Research article.
Published in *Open Differential Psychology* DATE of MONTH, 2014
Submitted 20 of July, 2014

The international general socioeconomic factor: Factor analyzing international rankings

Emil O. W. Kirkegaard[1]

Abstract
Many studies have examined the correlations between national IQs and various country-level indexes of country well-being. The analyses have been unsystematic and not gathered in one single analysis or dataset. In this paper I gather a large sample of country-level indexes and show that there is a general socioeconomic factor (S factor) which is highly correlated (about .86-87) with national cognitive ability using either Lynn and Vanhanen's dataset or Altinok's. Furthermore the method of correlated vectors showed that the correlation between variable loadings on the S factor and cognitive measurements were .99 in both datasets using both cognitive measurements indicating that it is the S factor that drives the relationship with national cognitive measurements, not the remaining variance.

Keywords: National IQs, social progress index, democracy ranking, intelligence, g-factor, group differences, general socioeconomic factor, method of correlated vectors.

1 Introduction

Recently, research has been done on general socioeconomic factors. Gregory Clark argued that there is a general socioeconomic factor which underlies socioeconomic performance at the individual-level.[1] Moving one step up, earlier I found evidence that among 71 Danish immigrant groups ranked on 4 different measures of socioeconomic variables (crime, use of social benefits, income, education attainment) there was a large (40% of variance explained) general socioeconomic factor[2]. Since groups are mostly just collections of individuals, this leads to the expectation that there may be a general socioeconomic factor at the country level as well. The general mental ability factor at the individual level has been termed "g" (often italicized "*g*"[3]), while the national-level group equivalent has been termed "G" ("big g factor").[4] Keeping in line with this terminology, one might refer to the general socioeconomic factor at the individual level as "s factor" and the group level version "S factor" (or "big s").

There are by now many different national measures of country well-being. Perhaps the most common is Human Development Index published by the United Nations, but there are plenty of others e.g. Social Progress Index, Democracy Ranking, Quality of Life Index, Where-to-be-born Index, Democracy Index. Previous studies have correlated some of these with national IQs but not in a systematic manner (see review in [5]. Most of these national indexes have subcomponents that can be analyzed as well to see whether there is common variance, or whether the presence of these variables on the index is merely a function of their desirability as judged by the authors. The index scores are not outputs from a factor analysis, but usually just some weighted average of the components based on how important the authors thought they were.

The goals were thus: 1) gather a large collection of country well-being indexes and their components,

[1]Corresponding author: emil@emilkirkegaard.dk

2) perform a factor analysis[2] to check whether there is an S factor in the data and how reliable it is, and 3) examine how national cognitive measures correlate with these factors.

Since I was only concerned with the question of a general factor, all analyses used only the first factor.

2 Building the dataset

After I began the task of compiling all the datasets into one, it was apparent one reason why others had not assembled the data like this: it is rather work-some to combine datasets manually. This is because the different datasets do not contain just the same countries, and not in the same order, and they are also often not spelled exactly the same. For instance South Korea is variously called "South Korea", "Korea, Republic of", "S. Korea", "Korea Rep.", "Korea South" and variations in other languages as well. This makes it a lot of work to rearrange countries and their data in a spreadsheet manually. To avoid this, I wrote Python code[3] that can combine two datasets into one. Then I used this code to combine datasets for among others GDP per capita (from the International Monetary Fund, 2014)[6], national IQs from Lynn and Vanhanen (2012)[5], Human Development Index, Social Progress Index, Democracy Ranking, Quality-of-Life Index, Where-to-be-born index, and Democracy Index as well as numerous others. The entire dataset as well as source code is available at http://emilkirkegaard.dk/megadataset.

After inspecting the dataset, I decided to limit my analysis to the components of the Social Progress Index and Democracy Ranking. This is because both of these are very comprehensive having more than 40 components that cover a wide area of life. Furthermore, these big composites usually included many of the more narrow national rankings, so using both would result in double-sampling some measures.

