
Multivariate Behavioral Research
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hmbr20

# More on Psychometric g and "Spearman's Hypothesis"

Arthur R. Jensen

Published online: 10 Jun 2010.

*Multivariate Behavioral Research*, *27* (2), 257-260


# More on Psychometric *g* and "Spearman's Hypothesis"

Arthur R. Jensen
University of California, Berkeley



What I have labeled "Spearman's hypothesis" was indeed only a "minor comment" on page 379 of his great work *The Abilities of Man* (1927). But that fact does not belittle its importance. Spearman's genius sparked a good many "minor comments" that are theoretically pregnant and waiting to be formalized as empirically testable hypotheses. This is obviously one of them. Was it such a heinous a crime that I chose to give him credit for the idea? I doubt that he would have objected to my empirical investigations of his "minor comment", and I imagine he would have been especially pleased that my effort has made others as well think about it. But it will make anyone happy, I will henceforth write "Spearman's hypothesis" in quotation marks.

Arguments about one-factor and two-factor theories of mental ability have long since been past history and to carry on about them in the present as if they were anything other than historic relics is fatuous pedantry. Everyone today is convinced of the existence of multiple factors, and few doubt that factors differ in *generality*, making it reasonable to represent them hierarchically in terms of that property.

Acknowledging the existence of a hierarchical general factor, or *g*, I believe Gustafsson's commentary (pp. 239-247) implicitly points up the most crucial issue, namely, what is the "best" *g* that any kind of factor analysis of psychometric tests of mental ability can give us? Is this a pseudo-question that in principle is unanswerable, like how many angels can dance on the point of a pin? Or is it theoretically or empirically a legitimate scientifically answerable question?

The problem is highlighted by Gustafsson's factor analysis of the correlations among the 13 subscales of the Wechsler Intelligence Scale for Children-Revised (WISC-R) given in Jensen and Reynolds (1982). I am not familiar with the particular procedure used by Gustafsson and so could not begin to suggest the reason for the discrepancy between his results and Jensen and Reynolds'. What argument can be made that Gustafsson's general factor (henceforth labeled *G*) is a better representation of the general factor of the WISC-R correlation matrix than Jensen and Reynolds' general factor (henceforth labeled *g*)? The Jensen and Reynolds *g* was derived by a Schmid-Leiman hierarchical factor analysis (done by Professor Schmid himself). The vector of Gustafsson's *G* loadings and the vector of Jensen and Reynolds' *g* loadings (averaged over the black and white samples) are correlated $r = .80$. The congruence coefficient is .99, indicating that *G* and *g* may be regarded as the same factor. But the correlation of .80 between the two sets of loadings is surprisingly low. Even between the black and white samples the *g* vectors have $r = .90$. The relatively low $r = .80$ between *G* and *g* has a surprisingly drastic effect on the test of "Spearman's hypothesis," that is, the correlation between the vector of general factor loadings and the vector of white-black mean differences (expressed in standard score units). (The standardized differences are henceforth labeled *D*.) The rank correlation between *g* and *D* is .67 ($p < .05$), bearing out "Spearman's hypothesis". But the $G \times D$ correlation is only .23 (n.s.). This is a much lower correlation, indeed an outlier, than any of the analogous correlations found in 15 independent studies of "Spearman's hypothesis". Is there something aberrant about Gustafsson's *G*? In all of the other studies the *g* was either a Schmid-Leiman hierarchical *g* or a first principal factor (PF1). It seems incumbent on Gustafsson to explain why his *G* is better than either the Schmid-Leiman *g* or the PF1. And can he point out anything *wrong* in the Jensen and Reynolds study, or explain why his result should be so discrepant from Jensen and Reynolds'?

Since Gustafsson (1988) correctly equates the *g* at the very top of the hierarchy of generality with fluid ability, *Gf*, he therefore regards a test such as Raven's Progressive Matrices as a good marker for assessing the construct validity of the *g* factor extracted from any test battery. I have found one published study (Vernon, 1983) in which Raven's Advanced Progressive Matrices (APM) was included in a factor analysis of the 11 subtests of the Wechsler Adult Intelligence Scale (WAIS), given to 100 university students. (The WAIS subtests are directly analogous to the WISC-R subtests, with the omission of Tappiing Span and Mazes.) The PF1 has its highest loading on the APM, followed by Block Design and Arithmetic. Hence this PF1 seems to accord with Gustafsson's specifications for a "good *g*." Yet the vector of loadings on Vernon's PF1 is correlated only .37 with Gustafsson's *G* loadings on the analogous subtests of the WISC-R. Despite the fact that Vernon's *g* (i.e., PF1) is based on university students, the vector of PF1 loadings (of 11 WAIS subtests) is correlated .79 ($p < .01$) with the vector of white-black *D*s on the corresponding WISC-R subtests in the Jensen and Reynolds (1982) study. Could it mean that Vernon's PF1 is a better representation of *g* than the *G* extracted by Gustafsson's method? (Or, for that matter, by the Schmid-Leiman method used by Jensen and Reynolds?) I would like to see Gustafsson's type of analysis applied to the matrix given in Vernon. And perhaps a more suitable application of it to testing "Spearman's hypothesis" than the Jensen and

