), and the references therein, for further details and summary statistics. For this data set, Jensen's test of Spearman's hypothesis yielded correlations of .75 and .64, using factor loadings estimated in the white and the black sample, respectively.

Exploratory Factor Analyses

To establish the pattern of the matrix of factor loadings ($\mathbf{\Lambda}$), we first carry out exploratory factor analyses. On the basis of the factor analysis published in Jensen and Reynolds (1982), and other analyses of the WISC-R (Gustafsson, 1992; Naglieri & Jensen, 1987), we expect 3 correlated 1st order factors: a verbal factor (VERB), a performance factor (PERF), and a memory span factor (MEM). We carry out exploratory factor analyses of the correlation matrices followed by an quartimax (oblique) rotation. These analyses are carried out using the program CEFA 1.01 (Browne, Cudeck, Tateneni, & Mels, 1998). In all analyses (here and below), we use normal maximum likelihood (ML) estimation to obtain estimates (Jöreskog, 1971; Sörbom, 1974; Lawley & Maxwell, 1971). Unlike many other programs for exploratory factor analysis, the CEFA program provides standard errors of parameter estimates, which can be used to evaluate the significance of the parameter estimates.

Before fitting the three factor models, we fit a single factor model. The $\chi^2(130)$ for this model equals 1544.3 [i.e., $\chi^2(65)$ = 279.2 in the black sample and $\chi^2(65)$ = 1265.09 in the white sample]. The $\chi^2(84)$ for the exploratory factor model with three common factors equals 195.5 [i.e., $\chi^2(42)$ = 112.3, in the white sample, and $\chi^2(42)$ = 83.2, in the black sample]. From these results it is clear that the single common factor model does not fit the data. We do not further consider the single factor model. Table 1 contains the factor loadings of the two groups for the three factor model. The underlined factor loadings are considered to be significantly greater than zero on the basis of the standard errors and the absolute magnitude. Naglieri and Jensen (1987) contains correlation matrices of 24 variables observed in a sample of 86 blacks and 86 whites. These 24 variables include 11 of the WISC-R sub-scales analyzed here (I, S, A, V, C, DS, PC, PA, BD, OA, CO). An exploratory factor analysis of these data confirmed the larger factor loadings in Table 1. The sample size of 86 within each group affords insufficient power to detect the smaller effects. On the basis of these results, we specify the matrix $\mathbf{\Lambda}$ as shown in Table 2.

The pattern of factor loadings in $\mathbf{\Lambda}$ is based on the union of the significant factor loadings in the white and the black sample (see Table 1). All models that we consider subsequently include this matrix of factor loadings.

Analysis Based on Multi-Group Confirmatory Factor Models[3]

All subsequent results are obtained from analyses of the covariance matrices and mean vectors. These analyses are carried out using LISREL 8.20 (Jöreskog & Sörbom, 1993). We fit the sequence of models summarized in Table 3. To assess goodness of fit we report a variety of fit indices, as recommended by Bollen and Long (1993). Although, there is some redundancy among these indices, we consider the non-normed fit index (NNFI), Akaike's information criterion (AIC), the related CAIC, the

[3] LISREL input files for all models (see Table 3) are available upon request.

Table 1
ML Estimates[a] of Quartimax Rotated Factor Loadings in Black (N = 305) and White (N = 1868) Sample (approximate standard errors in parentheses)

	blacks N = 305			whites N = 1868		
	VERB	PERF	MEM	VERB	PERF	MEM
I	.609	-.016	.215	.688	.036	.106
	(.074)	(.055)	(.073)	(.026)	(.023)	(.024)
S	.683	.119	-.026	.688	.109	-.007
	(.069)	(.061)	(.063)	(.026)	(.025)	(.023)
A	.363	.010	.402	.327	.056	.416
	(.087)	(.060)	(.086)	(.038)	(.026)	(.036)
V	.928	-.037	-.033	.874	-.054	.027
	(.034)	(.024)	(.031)	(.017)	(.013)	(.017)
C	.608	.138	.012	.718	.040	-.067
	(.072)	(.065)	(.069)	(.025)	(.025)	(.023)
DS	.064	.011	.700	.094	-.021	.620
	(.075)	(.041)	(.089)	(.035)	(.019)	(.037)
TS	-.088	.078	.598	-.075	.107	.545
	(.055)	(.065)	(.078)	(.024)	(.030)	(.033)
PC	.228	.471	.046	.205	.501	-.087
	(.079)	(.073)	(.075)	(.034)	(.031)	(.029)
PA	.136	.397	.104	.238	.387	-.022
	(.083)	(.076)	(.083)	(.035)	(.032)	(.032)
BD	.030	.657	.143	.003	.750	.124
	(.062)	(.073)	(.069)	(.021)	(.028)	(.022)
OA	-.019	.836	-.074	.000	.733	-.086
	(.033)	(.043)	(.032)	(.023)	(.023)	(.020)
CO	-.023	.409	.171	.066	.148	.326
	(.082)	(.078)	(.088)	(.037)	(.034)	(.036)
MA	.043	.204	.269	-.086	.482	.124
	(.088)	(.081)	(.091)	(.033)	(.032)	(.033)

Note. Underlined factor loadings are judged to be significant. The factors are identified as a verbal factor (VERB), a performance factor (PERF) and a memory factor (MEM).
[a] Estimates and standard errors obtained with the program CEFA 1.01 (Browne, et al., 1998).