2.1 Social progress Index (SPI) and Democracy Ranking (DR)

The SPI is very comprehensive and is a function of 54 basic components[7]. The structure is complicated and best shown visually as done below in Figure 1. A 56 page description of the SPI is found in the methodological report.[8]

The DR also has a large structure. It has 6 overall dimensions: political system, economy, environment, gender equality, health and knowledge. These have a total of 42 components.[9]

The DR sample of components overlaps heavily with that of SPI. As with the SPI, there is a 50 page methodological report that explains in further detail how the index works.[10] The main difference is between SPI and DR is that DR does include economic variables while the SPI only includes social and political. The SPI authors explicitly state they want to "move beyond GDP" (p. 8). However, the correlation between GDP per capita (IMF, 2013) and SPI is .860. The reader can judge for himself how well they succeeded.

2.2 Should one combine the SPI and DR?

One could combine the components from SPI and DR into one dataset by removing the duplicates. I did not do this because the indexes are already extremely comprehensive, it would reduce sample size and result in a still lower case-var ratio. The case-var ratio is already small: 132/54=2.44 in the SPI and 115/42=2.74 in the DR. Nathan Zhao has compiled an excellent overview of recommendations and practice concerning the case-var ratio.[11] He mentions that in 26% of a sample of studies using principal components in PsychINFO, the case-var ratio was between 2 and 5 as it is with my two datasets.

[2]For the purposes of this article, principal components analysis is considered a factor analytic method, even though some researchers argue that PCA should not be considered as such.

[3]Python is a free and versatile yet easy to use general purpose programming language. Learn more at https://www.python.org/. I used the Anaconda package. https://store.continuum.io/cshop/anaconda/

Figure 1: The structure of the Social Progress Index. Source: http://socialprogressimperative.org/

3 Initial analyses

I used R to do the analyses.[4]

The SPI dataset had missing data for some cases, between 1 and 5 missing values. I decided to examine the effect of using the reduced sample with complete data vs. imputing the means in the cases with missing values. The subsample with complete has N=78, while the full sample has N=132. In other words, using only the reduced sample implies a sample size decrease of 41%.

I performed principal components analyses (PCA) on both the reduced and the means-imputed datasets to examine the effect of the procedure. The Pearson correlation of factor loadings was 0.996 indicating that the procedure did not alter the structure of the data much. Similarly, the congruence coefficient was 1.0. I performed KMO tests (a measure of sampling adequacy) on both samples which showed that reducing the sample reduced the KMO (0.899 to 0.809). In comparison, KMO in the DR dataset was 0.884. All values are considered 'meritorious'.[12, p. 225]

Bartlett's test (tests whether the data is suitable for factor analysis) were extremely significant in all three datasets ($p<0.00001$).

In some cases PCA can show a general factor where none exists (Jensen and Weng, 1994 [13]). For this reason, I compared the first factor extracted via PCA in the SPI dataset to the first factors using minimum residuals (MinRes), weighted least squares (WLS), generalized least squares (GLS), principal axis factoring (PAF) and maximum likelihood estimation (ML). Did it make a difference?

[4]R is a free, powerful, and yet easy to use programming language designed for data mining and statistics. See http://www.r-project.org/

Not much. The mean Pearson intercorrelation between factor scores is 0.992. The mean Pearson intercorrelation between factor loadings is .999. The mean factor congruence is .999.

Jensen and Weng also noted that PCA tends to overestimate the variance accounted for by the first factor. For this reason I compared the variance accounted for in each method of analysis. There was no sign of PCA overestimating since all methods resulted in the first factor explaining between 40 and 42%.

Major[14] reported that PCA tends to overestimate factor loadings. However, the mean absolute factor loading was similar across all 6 extraction methods in both datasets (SPI: range .582-.603).

In the DR dataset results were almost identical. The mean Pearson intercorrelation between factor scores is .990. The mean Pearson intercorrelation between factor loadings is .998. The mean factor congruence is .999. The variance accounted is between 44 and 47%.

