What is the national *g* factor?

Jelte M. Wicherts

University of Amsterdam, Amsterdam, The Netherlands

Oliver Wilhelm

Humboldt Universität zu Berlin, Berlin, Germany

Correspondence concerning this manuscript should be addressed to Jelte M. Wicherts, Department of Psychology, Psychological Methods, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands, telephone: +31 205256880, fax: +31 206390026, J.M.Wicherts@uva.nl.

Abstract

Rindermann (in press) correlated the national averages of several student assessment studies and "national IQ" estimates and poses that these variables are all indicators of a common cognitive ability at the macro-social level, which he denotes the national *g* factor. We argue that Rindermann oversimplifies issues of individual differences and applies inappropriate statistical analyses. Therefore, we refute his conclusions.

Rindermann implicitly assumes that his national *g* factor is homogenous with the *g* factor at the inter-individual level. We argue that Rindermann has failed to establish that the nature and causes of this national *g* factor are similar (or identical) to the nature and causes of the inter-individual *g* factor (as for example in Carroll's [1993] widely accepted model).

Nature of the national g factor

Probably only a few educational or intelligence researchers would have issues with the idea that tasks from large scale educational studies share critical attributes with intelligence tasks. Nonetheless, most ability researchers will likely dismiss the notion that two different ability tasks such as Raven's Standard Progressive Matrices and the Reading Literacy test from PISA measure the same underlying ability. There is no serious doubt about the necessity of more specific lower order or nested factors to adequately account for covariances between ability tests (Carroll, 1993). If, for pragmatic reasons, only single indicators rather than a comprehensive intelligence test battery can be used, such measures should be decontextualized reasoning measures like the Raven's tests because there is little to no contribution to performance from intelligence factors other than *g* (at least in the west) (Gustafsson, 1984; Wilhelm, 2005). The literature on the factor analytical structure of inter-individual differences in cognitive ability is neglected when Rindermann factor analyzes the average scores of distinct achievement indicators across countries.

Rindermann discusses average scores on the level of nations in terms of notions drawn from studies at the level of individuals. However, he fails to establish a bridge between the individual and the national level. Multilevel data require the use of multilevel analyses (Muthén, 1991). Unfortunately, Rindermann uses nation's averages as his input and fits a

single factor model at the level of nations' averages, without verifying that test scores show comparable within-country factor structures across nations. The ecological correlations are indeed high, but it has long been known that ecological correlations may lead to results that are inconsistent with inter-individual correlations (Robinson, 1950).

For a comparison of test scores across nations to be meaningful, we have to establish that the tests measure the same underlying construct across nations and that the tests are not biased with respect to nations. Moreover, we have to establish that the underlying construct we are looking at across nations is indeed *g* and not another of the lower order factors that are known to play a role in cognitive test performance (Carroll, 1993). The starting point would be to study measurement invariance (Mellenbergh, 1989) across nations. Unfortunately, Rindermann does not discuss the issue of measurement invariance, he does not fit the relevant multilevel factor models (cf. Muthén, 1991), he does not consider model fit, nor does he study alternative factor models to the structure of aggregated data. Thus, Rindermann's analyses failed to show that national differences in average achievement and IQ test performance are due to national differences in average inter-individual *g*.

Causes of the national g factor

Even if the factor analytic results were trustworthy – and we must stress that they are not – we still need to consider what explains differences in national *g*. We do not know whether national differences in ability test performance have a genetic basis. This would require the finding of genes that account for intelligence and that differ in distribution across the globe (because cross-national twins do not exist). Lynn (2006) and Rushton (2000) pose

evolutionary theories related to race, but archeological findings are strongly at odds with these theories (MacEachern, 2006).

More importantly, a comparison of nations across the globe is fraught with many highly relevant confounds at the macro-social level. Many of these confounds are known or at least suspected to have strong effects on average cognitive ability test scores, which have risen substantially in developed countries over the course of the twentieth century (i.e., the Flynn Effect; Neisser, 1998). Health, nutrition, education, urbanization, trends towards smaller families, and the introduction of computers have all been proposed as causes for the Flynn Effect. However, developing countries like those in sub-Saharan Africa have not seen such developments. Wicherts (2007) has shown that estimates of national IQ (from Lynn & Vanhanen, 2002) correlate highly with basically all variables that have been proposed to have caused the Flynn Effect, such as secondary enrollment ratio (.78), pupil-to-teacher ratio (-.72), the number of PCs per 1000 persons (.66), fertility rate (-.86), urbanization (.67), general health as expressed in the child mortality rate (-.81), and nutrition as expressed in the amount of proteins in g per day per capita (.76). This begs the question whether the national *g* factor is indeed something that looks like *g* at the individual level, because on the basis of such strong correlations, national *g* looks suspiciously similar to the developmental status of countries.

We would like to add that the results by Rindermann are further compromised by many systematic flaws in Lynn and Vanhanen's (2002, 2006) IQ data. For example, the IQ values for African countries are consistently too low (Wicherts, 2007), which artificially inflates the correlations reported by Rindermann.

Conclusion

We think that Rindermanns' paper nicely illustrates three fundamental issues in the study of group differences. First, in order to understand group differences in some construct it is necessary to understand the nature of individual differences in this construct and it does not seem Rindermann adequately understood individual differences in the achievement data he reanalyzed. Second, when investigating group differences the use of adequate statistical tools (Mellenbergh, 1989; Muthén, 1991) is crucial, and unfortunately Rindermann fell short of using this set of appropriate tools. Third, when making controversial statements about the differences in national *g* through reanalysis of available data, these data must fit the problem and must be of unequivocal quality. The PISA, TIMSS, and PIRLS data used by Rindermann are of reasonable quality but don't fit to the problem. The IQ data used by Rindermann might fit to the problem, but are of poor quality.

References

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

Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8,* 179-203.

Lynn, R. (2006). *Race differences in intelligence: An evolutionary analysis.* Augusta, GA: Washington Summit Publishers.

Lynn, R., & Vanhanen, T. (2002). *IQ and the wealth of nations.* Westport, CT: Praeger.

Lynn, R., & Vanhanen, T. (2006). *IQ and global inequality.* Augusta, GA: Washington Summit Publishers.

MacEachern, S. (2006). Africanist archaeology and ancient IQ: racial science and cultural evolution in the twenty-first century. *World Archaeology, 38,* 72-92.

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

Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement, 28,* 338-354.

Neisser, U. (Ed.). (1998). *The rising curve: Long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.

Rindermann, H. (in press). The g factor of international cognitive ability comparisons: The homogeneity of results in PISA, PIRLS and IQ tests across nations. *European Journal of Personality.*

Robinson. W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 3,* 351-357.

Rushton, J. P. (2000). *Race, evolution, and behavior. A life history perspective*. Port Huron, MI: Charles Darwin Research Institute.

Wicherts, J. M. (2007). *Group differences in intelligence test performance*. Unpublished doctoral dissertation, University of Amsterdam, Amsterdam.

Wilhelm, O. (2005). Measuring reasoning ability. In O. Wilhelm, & R. W. Engle (Eds.), *Understanding and measuring intelligence* (pp. 373-392). London: Sage.