CPC, 1997, 16 (6), 721-728

COMMENTARY ON « FAMOUS ARTEFACTS »
(P. H. SCHÖNEMANN)

On the mathematical relationship between factor or component coefficients and differences between means

John Horn

University of Southern California

By my reckoning, Schönemann is basically correct, although perhaps not in every detail. If n variables are positively correlated and the means on these variables (M_1, M_2, ... M_n) for one group of subjects are larger than the means for another group (m_1, m_2, ..., m_n), then the correlation between a 'd' vector of differences between the means (M_1-m_1, M_2-m_2, ..., M_n-m_n) and a 'PC1' vector of the first principal component coefficients (C_1, C_2, ..., C_n) for those variables (whether calculated separately within the two groups or over the sample of the two groups combined) is mathematically constrained to be positive, nonzero. This is not so much an artefact (or, as more commonly spelled, artifact) as it is an algebraic necessity. Guttman (1992) said this in words and in mathematical demonstration in his 'last paper' (Mulaik, 1992). Schönemann cites that work, develops new demonstrations of the artefact, and points out that Jensen (1992), in his analyses of eleven large data sets, and several scientists who have accepted and praised Jensen's analyses, seem

---

Correspondence should be sent to John Horn, Psychology Department, University of Southern California, Los Angeles, CA 90089-1061, U.S.A. (e-mail: jhorn@rcf.usc.edu).

not to have recognized the algebraic constraint, and instead have incorrectly regarded the correlation between the d vector and the PC1 vector as entirely an 'empirical discovery' (Jensen, 1992, p. 232). I think the point of Guttman's and Schönemann's demonstrations – the algebraic restraint – can be made a bit more simply than they make it, and the point can almost certainly be made less caustically and sarcastically than Schönemann makes it, but I think he is quite correct in making the point and in lecturing scientists to be more careful and cautious in their analyses and writings, particularly when the results can be (or can be readily interpreted as) denigrating to a whole class of people. It is neither good science nor good ethics to proclaim that an outcome that is a result of mathematical constraint is entirely empirical evidence in support an hypothesis stipulating that Black persons, as a population, are inferior in respect to a highly valued attribute, intelligence.

The constraint occurs under several different conditions, and the magnitude of the constraint and hence of the correlation between the component weights and the mean differences depends on which of these conditions obtains. At one extreme are the conditions of factorial invariance across two groups. My 'simpler' way of making Schönemann's (and Guttman's) point stems from the case in which the conditions of factorial invariance obtain. These are the conditions Jensen assumes implicitly in his claim that the same *g* appears in different batteries of tests, and in his claims that groups differ in respect to this *g*, but he does not acknowledge these assumptions explicitly. These are the conditions Guttman assumed (explicitly) in his demonstration of the constraint for a case in which there is one and only one common factor – the requirements of the Spearman model for *g*. These are the conditions Schönemann assumed here (also explicitly) in his demonstration of the constraint for the case of the first principal component in what he identified as a Level I interpretation. Both Guttman and Schönemann explicitly assumed that the covariance matrices for two groups are equal.

The condition of equal covariance matrices for different groupings of people is a condition of strong factor invariance (Meredith, 1993). This means that, in common-factor parlance, the factor pattern matrix, the factor variances, the factor intercorrelations, and the variable communalities are the same – are invariant. For the case of principal components, the full component pattern matrix (including the PC1 coefficients, but also the PC2, PC3, ..., PCn coefficients) and component variances are invariant. The factor means – component means – and variable means may differ under these conditions, and, of course, do differ for the case under consideration. A single common factor, such as a *g* factor, is not required under these conditions, but the first common factor coefficients (whether or not the Spearman *g*-model obtains) will be invariant across the two samples. Most important the coefficients for estimating the variables from the common factors or principal components are the same for the two groups (as shown in some detail in Horn, 1991).