In the full analyses, method of extraction made very little difference for the first factor.

4 Subset x whole comparisons: How many components are necessary?

Since I found that regardless of method and dataset used, the first factor was a general factor accounting for about 40-47% of the variance, it was interesting to know how many components one needed to measure it well. To find out, I sampled subsets of components at random from the datasets, extracted the first factor, and then correlated the scores from it with the scores of the first factor using all the components. I repeated the sampling 1000 times to reduce sampling error to almost zero. Since recently there was interest in comparing g factors from different factor extraction methods, I used the 6 different methods mentioned before.

Results are shown in Figure 2 for SPI. Each step shows the increase in correlation adding another random component. They show that the number of components necessary to estimate the first factor from all 54 components well is small. For PCA, picking one component at random gives .603 which is the same as the average (absolute) loading on the first factor in the full analysis. Using the first factor of 5 components yields an average correlation of 0.877.

Since PCA is usually singled out for criticism, I compared PCA with the average of all the other methods. As one can see, PCA is a bit higher for the small samples. Even the largest difference is quite small, however. The difference in the SPI dataset with 3 components is .041.

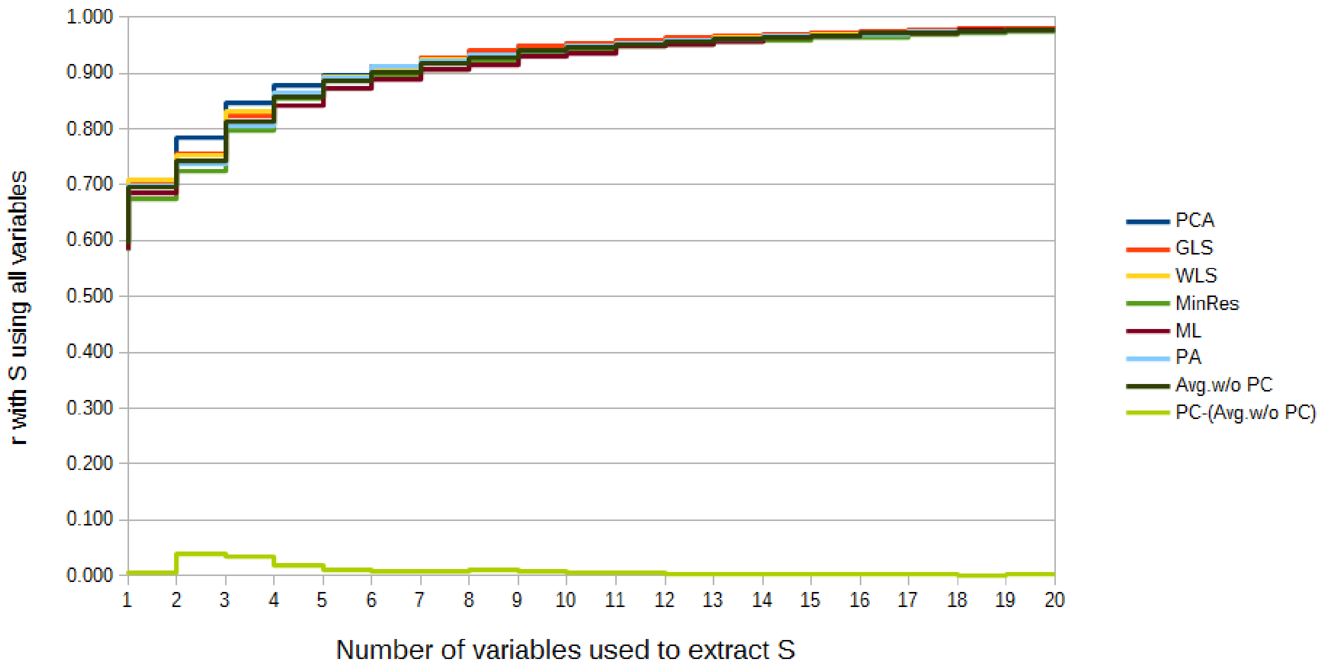

Figure 2: Component subset factor correlation with S factor in Social Progress Index dataset. The plot shows the average correlation of the first factor within a sample of X components with the first factor of all 54 components.

