Some new results on
the Spearman hypothesis artifact

PETER H. SCHÖNEMANN
Purdue University, West Lafayette, Indiana

This note constitutes a summary of the main results of two papers devoted to Jensen's (1980) "Spearman hypothesis": (1) Elementary geometry shows that positive ("Spearman") correlations between the mean black/white difference vector and the leading eigenvector of correlation matrices are artifacts, regardless of how one interprets Jensen's ambiguous definitions of "Spearman's hypothesis" (Level I versus Level II interpretation). (2) Empirically, the stronger Level II interpretation (which predicts positive correlations with *both within* eigenvectors) also arises with data that have nothing to do with g, such as *SES* variables indicating the number of toys and games. (3) Mathematically, the Level II interpretation implies not just approximate but perfect collinearity between the mean difference and the eigenvectors of all three covariance matrices, if one assumes multinormality, positivity of both subgroup covariance matrices, and an equal split into an HI and an LO group, regardless of whether Spearman's factor model holds or not.

The purpose of this advance note is to put some recent results regarding Jensen's Spearman hypothesis on record until more complete accounts can be published in the appropriate journals. In the meantime, the reader is referred to Schönemann (1988a, 1988b) for more detailed empirical evidence and proofs of formal claims.

Level I and Level II of Spearman's Hypothesis

In his book *Bias in Mental Testing*, Jensen (1980) devoted considerable space to "Spearman's interesting hypothesis that the magnitudes of white-black mean differences on various mental traits are directly related to the test's g loadings (Spearman, 1927, p. 379)" (p. 535). In a target article for *The Behavioral and Brain Sciences*, Jensen (1985) presents reanalyses of 11 data sets as evidence "that the varying magnitude of the mean difference between black and white populations on a variety of mental tests is directly related to the size of the test's loading on g, the general factor common to all complex tests of mental ability ... in accord with Spearman's hypothesis" (Jensen, 1980, abstract). On p. 206, Jensen spells out "an important practical implication of Spearman's hypothesis," at least as he sees it: "black people, statistically, will have a greater handicap in those educational, occupational, and military assignments that are most likely correlated with measures of general intelligence."

Jensen's ambiguous formulations of "Spearman's hypothesis" leave room for two technical interpretations of his definitions:

Level I: The mean difference vector correlates positively with the regression weights of the first eigenvector of the *pooled covariance matrix*.

Correspondence may be addressed to Peter H. Schönemann, Department of Psychological Sciences, Pierce Hall 365, Purdue University, West Lafayette, IN 47906.

Level II: The mean difference vector correlates positively with the regression weights of the first principal components of *both within-sample covariance matrices*.

Jensen (1980) appealed explicitly to the Level I version: "Probably the most compelling assemblage of evidence for the Spearman hypothesis ... is the massive data of the General Aptitude test of the US Employment Service" (p.549). Since "the correlations were not computed separately for black and white samples but are based on predominantly white samples" (Jensen, 1985, p. 216), this compelling evidence consisted of a positive correlation ($r = .71$) between the first principal component of the *pooled* correlation matrix and the mean difference vector, thus confirming only the Level I version.

Jensen (1985) added as one of several "methodological desiderata" that "ideally ... factor analysis should be performed in the two population groups separately," which points to the Level II interpetation. Since most of his *data*, though not his *definitions*, of "Spearman's Hypothesis" actually do point to the stronger Level II interpretation, it will be useful to clearly distinguish between these two technically different interpretations of Jensen's definitions.

Geometric Interpretations of the Level I and Level II Version

In 1985, I presented an algebraic argument, which shows that, under the Level I interpretation, "the predicted correlation is a psychometric artifact that arises with any data as long as the covariance matrices are equal and the mean vectors are sufficiently different" (Schönemann, 1985, p. 241). I illustrated this with two computer simulations (see Schönemann, 1985, Table 1). Here, a simpler, more direct geometric argument will be given here which easily extends to the stronger, Level II interpretation of Jensen's definition of "Spearman's hypothesis":

Level I: If we have two populations, HI and LO, and we pull them apart so that the length of the mean difference vector $d = \bar{y}_1 - \bar{y}_2$ increases, then the pooled population will contain a dominant first eigenvector (the main diameter of the pooled distribution) that will become more and more parallel to d, regardless of whether the two subpopulations form positive manifolds or not (Schönemann, 1985).

Level II: To extend this reasoning to Level II, one needs (1) a positive manifold and (2) the assumption that the pooled distribution is approximately multivariate normal (i.e., that the equidensity contours are ellipsoidal). Then any roughly even split into an HI and an LO group produces two attenuated within- covariance matrices whose principal axes (main diameters) will be approximately collinear with the principal axes of the pooled population. Hence the first eigenvectors will correlate highly with the mean difference vector in all three populations (Schönemann, 1986).

Concretely, if one cuts a banana in the middle, one is left with two elongated banana halves. As long as they are not moved, their main diameters point in the same direction as that of the whole banana (= pooled population), which, in turn, will be approximately collinear with the line through the points of gravity of both halves.

If, in the sample, the split of the pooled score matrix is very uneven (i.e., the sample size of one group is much larger than the sample size of the other), then the correlation between the mean difference vector and the first principal axes will be larger for the larger sample because (1) the joint distribution of the smaller sample will be more nearly circular, and (2) the within-sample correlation matrix for the smaller sample is less accurate. More generally, all these predictions rely on substantial sample sizes.

Illustration of the Artifact with Real (Head Start) Data

It is clear, then, that positive correlations between the mean difference vector of an HI and an LO subgroup and the leading eigenvectors will arise with any data that are roughly multinormal and sufficiently correlated to form positive manifolds in both subgroups.