Table 2
Pattern of Matrix **Λ** Used in MGCFA (based on results in Table 1)

Variable	Common Factors		
	VERB	PERF	MEM
I	$\boldsymbol{\lambda}_{1,1}$	0	$\boldsymbol{\lambda}_{1,3}$
S	$\boldsymbol{\lambda}_{2,1}$	$\boldsymbol{\lambda}_{2,2}$	0
A	$\boldsymbol{\lambda}_{3,1}$	0	$\boldsymbol{\lambda}_{3,3}$
V	$\boldsymbol{\lambda}_{4,1}{}^{*}$	0	0
C	$\boldsymbol{\lambda}_{5,1}$	$\boldsymbol{\lambda}_{5,2}$	0
DS	0	0	$\boldsymbol{\lambda}_{6,3}$
TS	0	0	$\boldsymbol{\lambda}_{7,3}{}^{*}$
PC	$\boldsymbol{\lambda}_{8,1}$	$\boldsymbol{\lambda}_{8,2}$	0
PA	$\boldsymbol{\lambda}_{9,1}$	$\boldsymbol{\lambda}_{9,2}$	0
BD	0	$\boldsymbol{\lambda}_{10,2}$	$\boldsymbol{\lambda}_{10,3}$
OA	0	$\boldsymbol{\lambda}_{11,2}{}^{*}$	0
CO	0	$\boldsymbol{\lambda}_{12,2}$	$\boldsymbol{\lambda}_{12,3}$
MA	0	$\boldsymbol{\lambda}_{13,2}$	$\boldsymbol{\lambda}_{13,3}$

Note: In fitting all models (see Table 3), the parameters accompanied by an asterisk are fixed to equal one. This parameterization allows us to estimate common factor variance-covariance matrices (rather than correlation matrices).

expected cross validation index (ECVI), the root mean square error of approximation (RMSEA), the χ^2, and χ^2/df. The χ^2 is treated as a measure of (badness of) fit, rather than as a formal test-statistic (Jöreskog, 1993). The index χ^2/df is a simple measure of fit which takes into account the degrees of freedom of the model. The RMSEA (Steiger, 1990; Browne & Cudeck, 1993) 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 an RMSEA of .05, or less, is indicative of a good approximation. The ECVI provides an indication of the discrepancy between the fitted covariance matrices in the analyzed samples and the expected covariance matrices that would be obtained in a second sample of the same size. The models with low values of ECVI are preferable to models with large values. This rule also applies to AIC and CAIC: lower values of AIC and CAIC indicate better fitting models. Compared to AIC, CAIC favors more parsimonious models. Finally, NNFI of .92, or higher,

Table 3
Summary of Models

Model:	Specification	Nesting[g]
1st order MGCFMs		
A1: pattern invariance	$\boldsymbol{\Lambda}_b \neq \boldsymbol{\Lambda}_w, \boldsymbol{\Psi}_b \neq \boldsymbol{\Psi}_w, \boldsymbol{\Theta}_b \neq \boldsymbol{\Theta}_w, \boldsymbol{\nu}_b \neq \boldsymbol{\nu}_w, \boldsymbol{\delta} = 0$	
A2: invariance $\boldsymbol{\Lambda}$	A1, but $\boldsymbol{\Lambda}_b = \boldsymbol{\Lambda}_w$	A1
A3: invariance $\boldsymbol{\Lambda}$, $\boldsymbol{\Theta}$	A1, but $\boldsymbol{\Lambda}_b = \boldsymbol{\Lambda}_w$ & $\boldsymbol{\Theta}_b = \boldsymbol{\Theta}_w$	A2
A4: invariance $\boldsymbol{\nu}$	A1, but $\boldsymbol{\Lambda}_b = \boldsymbol{\Lambda}_w$ & $\boldsymbol{\Theta}_b = \boldsymbol{\Theta}_w$ $\boldsymbol{\nu}_b = \boldsymbol{\nu}_w$, $\boldsymbol{\delta} \neq 0$	A3
2nd order MGCFMs		
B1: invariance $\boldsymbol{\Psi}$* and $\boldsymbol{\Gamma}$	$\boldsymbol{\Lambda}_b = \boldsymbol{\Lambda}_w, \boldsymbol{\Psi}_b^* = \boldsymbol{\Psi}_w^*, \boldsymbol{\Theta}_b = \boldsymbol{\Theta}_w, \boldsymbol{\Gamma}_b = \boldsymbol{\Gamma}_w,$ $\boldsymbol{\nu}_b \neq \boldsymbol{\nu}_w, \boldsymbol{\Phi}_b \neq \boldsymbol{\Phi}_w$ $\boldsymbol{\delta} = 0, \boldsymbol{\tau} = 0$	A3
B2[a]: Strong SH[f]	B1, but $\boldsymbol{\nu}_b = \boldsymbol{\nu}_w$, $\boldsymbol{\tau} \neq 0$	B1,B3 to B8
B3[b]: Weak SH I	B2, but VERB $\boldsymbol{\Psi}_b^* \neq \boldsymbol{\Psi}_w^*$, $\boldsymbol{\delta} \neq 0$	B1,B6,B8
B4[b]: Weak SH II	B2, but PERF $\boldsymbol{\Psi}_b^* \neq \boldsymbol{\Psi}_w^*$, $\boldsymbol{\delta} \neq 0$	B1,B6,B7
B5[b]: Weak SH III	B2, but MEM $\boldsymbol{\Psi}_b^* \neq \boldsymbol{\Psi}_w^*$, $\boldsymbol{\delta} \neq 0$	B1,B7,B8
B6[c]: Weak SH IV	B2, but PERF&VERB $\boldsymbol{\Psi}_b^* \neq \boldsymbol{\Psi}_w^*$, $\boldsymbol{\delta} \neq 0$	B1
B7[c]: Weak SH V	B2, but PERF&MEM $\boldsymbol{\Psi}_b^* \neq \boldsymbol{\Psi}_w^*$, $\boldsymbol{\delta} \neq 0$	B1
B8[c]: Weak SH VI	B2, but MEM&VERB $\boldsymbol{\Psi}_b^* \neq \boldsymbol{\Psi}_w^*$, $\boldsymbol{\delta} \neq 0$	B1
1st order aMGCFMs		
C1[d]: VERB	$\boldsymbol{\Lambda}_b = \boldsymbol{\Lambda}_w$ & $\boldsymbol{\Theta}_b = \boldsymbol{\Theta}_w$ VERB $\boldsymbol{\Delta} \neq 0$, $\boldsymbol{\pi} \neq 0$	A4,C5,C6
C2[d]: PERF	$\boldsymbol{\Lambda}_b = \boldsymbol{\Lambda}_w$ & $\boldsymbol{\Theta}_b = \boldsymbol{\Theta}_w$ PERF $\boldsymbol{\Delta} \neq 0$, $\boldsymbol{\pi} \neq 0$	A4,C4,C5
C3[d]: MEM	$\boldsymbol{\Lambda}_b = \boldsymbol{\Lambda}_w$ & $\boldsymbol{\Theta}_b = \boldsymbol{\Theta}_w$ MEM $\boldsymbol{\Delta} \neq 0$, $\boldsymbol{\pi} \neq 0$	A4,C4,C6
C4[e]: PERF&MEM	$\boldsymbol{\Lambda}_b = \boldsymbol{\Lambda}_w$ & $\boldsymbol{\Theta}_b = \boldsymbol{\Theta}_w$ PERF&MEM $\boldsymbol{\Delta} \neq 0$, $\boldsymbol{\pi} \neq 0$	A4
C5[e]: PERF&VERB	$\boldsymbol{\Lambda}_b = \boldsymbol{\Lambda}_w$ & $\boldsymbol{\Theta}_b = \boldsymbol{\Theta}_w$ PERF&VERB $\boldsymbol{\Delta} \neq 0$, $\boldsymbol{\pi} \neq 0$	A4
C6[e]: MEM&VERB	$\boldsymbol{\Lambda}_b = \boldsymbol{\Lambda}_w$ & $\boldsymbol{\Theta}_b = \boldsymbol{\Theta}_w$ MEM&VERB $\boldsymbol{\Delta} \neq 0$, $\boldsymbol{\pi} \neq 0$	A4