The results from the same analysis using the DR dataset are shown in Figure 3. They are nearly identical.

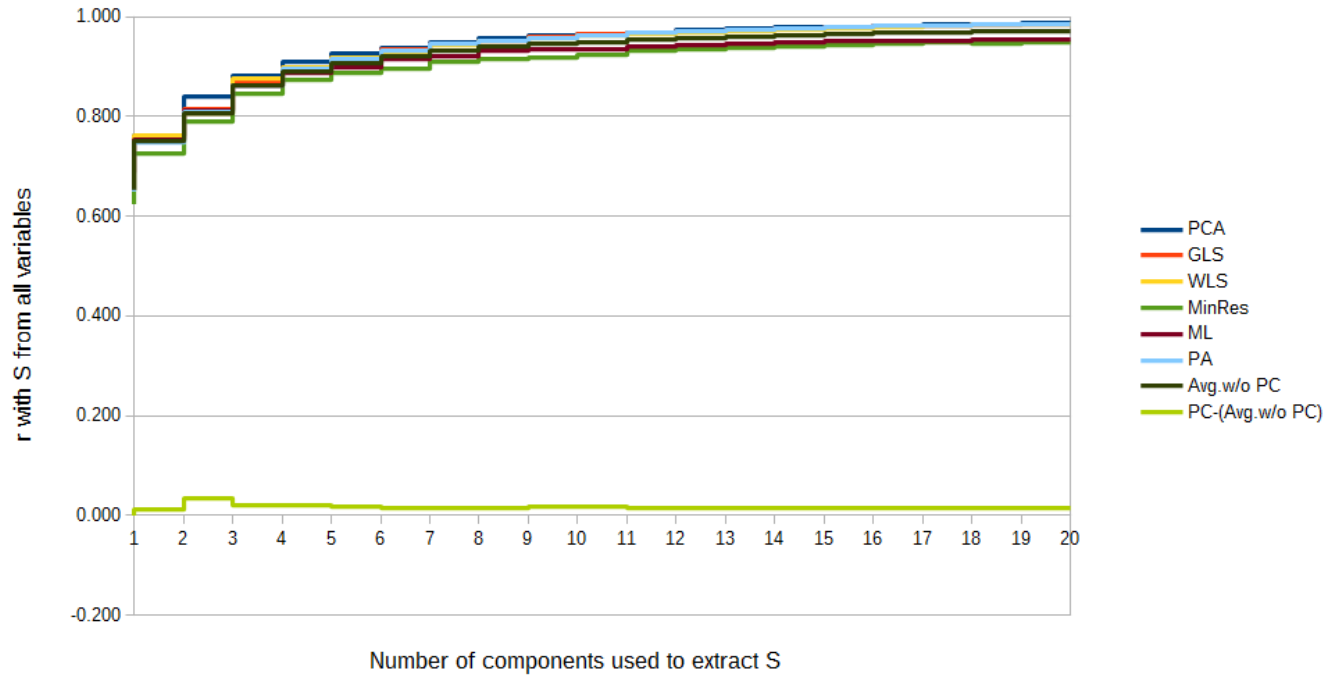

Figure 3: Component subset factor correlation with S factor in Democracy Ranking dataset. The plot shows the average correlation of the first factor within a sample of X components with the first factor of all 54 components.

5 Subset x subset comparisons

Another method similar to the above is to ask: How well does the first factor from a random subset correlate with that of another random subset, with no overlap in components? I tested this in both datasets with subset sizes 5 and 10 with PCA. Results are shown in Table 1.

N subsets / Dataset	SPI	DR
N=5	.758	.804
N=10	.874	.902

Table 1: First factors from different and random subsets of the same dataset.