For example, the same artifact arises with SES data contained in a report of the Westinghouse Learning Corporation (1969): *The Impact of Head Start* (I am grateful to V. Cicerelli, Department of Psychological Sciences, Purdue University, for providing me with a tape of these data, and to T. Putnam and W. Whitson of the Purdue University Computing Center for transferring them onto diskettes). Among many other variables, these data contain the following eight measures, which form positive manifolds in several subsamples of black and white children:

Variable 1: The number of toys that the child has that could be used in playing school.

Variable 2: The number of books that the child has to read.

Variable 3: How often the child reads by himself at home.

Variable 4: How often the respondent reads with the child.

Variable 5: The length of time that the child reads or was read to on the day before the interview.

Variable 6: The number of games that the child has.

Variable 7: How often the child plays with games.

Variable 8: How often the respondent plays games with the child.

Stratifying variable: Race.

These data were analyzed in eight smaller subsamples and also in six larger total samples. A detailed numerical example is given in Table 2 in Schönemann (1988b). The eight toy variables form positive manifolds in all subgroups.

For the smaller samples, most but not all Spearman correlations are positive, and their average is positive for both sets: .47 for the six large white samples, and .39 for the six large black samples. For the larger samples, all 12 correlations are positive.

Essentially the same picture emerges if one measures the collinearity between the standardized black/white difference vector and the two within-sample principal components in terms of cosines ("congruence coefficients"). The average cosines are .51 for the white samples and .45 for the black samples.

The prediction that, in the case of unequal splits, the Spearman correlations should be larger for the larger subgroups was borne out in all 5 of Jensen's (1985) 11 data sets that involve large enough sample sizes and uneven splits.

The predictions of the geometric model were also confirmed in 200 computer simulations involving various sample sizes and equal and unequal splits, as is described in more detail in Schönemann (1988b, pp. 13ff.; see Table 4).

Recent Theoretical Developments

In the fall of 1987, shortly before his death, L. Guttman gave me a copy of a manuscript he had submitted to *Behavioral and Brain Sciences* (Guttman, 1986). It contains a trenchant critique of many of Jensen's claims about Spearman correlations, positive manifolds, *g*, reaction time, and black/white differences.

In this 1986 paper, Guttman also proves the theorem that "actual proportionality must hold between the loadings and the standardized [mean] differences [if] Spearman's tetrad condition hold[s] for each of the subpopulations as well as for the total population" (p. 39; see also p. 46). However, as Guttman also notes (p. 13), this assumption will of course never be satisfied empirically: "Any reader of these lines can himself easily disprove 'g' by looking at almost any mental test correlation matrix at his disposal and checking for proportionality."

Guttman's result is not directly applicable to the Level II version of Spearman's hypothesis, since Jensen in effect abandoned Spearman's factor model long ago. He actually works with principal components, although he con-

tinues to *talk* about factors. However, in Schönemann (1988a) the following result about components rather than factors is proved (which does not require Guttman's assumption that Spearman's factor model holds in all three populations):

THEOREM: *If the range Re^p of a p-variate normal random vector $y \sim N_p(0, \Sigma)$, where Σ is positive with $diag(\Sigma') = diag(I)$, is partitioned into a High set (H) and a Low set (L) by the plane $\Sigma_k y_k = 0$, and both within-covariance matrices Σ_H, Σ_L remain positive, then (a) the mean difference vector $d := E(y|H) - E(y|L)$ is collinear with the largest principal component of Σ, and (b) d is also collinear with the PCIs of Σ_L, Σ_H.*

For the proof, see Schönemann (1988a).

Concretely, under the stated assumptions (in particular, multinormality and positive within-group covariance matrices), the *cosines* between the mean difference vector and the leading eigenvectors of both within-group *covariance* matrices are not just positive, but, except for sampling error, unity. This fact is obscured when the analyses are based on correlations instead of covariances and when collinearity is measured in terms of correlations instead of cosines.

Discussion

It thus emerges that positive Spearman correlations are artifacts under both interpretations of Spearman's hypothesis. They are a tautological consequence of two well-known facts: (1) that the subtests of IQ tests, largely by construction, correlate positively, and (2) that blacks, for whatever reasons, at present score lower than whites on these tests.

In particular, positive Spearman correlations tell us nothing whatsoever about "the nature of g": (1) for the Head Start toy data, the first principal component of the eight toy variables explained less than 10% of the variance of the first principal component of the Metropolitan Reading Readiness Test ($r = .27$); (2) by construction, the 200 score matrices used in the simulation did not contain any g (see Schönemann, 1988b, p. 14).

REFERENCES

GUTTMAN, L. (1986). *The irrelevance of factor analysis for the study of group differences*. Manuscript submitted for publication.

JENSEN, A. R. (1980). *Bias in mental testing*. New York: Free Press.

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

SCHÖNEMANN, P. H. (1985). On artifical intelligence. *Behavioral & Brain Sciences*, **8**, 241-242.

SCHÖNEMANN, P. H. (1986). *Level I and Level II artificial intelligence*. Manuscript submitted for publication.

SCHÖNEMANN, P. H. (1988a). A note on "Spearman's hypothesis" (Tech. Rep. No. 88-5). West Lafayette, IN: Purdue University, Purdue Mathematical Psychology Program.

SCHÖNEMANN, P. H. (1988b). Spearman's hypothesis and the General Toy factor (Tech. Rep. No. 88-2). West Lafayette, IN: Purdue University, Purdue Mathematical Psychology Program.

WESTINGHOUSE LEARNING CORPORATION. (1969). *The impact of Head Start*. Akron: Ohio University Press.