[a] Equations 17 to 20. [b] Single element of diagonal covariance matrices $\boldsymbol{\Psi}_b^*$ and $\boldsymbol{\Psi}_w^*$ free to vary over the groups; one component of $\boldsymbol{\delta}$ is estimated (see Equations 21 to 24). [c] Two elements of matrices $\boldsymbol{\Psi}_b^*$ and $\boldsymbol{\Psi}_w^*$ free to vary over the groups; two components of δ are estimated (see Equations 21 to 24). [d] Single diagonal element of Δ free; single component of $\boldsymbol{\pi}$ is estimated (see Equations 11, 12, and Equations 13 and14). [e] Three elements of Δ free (2 diagonals and 1 off-diagonal); two components of π are estimated (see Equations 11, 12, and 15, 16). [f] SH: Spearman's hypothesis. [g] For example, B5 is nested under B1, which is nested under A3, which is nested under A2, etc.

is viewed as indicative of a well fitting model. To judge the adequacy of selected models, we also report the range and median of the standardized residuals. The standardized residuals are a function of the difference between the observed data and the expected data under the specified model. If the model provides a good description of the data (strictly speaking, if the specified model is true, and the data are multivariate normally distributed), the standardized residuals follow a standard normal distribution.

The sequence of models that we consider is summarized in Table 3. The nesting among the models, which we discuss in some detail below, is also shown in Table 3. We first examine 1st order factor models. To establish whether the constraints associated with SFI are tenable, we first fit the model without the means, and without any equality constraints over the groups. The model is $\mathbf{\Sigma}_w = \mathbf{\Lambda}_w \mathbf{\Psi}_w \mathbf{\Lambda}_w^t + \mathbf{\Theta}_w$ and $\mathbf{\Sigma}_b = \mathbf{\Lambda}_b \mathbf{\Psi}_b \mathbf{\Lambda}_b^t + \mathbf{\Theta}_w$, where the pattern (the positions of fixed zeros and fixed ones) of $\mathbf{\Lambda}_w$ and $\mathbf{\Lambda}_b$ is identical, and $\mathbf{\Theta}_w$ and $\mathbf{\Theta}_b$ are diagonal. The χ^2 for this model (Model A1 in Table 3; pattern invariance) equals 239.9 (df = 106). We next introduce the constraints that $\mathbf{\Lambda}_w = \mathbf{\Lambda}_b$ and $\mathbf{\Theta}_w = \mathbf{\Theta}_b$. Following the imposition of the equality constraint that $\mathbf{\Lambda}_w = \mathbf{\Lambda}_b$, we observe a χ^2 of 281.2 (df = 125; Model A2), and, following the imposition of both $\mathbf{\Lambda}_w = \mathbf{\Lambda}_b$ and $\mathbf{\Theta}_w = \mathbf{\Theta}_b$, we observe a χ^2 of 305.7 (df = 138; Model A3). Finally we introduce the means as shown in Equations 7 and 9; this is model A4. The χ^2 of this model equals 327.7 (df = 148).

Table 4 contains a summary of the fit indices of the various models. On the basis of these, we conclude that model A4 provides an acceptable fit. The nesting among the models (A4 under A3, A3 under A2; A2 under A1) allows us to consider loglikelihood ratios. In comparing A4 to A3, A2, and to A1, we find that the increase in χ^2/df is fairly constant at about 2.1 (22.0/10, 46.5/23, and 87.8/42). CAIC clearly identifies A4 as the best model in the subset comprising A1 to A4. The AIC is slightly better for model A3, but the difference is very small (438.6 vs. 439.8). Neither RMSEA and NNFI discriminate well; they indicate that models A1 to A4 fit well. On the basis of the fit indices, we conclude that model A4 fits the data adequately and that therefore SFI is tenable. In the white sample the standardized residuals, which can be viewed as z-scores, range from -3.56 to 3.26 (median -.43); in the black sample, they range from -2.77 to 2.66 (median .07).