Results are high but not near unity. They show again that, as long as one picks a reasonable number of them, e.g. 10., it is not so important which subset of national indexes one chooses since they measure to a large degree a common factor: S.

6 Factor loadings with different methods

It has also been claimed that PCA tends to assign higher loadings to variables than other methods, especially when used on a small number of variables.[15, 16] To see if this was also true for my international data, I extracted the mean absolute loading from 1000 analyses of each sample size 2 to 20 with each of 6 different methods. Results are shown in Figures 4 and 5.

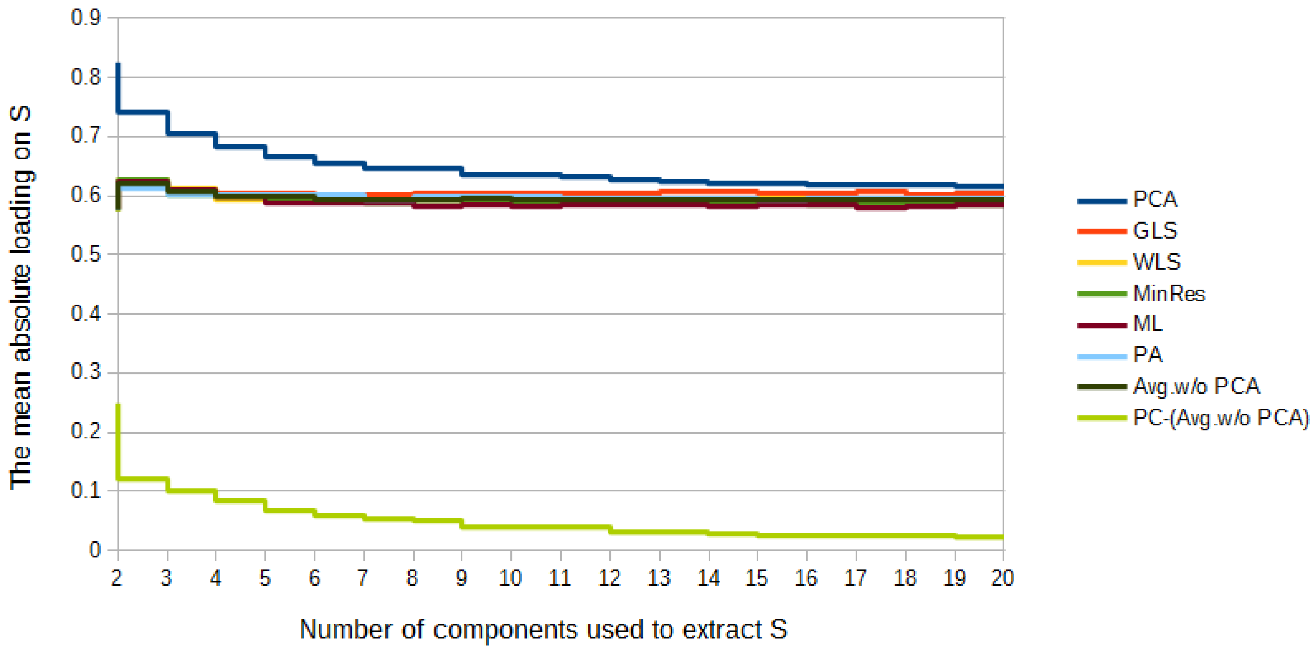

Figure 4: Mean absolute loadings on the first factor with different methods and different sample sizes of variables. Results for SPI dataset.