For the principal components case, any variable, Z_j (for i = 1, 2, ..., N_1 individuals in group 1 and i = 1, 2, ..., N_2 individuals in group 2) can be calculated as a weighted linear composite of the principal components as

$$Z_j = a_{j1} C_{1i} + a_{j2} C_{2i} + \dots + a_{jn} C_{ni} \quad [1]$$

The same linear composite, using the same a_{jk} weights (k = 1, 2, ..., n for n variables), is calculated in both groups under the factor invariance conditions: all that varies across groups are the component scores, C_{ki}. (These scores can be calculated as linear combinations of the variables, the weights being the values in the inverse of the matrix of the a_{jk}, again the same for the two groups). The mean for Z_j is the sum of each a_{jk} times the corresponding mean of C_{ki} (k = 1, 2, ..., n). Thus, if each of the means for the C_{ki} is larger for the i = 1, 2, ..., N_1 individuals in group 1 than for the i = 1, 2, ..., N_2 in group 2, as is required under the conditions specified by Jensen and Schönemann, then, for each instance of j variables (j = 1, 2, ..., m), necessarily the mean of Z_j for the i = 1, 2, ..., N_1 individuals in group 1 is larger than the mean of Z_j for the i = 1, 2, ..., N_2 individuals in group 2. More pointedly, the largeness of the means is a direct function of the a_{j1} coefficients for PC1 and in the order of magnitude of these a_{j1} coefficients.

This is seen most simply and clearly in the extreme case of a non-error *g*-factor and true specific factors, in which case essentially only the first principal component is a non-error set of scores. In this case, equation [1] reduces to essentially only

$$Z_j = a_{j1} C_{1i} + e, \quad [2]$$

where e is random error: because the components beyond the first, C_{ki} (k = 2, ..., n), are random variables, the weighted sum of these ($a_{j2} C_{2i}$ +...+ $a_{jn} C_{ni}$) is e. This e can be set to an expected value of zero by, for example, assuming the C_{ki}'s are in standard-score form or by scaling the C_{ki}'s to standard-score form: the negative C_{ki}'s then cancel the positive C_{ki}'s in the summation. In any case, the mean of Z_j is a_{j1} times the mean of C_{1i} (give or take some error). So if the mean of C_{1i} is larger for the group 1 sample than for the group 2 sample, then necessarily the mean for each variable, Z_j, for group 1 is larger than the mean of that variable in group 2 and, most important, the variable means are larger in precisely the amount, and thus in the order of, the a_{j1} coefficients.

Concretely, if the component means for groups 1 and 2 are, say, 4 and 3, respectively, and the coefficients for the first principal component are .70, .60, .50, .40, .30, .20 for variables Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, respectively, then the variable means for the first group are 2.8, 2.4, 2.0, 1.6, 1.2, and 0.8, while for the second group the means are 2.1, 1.8, 1.5, 1.2, 0.9, 0.6, and the differences between the means are 0.7, 0.6, 0.5, 0.4, 0.3, and 0.2, which are in precisely the order the principal component coefficients. This will be true for any difference in the component means for the two groups, however small, and for any range of values of the component regression coefficients.

Thus we see in this case that there is a perfect algebraic relationship between the mean difference vector and the PC1 regression coefficients. The relationship holds for both the Level I and Level II interpretations of Schönemann's article (since invariance across samples means that the factor pattern matrix for the combined sample is the same as for the separate samples). The only statistical assumption on which this demonstration of the relationship is based is the assumption that the specific factors – components beyond the first – are random variables. It is easy to see the relationship. It requires no complex mathematical reasoning. One would think that Jensen would readily see it and acknowledge it. And given that he saw it, he presumably would not then call Guttman's critique 'peevish' and resort to a specious argument that « Ipso facto, nothing can be mathematically inferred about the rank order of tests' means (or mean group differences) from a knowledge that the tests' loadings on *g* or on any other factors extracted from the correlation matrix. » (Jensen, 1992, p. 232). Unfortunately, Guttman's and Schönemann's demonstrations of this relationship are a good deal more complex – as well as more mathematically and statistically elegant – than this simple demonstration. A consequence of this, it seems, is that Jensen does not 'get it' and thus continues to proclaim that since his « Spearman's hypothesis has been consistently borne out on many independent sets of appropriate data , and no contrary data have been found, it may legitimately claim the status of empirical fact » (Jensen, 1992, p. 232). This is unfortunate.