The next subset of models that we fit includes the second order common factor, g (Table 3, models B1 to B8). We first fit the model without the means. We retain the equality constraints on $\mathbf{\Lambda}$ and $\mathbf{\Theta}$, and add the constraints that $\mathbf{\Psi}_w^* = \mathbf{\Psi}_w^*$ and $\mathbf{\Gamma}_w = \mathbf{\Gamma}_b$ (see Equations 18 & 20; as mentioned, the asterisks indicate that the matrices are diagonal). The difference in covariance structure is solely a function of the variance of the

Table 4
Fit Indices for Model in Table 3

	χ^2	*df*	χ^2/df	CAIC	AIC	ECVI	RMSEA	NNFI	Model[a]
EFA	195.5	84	2.33						
1st order	239.9	106	2.26	1124.2	442.4	.204	.034	.979	A1
MGCFA	281.2	125	2.25	1001.7	446.9	.206	.034	.971	A2
	305.7	138	2.21	906.4	438.6	.202	.033	.968	A3
	327.7	148	2.21	840.8	439.8	.203	.033	.966	A4
2nd order	310.7	143	2.17	867.6	433.1	.200	.032	.967	B1
MGCFA	392.7	155	2.53	849.0	494.7	.228	.037	.959	B2
	379.9	153	2.48	850.6	483.0	.222	.036	.961	B3
	337.4	153	2.20	807.5	439.9	.203	.033	.965	B4
	367.3	153	2.40	836.6	468.9	.216	.035	.962	B5
	330.1	151	2.19	816.8	435.8	.200	.032	.966	B6
	332.4	151	2.20	819.3	438.3	.202	.032	.965	B7
	330.3	151	2.19	816.6	435.7	.201	.032	.965	B8
1st order	425.2	155	2.74	885.5	531.2	.240	.040	.960	C1
aMGCFA	363.7	155	2.34	810.1	455.8	.210	.034	.960	C2
	548.2	155	3.53	1020.3	666.1	.310	.049	.940	C3
	360.9	152	2.37	833.8	459.5	.210	.034	.960	C4
	334.7	152	2.20	813.1	438.8	.200	.033	.970	C5
	420.8	152	2.77	907.4	533.1	.250	.040	.960	C6

Note: See text for an explanation of these fit indices.
[a] See Table 3 for description of the models. Fit indices of models that are judged to fit relatively well are underlined.

second order common factor *g* (i.e., $\mathbf{\Phi}_w \neq \mathbf{\Phi}_b$). We refer to this model as model B1. The χ^2 for this model is 310.7 (df = 143). As model B1 and model A3 are nested, we can carry use the likelihood ratio to assess the tenability of the equality constraints. The difference in χ^2 equals 5.0 (df = 5), so we conclude that the constraints that $\mathbf{\Psi}_w^* = \mathbf{\Psi}_b^*$ and $\mathbf{\Gamma}_w = \mathbf{\Gamma}_b$ are reasonable. Next we introduce the mean structure into the model as shown in Equations 17 and 19. In model B2, the differences in means are modeled solely as a function of the difference in the mean of the second order factor, *g* (i.e., $\boldsymbol{\tau}$ in Equation 19). This model represents the strong version of

Spearman's hypothesis. The χ^2 for model B2, which is nested under model B1, equals 392.7 (df = 155).

Models B3 to B5 represent the weak versions of Spearman's hypothesis (see Equations 21 to 24). These models are nested under model B1. In addition to the difference in means of g, we introduce a difference in means of a single first order residual (i.e., 1 component of $\boldsymbol{\delta}$ in Equation 23). In model B3, B4, and B5, we allow a difference in means of the residual of the first order factors VERB, PERF, and MEM, respectively. As mentioned, we consistently let a difference in means be accompanied by a difference in variance. Compared to model B2, we therefore loose two df in fitting B3, B4, and B5. The observed χ^2 goodness of fit indices are 379.9 (df = 153) for model B3 (g + VERB), 337.4 (df = 153) for model B4 (g + PERF), and 367.3 (df = 153) for model B5 (g + MEM).

Other weak versions of Spearman's hypothesis may be investigated by including two first order residual factor means. These models, designated models B6 to B8 (see Table 3), were fit to make comparison with the other models possible. However, as models B6 to B8 are hardly more parsimonious than model A4 with respect to the parameterization of the means, they are judged to be of little use.

In the subset comprising models B2 to B5, CAIC clearly identifies model B4 (g + PERF) as the best model. The AIC of this model is as good as that of model A4, and does not differ greatly from the less parsimonious models B6 to B8. The same applies to the other fit indices, although again NNFI and RMSEA discriminate poorly. Models B2 to B5 are nested under model B1. The loglikelihood ratios equal 82.0 (B2 vs. B1, df = 12), 69.2 (B3 vs. B1, df = 10), 26.7 (B4 vs. B1, df = 10), and 56.6 (B5 vs. B1, df = 10). These comparisons also suggest that model B4 provides the best fit in the set comprising models B2 to B5. The standardized residuals for model B4 range between -3.44 and 2.81 (median -.36) in the white sample, and between -2.32 and 3.28 (median .48) in the black sample.

Finally we fit the subset of models based on the 1st order aMGCFM (Equations 7, 8, 11, 12). In the models C1, C2, and C3, the differences between the blacks and the whites in means and covariance structure are attributable solely to differences in the VERB, in the PERF, and in the MEM common factor, respectively. Judging by the fit indices, models C1 (VERB) and C3 (MEM) do not fit the data. The χ^2s of these models are 425.2 and 548.2 (df = 155), respectively. The hypothesis that the groups differ mainly with respect to the PERF factor (C2), provides a better fit: the χ^2 (df = 155) equals 363.7. Models C1 to C3 are nested under model A4. The loglikelihood ratios (df = 7) equal 97.5 (C1 vs. A4), 36.0 (C2 vs. A4), and 220.5 (C3 vs. A4). Models C4 to C6 represent less parsimonious 1st order aMGCFMs. Specifically, models C4, C5, and C6 specify that the groups differ with respect to factors PERF and MEM, the factors VERB and PERF, and the factors MEM and VERB, respectively. Model C6 (MEM & VERB) does not fit the data (χ^2 = 420.8, *df* = 152), and model C4 (PERF & MEM; χ^2 = 360.9, *df* = 152) does not provide any substantial improvement over model C2 (PERF; χ^2 = 363.7, *df* = 155). We identify model C5 as the best fitting model (VERB and PERF; χ^2 = 334.7, *df* = 152). Comparing the χ^2s of models C5 and A4, we observe an acceptable loglikelihood ratio of χ^2 = 7.0 (*df* = 4). The standardized residuals for model C5 range from -3.50 to 3.15 (median -.38) in the white sample, and from -2.45 to 3.11 (median .30) in the black sample.