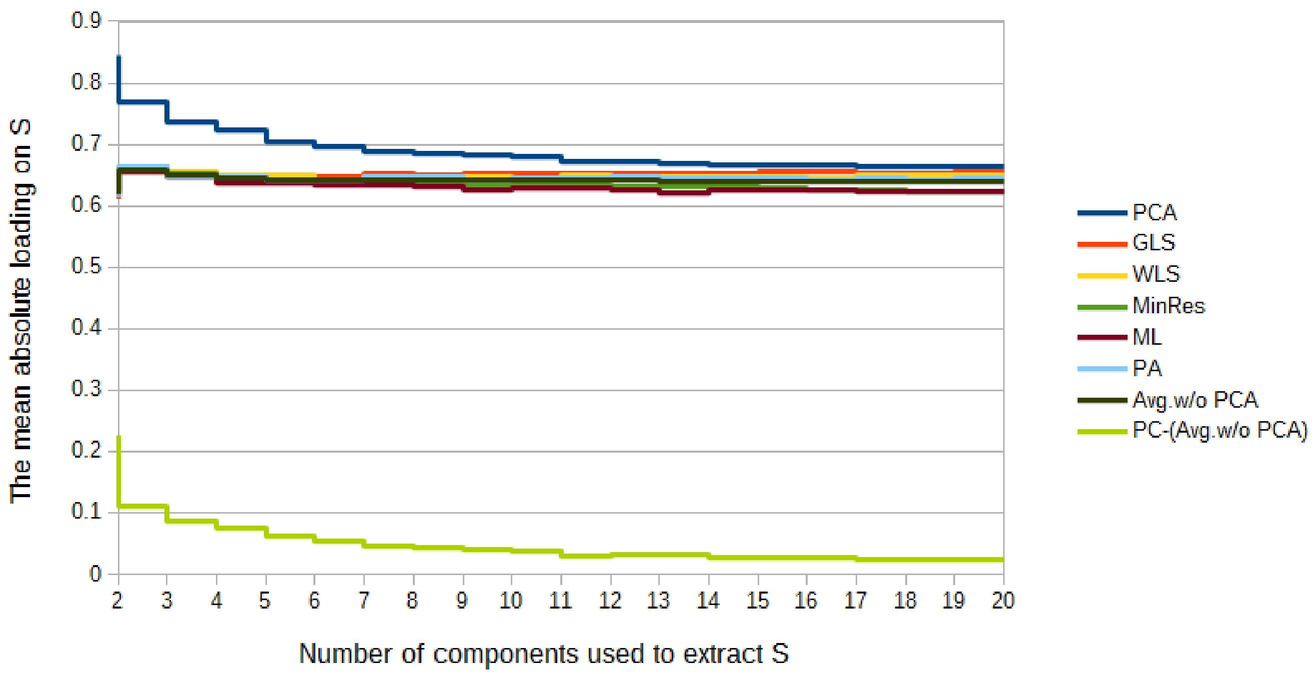

Figure 5: Mean absolute loadings on the first factor with different methods and different sample sizes of variables. Results for DR dataset.

Results were nearly identical for both datasets. As claimed, the loadings using PCA were higher than with the other methods and this was especially the case for smaller samples of variables. Thus, Major et al[15] seem to be right: It is unwise to base interpretations on the strength of loadings in small samples when using PCA.

7 Relationship with national cognitive ability measures

As it is usually done, one finds some interesting country-level variable and then regresses it on national IQ, perhaps with some controls. Often authors will also argue for a causal connection from national IQ/G to country-level variables. The typical example of this is wealth (e.g. [17, 18, 19, 20, 21]). Since it is known that g causes greater wealth at the individual level, and that nations can generally be considered a large group of individuals, it would be very surprising, though not impossible, if there was no causation at the group level as well.

I don't want to argue at length for any causal role in this paper, so I merely present the correlations and leave the interpretation to the reader. I chose to look at two cognitive ability measures, the total scores from SPI and DR, as well as the first (PCA) factors from the datasets. For cognitive measures I used Lynn and Vanhanen's 2012 national IQ estimates[5] and Altinok's educational achievement estimates[22] as an alternative cognitive measurement. Results are shown in Table 2.