Reynolds data would be the correlation matrix of 25 different tests in black and white samples matched in age, school, class, sex, and SES, in which the Schmid-Leiman $g \times D$ correlation was .78, bearing out "Spearman's hypothesis" (Naglieri & Jensen, 1987).

In Jensen and Reynolds (1982) it was noted that the mean white-black differences were significant for factor scores on each of the four independent factors (*g*, Verbal, Performance, and Memory) but was very much larger on *g* than on the other factors. Gustafsson's factor scores also show a large white-black difference on *G*, but the race differences are nearly the same on two other factors (*Gc*' and *Gv*', corresponding roughly to Verbal and Performance in the Jensen & Reynolds analysis). The reason for this difference is mysterious to me, and would have to be explained in terms of Gustafsson's model (see pp. 239-247), which seems to assign less variance to the general factor and more to the residualized first-order factors, *Gc*' and *Gv*', than is the case in the Jensen and Reynolds Schmid-Leiman analysis. Gustafsson's result with factor scores is at odds with numerous other analyses of this issue as well (see Jensen, 1987). Again, is Gustafsson's method of analysis better than the others? Why? It can't simply be a matter of personal preference, of course. Right here we can see real work cut out for the factor analysis experts, if indeed this is a problem that can be solved at that level. It can be seen that "Spearman's hypothesis" has two aspects: (a) prediction of the relative sizes of various tests' *g* loadings, and (b) determining the size of the mean group difference on the *g* factor in comparison with other factors. The first is much more sensitive to the reliability (and construct validity) of the rank order of tests' *g* loadings than is the second, which is more sensitive to how a test's variance is apportioned to the various common factors. The ideal method presumably would maximize the *correctness* of both aspects, whatever correctness may mean. I must leave that to be debated by experts in factor analysis.

In an earlier comment, Gustafsson (1985) wrote, "The tasks most clearly related to *g* seem to be complex nonverbal reasoning problems that are new to the examinees, the Raven Progressive Matrices being the archetypical example" (p. 232). Then there was his impressive finding that the highest-order *g* is indistinguishable from the lower-order *Gf* (Gustafsson, 1988). And I recalled, too, that the Raven matrices were found to be in the dead center in Guttman's *radex* map of abilities, exactly where *g* should be in his model (Snow, Kyllonen, & Marshalek, 1984). These points coming together suggested to me an entirely *fluid* test of "Spearman's hypothesis": Obtain (a) a vector (**A**) of the correlations of each of a dozen different reaction time (RT) variables with the Raven matrices, and (b) a vector (**B**) of the standardized mean black-white differences on each of the RT variables. I have now done this experiment with large samples totalling more than 800 white and black elementary school pupils in grades 4 to 6 (Jensen, in press). All the RT tasks were extremely *fluid*, involving no intellectual content, just lights and buttons, and the most complex of the tasks had a mean RT of only 0.7 second in these children. The correlation between vectors **A** and **B** is .80 ($p < .01$). So again it appears there is actually something to this phenomenon I've called "Spearman's hypothesis." I hope there will be more theoretical and empirical investigation of it.

## References


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

Gustafsson, J.- E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence, Vol. 3*. Hillsdale, NJ: Erlbaum.

Jensen, A. R. (1987). Further evidence for Spearman's hypothesis concerning black-white differences on psychometric tests. *Behavioral and Brain Sciences, 10*, 512-519.

Jensen, A. R. (in press). Spearman's hypothesis tested with chronometric information processing tasks. *Intelligence*.

Jensen, A. R., & Reynolds, C. R. (1982). Race, social class, and ability patterns on the WISC-R. *Personality and Individual Differences, 3*, 423-438.

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

Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence, Vol 2*. Hillsdale, NJ: Erlbaum.

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

Vernon, P. A. (1983). Speed of information processing and general intelligence. *Intelligence, 7*, 53-70.