This demonstration of the dependent relationship between the vectors of means and the vector of factor or component coefficients is essentially only a restatement of the Spearman *g*-factor model stipulating one and only one common factor: it follows from the statements of that model. If the one-and-only-one common factor hypothesis is supported by data, then the differences between means will be proportional to the factor or component coefficients.

Jensen (1992) referred to the Spearman *g*-factor model as a « long defunct 'two-factor' theory of mental abilities ») and added that he « ... never used Spearman's single factor method in testing what [he] termed the Spearman's hypothesis » (p. 226). This is unfortunate, for had he worked with Spearman's model, he very possibly would have noticed the simple relationship shown here.

This mathematical relationship does not prove anything empirically, no matter how many times it is calculated on independent sets of data. The relationship will always be found. It will always be nearly perfect (except for the random e variance) if the Spearman *g*-factor model fits the data.

The problem is the *g*-factor model never fits cognitive ability data if the sample of abilities selected to test the model are well chosen to represent the diversity of what is referred to as human intelligence. In such samples of abilities, it has always been found that more than one common factor is required to account for the covariance among the tests. (Narrow samples of abilities carefully selected to indicate particular common factors can be shown to fit the model, as in Horn (1997), but when other abilities are added into these samples, the model no longer fits the data).

It is found, also, that in different batteries of tests, broad and narrow, the first linear composite, whether weighted as a first principal component or not, correlates with the first linear composite of other batteries of tests at varying levels from a low of close to zero to a high of close to the reliabilities of the composites. The same common factor – the same *g* – simply does not show up in different batteries of tests all of which are said to indicate important features of human intelligence.

This is true of the general factors calculated at higher orders by Schmid-Leiman (1957) transformation, as in Carroll's (1993) tour de force study of hundreds of sets of data. Carroll identified a general factor by Schmid-Leiman transformation in 33 separate analyses, but the factor of one analysis was not the same as the factor of other analyses. Some of these general factors involved crystallized knowledge and retrieval from long-term storage to a major extent; some involved mainly fluid reasoning and broad visualization. Some involved still other conglomerations of common factors identified in the various studies Carroll reviewed. The general factors did not meet the standards of even the weakest form of factor invariance (namely, configural invariance, Horn & McArdle, 1992). Although referred to as *the* (singular) general factor, the Schmid-Leiman factors were not replications of one factor; they were different factors.

In his 'Last Paper' criticism of Jensen's claims of support for the so-called Spearman hypothesis, Guttman detailed much of the evidence from structural (factor analytic) research indicating that the *g*-factor model does not fit data. Schönemann here and in his 1992 paper cites this evidence in his criticism of Jensen's claims. I have cited this structural evidence, also, in criticisms of Jensen's work (Horn & Goldsmith, 1981), and have reviewed a larger body of research the results of which indicate that we have yet to find a general principle uniting all features of human intelligence (e.g., Horn, 1989, 1991, 1997). This larger body of research provides evidence not only from structural studies, but also from studies of human development, behavior genetics, neurological function, and prediction of achievement. The evidence is pervasive and compelling, but it is almost entirely ignored by Jensen in his claims that a particular *g* runs through large and diverse batteries of tests.