On the basis of the results in Table 4, we conclude that models C1, C3, and C6 do not fit the data. Primary differences with respect to the VERB factor and (or) the MEM factor cannot account for the differences between the groups. Model B2, representing the strong version of Spearman's hypothesis does not fit the data well either. It is notable that model C2, in which the PERF factor is the only source of B-W differences, fits the data better than model B2 (χ^2 of 363.7 vs. 392.7, both models have 155 *df*). We identify models A4, B4, and C5 as providing the best description of the data. A preference for any one of these three is hard to justify. All three models are nested under model A3. The differences in χ^2 of these models and model A3, are 22.0 (A4; *df* = 10; χ^2/df = 2.20), 31.7 (B4, *df* = 15; χ^2/df = 2.11), and 29.0 (C5; *df* = 14; χ^2/df = 2.07). Given the large total sample size, these likelihood ratios are not judged to be particularly large.

Results of Fitting Models A4, B4 and C5

In the present section we consider the results of fitting models A4, B4, and C5 in more detail. Table 5 contains the standardized residuals of the means, and the modification indices (MIs) associated with the parameter vector $\boldsymbol{\nu}$ (i.e., the constant intercepts). The standardized residuals in the white group are reported. These equal the standardized residuals in the black group, but have opposite sign. The MI equals the expected decrease in χ^2 (*df* = 1) resulting from the removal of the constraint that a given component of $\boldsymbol{\nu}$ is equal over the groups (Sörbom, 1989). Removal of this constraint essentially removes the relevant variable from the model for the means.

In Table 5, we find that the MI's associated with the constant intercept of the variable *C* (comprehension) are by far the largest (10.8, 12.52, and 12.18, in models A4, B4, and C5, respectively). The other MI's do not merit concern, given the present sample size. It is notable that variable *C* is also associated with the largest discrepancy in the results of the exploratory factor

Table 5
Standardized Residuals[a] (SR) and Modification Indices[b] (MI) for Means in Models A4, B4, and C5

	Model A4		Model B4		Model C5	
	SR	MI	SR	MI	SR	MI
I	.15	.02	0.06	0.04	-0.04	0.04
S	-2.10	4.40	-1.67	3.13	-1.83	3.36
A	.93	.87	0.29	0.13	-0.31	0.13
V	-1.17	1.38	-.49	0.44	-0.73	0.53
C	3.30	10.8	3.59	12.52	3.49	12.18
DS	-1.62	2.64	-2.71	5.69	-2.48	5.36
TS	1.15	1.31	-0.16	0.27	-0.19	0.22
PC	-.65	.43	-0.47	0.25	-0.50	0.25
PA	.55	.31	0.72	0.49	0.69	0.48
BD	-1.74	3.03	-2.41	3.95	-2.29	3.60
OA	1.00	1.01	1.14	1.31	1.19	1.42
CO	.32	.11	0.54	0.01	-0.48	0.00
MA	1.50	2.24	1.15	1.92	1.18	1.98

[a] SR may be interpreted as a *z*-scores. [b] MI represents the expected drop in χ^2 if mean of the variables is dropped from the model.

analysis. Following rotation, the factor loading of *C* on the PERF factor is not significant in the white sample (.040, s.e.: .025), but it is significant in the black sample (.138, s.e.: .065). These results suggest that the comprehension sub-test, an indicator mainly of the VERB factor, may be biased with respect to group. The standardized residuals can be interpreted as *z*-scores. The mean (*sd*) of these equal .124 (1.15), -.032 (1.65), and -.177 (1.60), in models A4, B4, and C5, respectively. None of the standardized residuals are particularly large, except for the standardized residual of the mean of the variable *C*. The standardized residuals and MI's suggest that the three models fit the observed means well.

Table 6 contains the raw factor loadings and the squared multiple correlation (SMC) of each indicator for model A4. The SMC is the ratio of the variance explained by the common factors to the variance explained by the common factors plus the specific variance. The factor loadings and the

Table 6
Raw Factor Loadings (Standard Errors in Parentheses) and Square Multiple Correlations (SMC) in Model A4

	Factor Loadings			SMC
	VERB	PERF	MEM	
I	.80 (.03)	0*	.24 (.07)	.57
S	.82 (.03)	.16 (.03)	0*	.53
A	.36 (.03)	0*	.84 (.07)	.48
V	1*	0*	0*	.70
C	.75 (.03)	.11 (.03)	0*	.47
DS	0*	0*	1.29 (.07)	.45
TS	0*	0*	1*	.30
PC	.24 (.03)	.63 (.04)	0*	.36
PA	.29 (.03)	.52 (.04)	0*	.29
BD	0*	.89 (.04)	.43 (.05)	.63
OA	0*	1*	0*	.50
CO	0*	.25 (.04)	.77 (.07)	.22
MA	0*	.56 (.04)	.32 (.06)	.23

Note: Parameters accompanied by an asterisk are fixed values. Factor loadings and SMC's in models B4 and C5 do not differ appreciably from those of model A4. SMC is the squared multiple correlation, that is, the percentage of variance of each test attributable to the common factors.

SMC's display little variation over the models A4, B4, and C5. Within the models, the SMC's are quite variable, ranging from 22% (CO) to 70% (V). The SMC's of the variables CO (22%), PC (36%), PA (29%), TS (30%), and MA (23%) are low. Low SMC's are not necessarily a source of concern in standard factor analysis. Here they are, because we have assumed that the specific variance terms, components of the vector $\boldsymbol{\varepsilon}$, do not contribute significantly to the means. If $\boldsymbol{\varepsilon}$ represents pure error, this is a safe assumption. However, usually in covariance structure modeling $\boldsymbol{\varepsilon}$ consists both of error and a systematic specific term. A SMC of 22% implies that 78% of the variance is not accounted for by the common factor(s). Given the nature of the present tests, it is unlikely that this 78% of the variance will be due to pure error. If a substantial part of the variance is due to a specific, systematic factor, it is possible that this factor is contributing to the means. This contribution may differ in the black and the white population. One would require repeated measures to identify the possible contribution to the mean of such specific effects (Meredith, 1991; Sörbom, 1975).