If population differences in G is a main cause of national differences in many socioeconomic areas, then aggregating measures should increase the correlation with G, since measurement specificity averages out.

r	SPI	DR	IQ	Altinok	SPI_1	DR_1
SPI	1.000	0.928	0.819	0.839	0.981	0.965
DR		1.000	0.712	0.760	0.893	0.909
IQ			1.000	0.910	0.856	0.868
Altinok				1.000	0.870	0.872
SPI_1					1.000	0.975
DR_1						1.000

Table 2: Correlation matrix with Social Progress Index, Democracy Ranking, their first factors, national IQs and Altinok's achievements scores.

As expected, the correlations with the aggregated country-level measures have a very strong correlation with proxies for G. All correlations are significant beyond the $p<0.001$ level (N's 100-132). Furthermore, the correlations between cognitive measures and the S factors from both datasets were stronger than the indexes as made by the authors. This indicates that it is the first factor that drives the relationship, not the remaining variance. Global hereditarians[23] may interpret this result as being in line with predictions, while non-hereditarians may interpret as showing that national differences in S cause national differences in G, or something else entirely.

I don't report the results using S extracted from the other methods because they are nearly identical due to the very high intercorrelations between factor scores from each method.

The curious reader can also examine tables 4 and 5 in the appendix for correlations between each single variable and national IQs (96 correlations).

8 The method of correlated vectors and the S factor

As with testing of mental abilities, it is known that different tests have different loadings on the first factor, their g-loading.[3]. In other words, some tests measure the underlying factor better than others. Jensen invented the method of correlated vectors to test whether it was the g factor that was related to some other variable, or some other part of the variance. I can use the same method here in reverse fashion to test whether it is the S factor found in the SPI and DR datasets that drives the correlation with IQ or the non-S variance.

I thus calculated the correlation of both cognitive national measures with every component in both datasets and correlated this with the loadings from the PCA. In both datasets using both national cognitive measures the correlation was .99 confirming the indication from before. Nevertheless, it is possible to have a very high correlation and not have a very linear relationship as demonstrated by Anscombe's Quartet.[24] Inspection of all the scatterplots, however, reveal that the relationships really are very linear. Figure 6 shows a plot of one of these MCV analyses. Tables 4 and 5 in the Appendix contains both the factor loadings on the S factor and the correlations of each variable with national IQs.

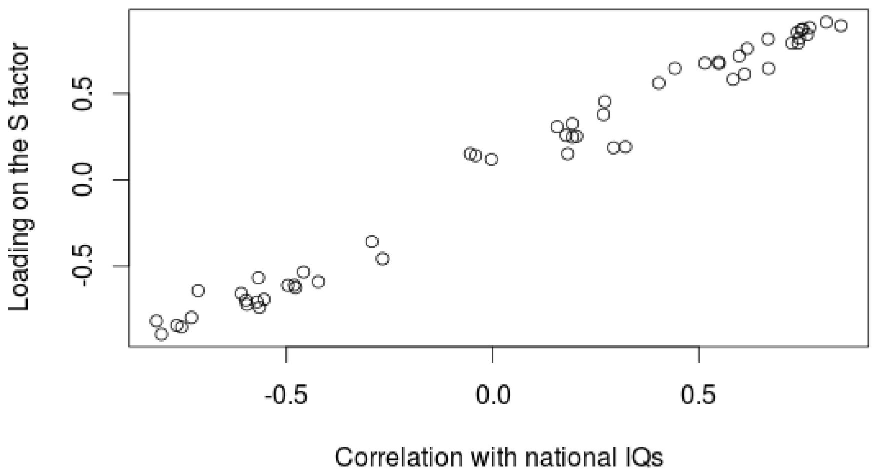

Figure 6: Variable loadings on the S factor (PCA) and their correlation with national IQs in the SPI dataset.

These very high correlations resulting from the use of method of correlated vectors in these two datasets give indirect support for researchers who have been arguing that the heterogeneous and lower than unity results with the method used on g-loadings are due to statistical artifacts, especially sampling error and restriction of range of subtests (e.g. [25, 26, 27, 28] see also [29]). None of these sources of errors appear large in the present analysis since I used 54 and 42 components varying wildly in their absolute loading from almost 0 to almost 1. In the SPI dataset, the min is .12 and the max .92 with a mean of .60 and SD of .24. The results are almost identical in the DR dataset with min, max, mean, SD at .11, .92, .65, .22 respectively (all values using PCA).