When the *g*-factor model does not fit data, there remains an algebraic relationship between the mean difference vector and the PC1regression coefficients, but the correlation will no longer be a perfect 1.0. It will vary lower than 1.0 depending on the non-error variance of the $a_{j2} C_{2i} + \dots + a_{jn} C_{ni}$ components beyond the PC1 component. The e in equation [2] is replaced in part by non-error components. For example, if there were one non-error component beyond the first, the equation would be

$$Z_j = a_{j1} C_{1i} + a_{j2} C_{2i} + e. \qquad [3]$$

The mean of the jth variable would now be function not only of a_{j1} times the mean of C_{1i}, but also a function a_{j2} times the mean C_{2i}. The adding-in of the second component in Z_j dilutes the amount by which the means (for groups 1 and 2) are dependent on the product of a_{j1} and the mean of C_{1i}. Thus, the correlation between the differences between the means for groups 1 and 2 and the PC1 coefficients will be reduced. How much this correlation will be reduced will depend on how large is (the variance of) component 1 relative to component 2 or, more generally, relative to the nonerror components beyond component 1 that are added in and dilute the amount by which the means are dependent on component 1.

The average of the Spearman hypothesis correlations in the calculations Jensen did on 11 sets of data was .59 (approximately the same value Schönemann found in comparable analyses for 8 items relating to the toys, the playing and the reading of children and among which there was clearly no *g*-factor of general intelligence), This .59 is notably smaller than 1.0. It indicates that the components beyond PC1 contributed substantial amounts of variance to the variables and thus substantially affected the differences between the means for the two groups in Jensen's studies. The results of Jensen's analyses thus add to the evidence that test data are not well accounted for by one general factor (even when the general factor is not the same in different batteries of tests).

Thus it is clear that Schönemann is correct in calling attention to largely incorrect but widely announced claims that empirical evidence demonstrates that scientists know how to specify and define a single factor that reliably represents the sine qua non of all putative indicators of human intelligence – the same factor in different samples of such indicators – and that this factor reliably distinguishes average differences between populations of Black and White people. The empirical evidence does not support these claims. Schönemann's statement of this fact is vigorous and usually clear. I would hope that it will be read widely. It should help to break the field of scientific psychology away from the preoccupation it has evinced in theory of, research on, and use of, the much over simplified concept of general intelligence. The evidence on human cognitive capabilities indicates that this domain is multidimensional. Results from studies of methods has brought us to a point of understanding that research on cognitive capabilities can now, and should be, based on multigroup analyses of invariance of multivariate equational models (Gustafsson, 1992; McArdle, 1988; Meredith, 1993). Such research, not further divisive debate on race differences, should direct our efforts in the future.

REFERENCES

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.

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

Guttman, L. (1992). The irrelevance of factor analysis for the study of group differences. *Multivariate Behavioral Research, 27,* 175-204.

Horn, J. L. (1989). Models for intelligence. In R. Linn (Ed.), *Intelligence: measurement, theory and public policy* (pp. 29-73). Urbana, IL: University of Illinois Press.

Horn, J. L. (1991). Discussion of issues in factorial invariance. In Collins, L. M., & Horn, J. L. (Eds.), *Best methods for the analysis of change?* (pp. 114-125). Washington, DC: American Psychological Association.

Horn, J. (1997). A basis for research on age differences in cognitive capabilities. In J. J. McArdle & R. Woodcock (Eds.), *Human cognitive abilities in theory and practice*. Chicago, IL: Riverside Press.

Horn, J., & Goldsmith, H. (1981). Reader be cautious: A review of bias in mental testing. *American Journal of Education, 89,* 305-329.

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

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

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), *The handbook of multivariate experimental psychology* (Vol. 2., pp. 561-614). New York: Plenum Press.

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

Mulaik, S. A. (1992). Guttman's "Last Paper": A commentary and discussion. Editor's introduction. *Multivariate Behavioral Research, 27,* 173-174.

Schönemann, P. H. (1992). Extension of Guttman's result from g to PCl. *Multivariate Behavioral Research, 27,* 219-224.

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