Table 7 contains the differences in common factor means. In model A4, these are estimated directly. In models B4 and C5, these differences are derived from the parameter estimates (see column 5 in Table 7). For instance, in models B4, the difference in means of common factor MEM is due to the regression of MEM on the second order factor *g*. The source of this difference in the mean of MEM is *g*. In model C5, the difference is due to the fact that MEM is correlated with VERB and PERF. As expected now, the differences in factor means are similar in the three models. At the level of the common factors, the VERB and the PERF factor make comparable contributions to the B-W differences. The MEM factor makes the smallest contribution to the B-W differences.

Model B4 is compatible with the weak version of Spearman's hypothesis. However, the weak version of Spearman's hypothesis states that the differences in means are *mainly* a function of the second order factor *g*, and, to a lesser extent, a function of the first order factors. Table 8 contains the decomposition of the vector of differences in means in models A4, B4, and C5. Limiting attention to model B4, we find that the percentage of the B-W differences in the means of the 13 variables attributable to *g* equals: 100 (i), 92 (s), 100 (a), 100 (v), 93 (c), 100 (ds), 100 (ts), 62 (pc), 67 (pa), 56 (bd), 49 (oa), 76 (co), 58 (ma). The percentage for the variable ma, for example, is calculated as 1.07/(1.07+0.78) (see Table 8, column 6 and 7). It would appear that in five of the variables (pc, pa, bd, oa, ma), the weak version of Spearman's hypothesis is questionable, even though the prescribed model

Table 7
Contributions of First Order Factors to B-W Differences in Latent Means

Model	VERB	PERF	MEM	model
A4	-2.63	-2.75	-0.82	$\boldsymbol{\delta}$[a]
B4	-2.59	-2.74	-0.99	$\boldsymbol{\delta} + \boldsymbol{\Gamma\tau}$[b]
C5	-2.61	-2.73	-1.00	$\boldsymbol{\Psi}_w\boldsymbol{\pi}$[c]

[a] See Equation 9. The vector $\boldsymbol{\delta}$ has three non-zero components, as all three first order factors contribute to the differences in means. [b] See Equation 23. The vector $\boldsymbol{\delta}$ has one non-zero component, that is, the difference in mean of the first order residual of PERF. See footnotes of Table 8 for actual values of $\boldsymbol{\delta}$ and $\boldsymbol{\tau}$. [c] See Equation 11 (and Equation 16). The vector $\boldsymbol{\pi}$ has two non-zero components, that is, difference in means of PERF and VERB. See footnotes of Table 8 for actual values of $\boldsymbol{\pi}$.

Table 8
Decomposition of Mean Differences in Models A4, B4, and C5

	obs.[a]	Model A4[b]			Model B4[c]		Model C5[d]	
		VERB	PERF	MEM	*g*	PERF	VERB	PERF
i	-2.32	-2.11	0	-.19	-2.31	0	-1.15	-1.18
s	-2.38	-2.15	-.44	0	-2.33	-.22	-1.15	-1.40
a	-1.74	-.96	0	-.69	-1.77	0	-.85	-.92
v	-2.56	-2.63	0	0	-2.59	0	-1.29	-1.31
c	-2.61	-1.97	-.30	0	-2.09	-.15	-1.04	-1.22
ds	-.90	0	0	-1.06	-1.28	0	-.59	-.70
ts	-.97	0	0	-.82	-.99	0	-.46	-.54
pc	-2.29	-.62	-1.74	0	-1.47	-.88	-.69	-1.65
pa	-2.27	-.76	-1.43	0	-1.45	-.72	-.69	-1.48
bd	-2.69	0	-2.46	-.35	-1.63	-1.25	-.74	-2.13
oa	-2.84	0	-2.75	0	-1.34	-1.39	-.61	-2.12
co	-1.36	0	-.67	-.64	-1.10	-.34	-.50	-.94
ma	-2.02	0	-1.54	-.26	-1.07	-.78	-.49	-1.36

[a] Observed B-W differences in means. [b] Decomposition is obtained in calculating $\mathbf{\Lambda\delta}$ (see Equation 9). The matrix $\mathbf{\Lambda}$ is given in Table 6, the vector $\mathbf{\delta}$ in Table 7. [c] Decomposition is obtained in calculating $\mathbf{\Lambda\delta} + \mathbf{\Lambda\Gamma\tau}$ (see Equation 23). The matrix $\mathbf{\Gamma}^t$ equals (1 .519, .384), the vectors $\mathbf{\delta}^t$ and $\mathbf{\tau}^t$ equal (0, -1.393, 0) and (-2.593), respectively. [d] Decomposition if obtained in calculating $\mathbf{\Lambda\Psi}_w\mathbf{\pi}$ (see Equation 11). The matrix $\mathbf{\Psi}_w$ is given in Table 9, the vector $\mathbf{\pi}^t$ equals (-.206, -.441, 0).

fits acceptably. This judgement is necessarily subjective, because the weak version of Spearman's hypothesis is ill-defined.

Table 9, finally, displays the covariance matrices of the common factors. It is striking that in model B4 the VERB factor correlates highly with the *g* factor (.96 in the white sample and .95 in the black sample). This large correlation was also observed in similar analysis of the Naglieri and Jensen (1987) data set.

Discussion

The aims of the present article are (a) to present suitable multi-group confirmatory factor models to investigate Spearman's hypothesis; (b) to discuss the relationship between factorial invariance and Spearman's

Table 9
Covariance Matrices of Common Factors (correlations shown in parentheses)

Model A4

whites[a] VERB	PERF	MEM	blacks[a] VERB	PERF	MEM
6.29			5.30		
2.98	4.79		2.70	4.43	
(.54)			(.56)		
2.22	1.25	2.41	2.32	1.54	2.53
(.57)	(.36)		(.63)	(.46)	

model B4

whites[b] VERB	PERF	MEM	*g*	blacks[b] VERB	PERF	MEM	*g*
6.26				5.57			
3.00	4.85			2.65	4.38		
(.54)				(.53)			
2.22	1.15	2.41		1.96	1.02	2.30	
(.57)	(.34)			(.45)	(.26)		
5.78	3.00	2.22	5.79	5.11	2.65	1.96	5.11
(.96)	(.56)	(.59)		(.95)	(.56)	(.47)	

model C5

whites[c] VERB	PERF	MEM	blacks[c] VERB	PERF	MEM
6.30			5.43		
2.97	4.81		2.80	4.54	
(.54)			(.56)		
2.22	1.23	2.40	1.93	1.16	2.30
(.57)	(.36)		(.54)	(.36)	

