Multivariate Behavioral Research, 27 (2), 265-267

The "Spearman Hypothesis" is False

Jan-Eric Gustafsson
University of Göteborg

The major part of this second commentary will discuss Jensen's response to the Guttman criticism. In my brief commentary to Jensen's (1985) original article, I argued that methodological problems in Jensen's analyses prevent any strong conclusions to be drawn, so "the hypothesis can neither be accepted nor rejected on the basis of the analysis performed" (Gustafsson, 1985, p.232). In the response to Guttman's criticism, Jensen strongly defends his procedures and conclusions, claiming that "Spearman's hypothesis" has so consistently been supported that "... it may legitimately claim the status of empirical fact" (p. 232). However, if anything, the new round of discussion and thought stimulated by Guttman's article has made me even more critical of Jensen's procedures and results, and I now feel that both the methods and the conclusions should be rejected. In my opinion, there are serious problems in both steps of the two-step procedure employed by Jensen.

In the first step the tests' loadings on the general factor are estimated. However, it does seem that the general factor which comes out of most of the studies reanalyzed by Jensen (1985) is somewhat biased in the sense that verbal-educational tests (i.e., *Gc* tests) have too high loadings, and that tests of inductive reasoning (i.e., *Gf* tests) have too low loadings on this factor. This bias seems to be a consequence of the procedure of operationalizing the general factor as the first principal factor or the first principal component. Particularly when employed on small matrices these procedures yield a first factor which is influenced by the composition of the test battery. Most of the matrices reanalyzed by Jensen only include about a dozen variables, very few of which are good *Gf* tests, but many of which are good *Gc* tests. A possible effect of a verbal bias in the general factor is that the verbal factor may be underestimated. The verbal bias of the general factor may thus be responsible for Jensen's conclusion that "Contrary to popular belief, the mean black-white differences on the verbal factor (independent of *g*) is nil" (p. 231).

In the second step of Jensen's (1985) procedure, the factor loadings are correlated with the tests' observed black-white mean difference. In this step the very considerable overlap of tests between the studies is a problem, because it implies that the 11 empirical sets of data considered by Jensen do not provide

independent tests of the "Spearman hypothesis". The limited number of tests in each study implies, furthermore, that characteristics of a few tests will exert an unduly strong influence on the estimated size of the relationship between factor loadings and mean differences. For example, the Wechsler Digit Span tests, which has a low mean difference and a low *g*-loading, is included in no less than 6 of the 11 studies, and contributes thus considerably to the observed correlation. There is thus reason to believe that Jensen's results may not be as consistently supportive of the "Spearman hypothesis" over a wide range of tests as it is claimed by Jensen.

However, even if Jensen's (1985) empirical results are accepted, they do not seem to provide as clear support for "Spearman's hypothesis" as is claimed by Jensen. The weak version of the "Spearman hypothesis" states that "... the mean black-white differences on various tests is associated *predominantly* (rather than exclusively) with the tests' *g* loadings" (p. 231). However, as Roskam and Ellis observe in their commentary, the correlation between the vector of black-white mean differences and the vector of factor loadings is only about .6, which "seems to indicate that other factors besides *g* have a substantial effect on black-white differences, and this is perhaps a more serious criticism of Jensen's article than anything else" (p. 217). Thus, the strongest conclusion allowed by Jensen's results seems to be that the mean black-white difference is associated *to some extent* with the tests' *g* loadings.

There are two reasons why not even the weak form of Spearman's hypothesis is true, even if there is a substantial mean difference between blacks and whites on the *g* factor. The first reason is that the *g* factor only accounts for a part of the observed score variance, and in most cases only a limited part. For example, according to the model of the WISC-R subscales presented in my previous commentary the amount of *g* variance in the subscales varies between a high of 53% (Arithmetic) and a low of 12% (Mazes), with a mean of 29%. This rather low amount of influence of the *g* factor on each of the subscales limits the possibility for differences in the *g* factor to show up as mean differences in observed performance. The other reason why the "Spearman hypothesis" runs into problems is that there are differences in other factors, which are orthogonal to *g* and which are about equally influential as sources of variance in performance. To take the WISC-R example again, the Performance subscales tend to be about as highly related to *Gv*' (or spatial ability) as they are to *g*, and the black-white difference seems to be even larger on *Gv*' than it is on *g*. The WISC-R model also indicated that there are black-white differences on *Gc*' (or verbal ability) and, in the opposite direction, in a residual memory span factor. These factors are important sources of variance in some of the subtests and they thus reduce the relative influence of the *g* factor in the observed performance differences.

The "Spearman hypothesis" must thus be rejected even though there does seem to be a black-white mean difference in level on the general factor. However, this difference is only to a limited extent responsible for the variability of observed black-white differences on psychological tests. Jensen (1985) obviously has made the mistake of over-rating the importance of the general factor as a source of variance in performance. However, one of the most important and paradoxical characteristics of the general factor is that it is relatively unimportant as a source of individual differences in most narrow domains of performance, while it is a most important source of variance when performance over broad domains is assessed. Thus, the general factor is the major source of variance in any measure which is an aggregate of performance in several different domains, such as a factor score or a Full Scale IQ score computed from the WISC-R battery, but it has a limited influence on each of the subscales.

Among the commentaries of the other discussants, I especially enjoyed the brilliant contribution by Roskam and Ellis. In my opinion they clarify and settle most of the technical issues raised by Guttman. In particular, I found their clear exposition of Guttman's fundamental mistake of disregarding the influence of the specific factor in the one-factor model to be most enlightening. It thus seems that if Guttman and Jensen had remembered that Spearman's (1927) theory of intelligence was really a theory about *two factors*, and not only about one *g* factor, neither the Jensen (1985) article, nor the Guttman criticism would have been written.

References

Gustafsson, J.-E. (1985). Measuring and interpreting *g*. *The Behavioral and Brain Sciences*, *8*, 231-232.

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

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