[a] $\mathbf{\Psi}_w$ in Equation 18, and $\mathbf{\Psi}_b$ in Equation 20. [b] $\mathbf{\Gamma\Phi}_w\mathbf{\Gamma}^t + \mathbf{\Psi}_w$ in Equation 22, and $\mathbf{\Gamma\Phi}_b\mathbf{\Gamma}^t + \mathbf{\Psi}_b$ in Equation 24. [c] $\mathbf{\Psi}_w$ in Equation 18, and $(\mathbf{\Psi}_w\mathbf{\Delta\Psi}_w{}^t + \mathbf{\Psi}_w)$ in Equation 12.

hypothesis within the context of these models and the advantages of MGCFA over Jensen's test of Spearman's hypothesis; and (c) to apply these models to a real data set. We have emphasized that the crux of Jensen's test of Spearman's hypothesis, namely the relationship between differences in means and factor loadings, is a single aspect of the hypothesis of SFI (see also Millsap, 1997b). Using MGCFA to investigate Spearman's hypothesis has a number of clear advantages compared to Jensen's test. First, MGCFA allows one to investigate Spearman's hypothesis in an integrative, more elegant manner: the various steps involved in Jensen's procedure are included in a single model. The hypothesis of SFI (Meredith, 1993), which we view as crucial to the meaningful interpretation of group differences within the factor model, is tested explicitly in MGCFA. In Jensen's test, this hypothesis is not tested directly, rather, some (but not all) aspects of SFI are investigated piecemeal in the prescribed procedure of Jensen's test (Jensen, 1992). Second, MGCFA allows one to fit a variety of competing models, which may, or may not include *g*. To prove unambiguously that *g* plays a central role in black-white differences in cognitive test scores, one has to demonstrate that the model, in which *g* is accorded a central role, fits acceptably, and that it fits better than competing models. A considerable advantage of MGCFA is its flexibility. MGCFA may be extended to several groups. For instance, one can include sub-groups of blacks and whites (e.g., urban and rural, high and low SES), as well as other groups, such as Asian-Americans. In addition, models for repeatedly measured groups can tested (e.g., Jöreskog, 1979; Marsh & Grayson, 1994, Hanna & Lei, 1985), which allows one to investigate black-white (say) differences in cognitive development.

Fitting the models presented (see Table 4) produced two important findings. First, SFI was found to be tenable. Second, it proved hard to distinguish between models. Regardless of the exact choice of model, the finding that SFI is tenable has important implications. As mentioned, this finding suggests that the tests are unbiased with respect to group (Millsap, 1997a, 1997b; Mellenbergh, 1989; Meredith, 1993). Another, closely related implication of strict factor invariance pertains to sources of within and between group differences. It has been emphasized many times that the causes of individual differences within two groups do not necessarily have any bearing on the causes of differences between the groups (e.g., Furby, 1973; Rose, Kamin, & Lewontin, 1984). However, if SFI is found to be tenable, we propose that this does provide support the hypothesis that within and between group differences are attributable to common causes. If the sources of between group differences are different from the sources of within group differences, we believe that it is unlikely that the vector of differences in means could be successfully modeled within the MGCFM (for a related argument, see Turkheimer, 1991).

The difficulty in distinguishing between models is disappointing. It is notable that RMSEA and NNFI fail almost completely to distinguish between models. On basis of the other fit indices in Table 4, we can reject models B2 (strong version of Spearman's hypothesis), C1, C3, and C6. However, models A4, B4, B6, B7, B8, C2, and C5 all fit the data comparatively well. The failure to distinguish between these models is due to the essentially small differences between the models. In addition, unequal sample sizes are known to adversely affect power in tests of factorial invariance (Kaplan & George, 1995). Finally, the failure to distinguish between models is also attributable to low dimensionality of the common factor space. To better distinguish between the models one requires a greater variety and a greater number of indicators of the common factors, and a large number of subjects that is more equally distributed over the two groups. Although the failure to discriminate between models is disappointing, it is an advantage of MGCFA that we are at least clearly confronted with this inability. This encourages us to be reluctant to embrace a given model too readily.

Jensen (1992) has written of Spearman's hypothesis that "the hypothesis, if true, would mean that understanding the nature of the statistical B-W differences on various psychometric tests in the cognitive domains depends fundamentally on the nature of *g* itself" (p. 229; see also Jensen, 1998, chapter 11). Establishing this would indeed be very informative. However, on the basis of the present results, we contend that it is questionable whether Jensen's test can be trusted to demonstrate the hypothesized central role of *g* in B-W differences. The interpretation of the correlations of .75 and .64, as observed in the present data, in support of the role of *g* in B-W differences is problematic. Having fitting the sequence of models in Table 3, we know that quite a number of models, both including and excluding *g*, fit well. Although a number of these models are compatible with the weak version of Spearman's hypothesis (notably model B4), the correlations of .75 and .65 would appear to be possible with models that do not include *g* (model C5, A4). Schönemann (1997b) has raised a similar point within the context of principal component analysis (PCA). Using simulated data, he demonstrated that in the absence of *g*, Jensen's test still produced substantial Spearman correlations. Within the context of MGCFA, which we consider to be a more appropriate approach (than PCA), Lubke, Dolan and Kelderman (1999), likewise, demonstrated that large Spearman correlations may be observed even though the underlying hypothesis, namely the central role of *g* as a source of within and between group differences, is severely violated. In conclusion, we believe that there are good reasons to abandon Jensen's test in favor of MGCFA in investigating group differences in cognitive ability test scores.

References

Bentler, P. (1992). *EQS: Structural equations program manual*. Los Angeles, CA: BMPD Statistical Software.

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

Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In: K.A. Bollen and J. Scott Long (Eds.). *Testing Structural Equation Models*. Newbury Park: Sage Publications.

Browne, M. W., Cudeck, R., Tateneni, K., & Mels, G. (1998). *CEFA: Comprehensive Exploratory Factor Analysis*. [WWW document and computer program]. URL http://quantrm2.psy.ohio-state.edu /browne/.

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

Dolan, C. V. (1997). A note on Schönemann's refutation of Spearman's hypothesis. *Multivariate Behavioral Research, 32*, 319-325.

Dolan, C. V. & Molenaar, P. C. M. (1994). Testing specific hypotheses concerning latent group differences in multi-group covariance structure analysis with structured means. *Multivariate Behavioral Research, 29*, 203-222.

Furby, L. (1973). Implications of within-group heritabilities for sources of between group differences: IQ and racial differences. *Developmental Psychology, 9*, 28-37.

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* (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.

Hanna, G. & Lei, H. (1985). A longitudinal analysis using the LISREL model with structured means. *Journal of Educational Statistics, 10*, 161-169.

Harman, H. H. (1976). *Modern factor analysis* (3rd edition). Chicago: University of Chicago Press.

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

Jensen, A. R. (1985). The nature of the black-white difference on various psychometric tests: Spearman's hypothesis. *Brain and Behavioral Sciences, 8*, 193-263.

Jensen, A. R. (1992). Spearman's hypothesis: Methodology and evidence. *Multivariate Behavioral Research, 27*(2), 225-233.

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

Jensen, A. R. & Reynolds, C. R. (1982). Race, Social Class and Ability Patterns on the WISC-R. *Personality and Individual Differences, 3*, 423-438.

Jensen, A.R. & Weng, L.-J. (1994). What is good *g*? *Intelligence, 18*, 231-258.

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

Jöreskog, K. G. (1979). Statistical estimation of structural models in longitudinal developmental investigations. In: J.R. Nesselroade and P.B. Baltes (Eds), *Longitudinal research in the study of behavior and development* (pp. 303-352). New York: Academic.

Jöreskog, K. G. (1993). Testing Structural Equation Models. In: K.A. Bollen and J. Scott Long (Eds.). *Testing Structural Equation Models.* Newbury Park: Sage Publications.

Jöreskog, K. G. & Sörbom, D. (1993). *LISREL 8: Structural Equation Modeling with the SIMPLIS command language*. Chicago: Scientific Software International.

Kaplan, D. & George, R. (1995). A study of the power associated with testing factor mean differences under violations of factorial invariance. *Structural Equation Modeling, 2,* 101-118.

Lawley, D. N. & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. London: Butterworth.

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

Lubke, G. L., Dolan, C. V. and Kelderman, H. (1999). Investigating black-white differences on cognitive tests using Spearman's hypothesis: an evaluation of Jensen's method. *Submitted for publication*.

Lynn, R. & Owen, K. (1994). Spearman's hypothesis and test scores differences between White, Indian, and Blacks in South Africa. *The Journal of General Psychology, 121,* 27-36.

Marsh, H. W. & Grayson, D. (1990). Public/Catholic differences in the high school and beyond data: A multi-group structural equation modelling approach to testing mean differences. *Journal of Educational Statistics, 5,* 199-235.

Marsh, H. W. & Grayson, D. (1994). Longitudinal stability of latent means and individual differences: A unified approach. *Structural Equation Modeling, 1,* 317-359.

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

Meredith, W. (1991). Latent variable models for studying differences and change. In L.M. Collins and J.L. Horn (Eds.), *Best methods for the analysis of change. Recent advances, unanswered questions and future directions*. Washington, DC: American Psychological Association.

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

Millsap, R. E. (1997a). Invariance in measurement and prediction: Their relationship in the single factor case. *Psychological Methods, 2,* 248-260.

Millsap, R. E. (1997b). 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.

Muthén, B. & Lehman, J. (1985). Multiple group IRT modeling application to item bias analysis. *Journal of Educational Statistics, 10,* 133-142.

Naglieri, J. A. & Jensen, A. R. (1987). Comparison and black-white differences on the WISC-R and the K-ABC: Spearman's Hypothesis. *Intelligence, 11,* 21-43.

Neale, M. C. (1997). *Mx: Statistical Modeling*. Richmond: Medical College of Virginia.

Oort, F. J. (1996). *Using restricted factor analysis in test construction*. Ph.D. Thesis, Psychology Faculty, University of Amsterdam.

Rock, D. A., Werts, C. E., & Flaughter, 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.

Rose, S., Kamin, L. J., & Lewontin, R. C. (1984). *Not in our genes: Biology, Ideology and Human Nature*. Harmondsworth: Penguin.

Rushton, J. P. (1999). Secular gains is IQ not related to the *g* factor and inbreeding depression — unlike Black-White differences: a reply to Flynn. *Personality and Individual Differences, 26*, 381-389.

Saris, W. E., de Pijper & Mulder, J. (1978). Optimal procedures for estimating factor scores. *Sociological Methods and Research, 7*, 85-106.

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

Schönemann, P. H. (1997a). Famous artefacts: Spearman's hypothesis. *Cahiers de Psychologie Cognitive (Current Psychology of Cognition), 16*, 665-694.

Schönemann, P. H. (1997b). The rise and fall of Spearman's hypothesis. *Cahiers de Psychologie Cognitive (Current Psychology of Cognition), 16*, 788-812.

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.

Sörbom, D. (1975). Detection of correlated errors in longitudinal data. *British Journal of Mathematical and Statistical Psychology, 27*, 229-239.

Sörbom, D. (1989). Model modification. *Psychometrika, 54*, 371-384.

Steiger, J. H. (1990). Structural model evaluation and modification: and interval estimation approach. *Multivariate Behavioral Research, 25*, 173-180.

te Nijenhuis, J. & van der Flier, H. (1997). Comparability of the GATB scores for immigrants and majority group members: Some Dutch findings. *Journal of Applied Psychology, 82*, 675-687.

Turkheimer, E. (1991). Individual and group differences in adoption studies of IQ. *Psychological Bulletin, 110*, 392-405.

Accepted May, 1999.