There is a question concerning whether it is proper to analyze do the MCV analyses without reversing the variables that have negative loadings on the S factor first. Using the non-reversed variables means the variance is higher which increases the correlation. Reversing them would decrease the correlations. I decided to use the data as they were given by the authors i.e. with no reversing of variables. As one can see from the plots, reversing them would not substantially change the results.

Furthermore, it is not quite obvious just which ones to reverse. With cognitive data, there is never any disagreement about which direction of a value is associated with better performance. With social, political and economic variables this is not always the case. For instance, in the DR dataset, there are two variables (28-29) dealing with expenditure on health: public and private. The public has a strong positive loading (.658) and the private has a small negative loading (-.210). Which directions are these to be reversed if they are? What about electric power consumption per capita (17)? It seems best not to get bogged down into discussions of which of these should be reversed and just use the data as given by authors. Recall that they could not have biased the data to fit my analyses since they put together the datasets for other purposes.

If, however, one insists on reversing the obvious cases, I have done this for those that I considered obvious enough. These are variables 1:6,8,13:18,26,28:33,35,42,48,53 from the SPI dataset. Re-doing the PCA and MCV analysis gives a correlation between S factor loading and national IQ of .98. Very little difference.

9 S loading and the desirability of component: Are the first factors general factors?

The mean absolute loading on the S factors is quite high (.6-.65). If one inspects the loadings (as shown in the Appendix, Tables 4 and 5), one will note that what loads positively on the S factor is generally considered something desirable, and reversely for negative loadings. For example, the first variable in the SPI dataset is percent of the population that is undernourished. This is clearly something generally considered undesirable and the loading is strongly negative (-.71).

There are a few components, however, where the S-loading and the desirability of the variable are opposite. In the SPI dataset, obesity rate has a loading of .65 on S but it is generally thought to be something undesirable, even if it indicates, evolutionarily speaking, that food is plentiful and absence of non-intentional starvation. The suicide rate from the same dataset shows the same pattern (loading .19). Since S and national cognitive measures (proxy for G) are so highly correlated, the fact that other studies have found that higher national IQ predicts higher suicide rates fits quite well with this finding.[30, 31] There are two more variables of interest: Death per capita by air pollution has an S loading of .192. Finally, use of water resources has a loading of .259. There are no variables in the DR dataset like this. All in all, there are 4 variables out of about 100 that having opposite signs on their S loading and desirability.

It is worth noting that group-level correlations need not be the same or even in the same direction as individual-level correlations. In the case of suicide, there does appear to be a negative correlation at the individual level as well.[32] However, in the case of obesity, it is commonly found that higher s (and g) predicts lower levels of obesity, not higher.[33]

In cognitive data if one transforms variables so that a positive value corresponds to better performance (e.g. by reversing response time variables since longer reaction times indicate worse performance), then there is a positive manifold: Every variable has a positive correlation with any other variable. This pattern is seen both in individual-level data with the g factor and in country-level data with the G factor.

In light of the above, one might wonder whether it is fair to call the first factors ”general factors” when there are 4 variables that go against the pattern. It comes down to which meaning of ”general” is used. Dictionary.com gives among others these two definitions:

1. of or pertaining to all persons or things belonging to a group or category: a general meeting of the employees.

2. of, pertaining to, or true of such persons or things in the main, with possible exceptions; common to most; prevalent; usual: the general mood of the people.

When I say ”general”, I mean the second sense, not the first, and in this sense they are clearly general pertaining to about 95% of the variables. Note that in the cognitive domain, there is a general factor no matter which interpretation is chosen.

10 Number of factors to extract and their size

All previous analyses used only the first factor however, one might want to know how many factors standard methods indicate that one should extract and if doing so changes results. The R package nFactrors includes a function[5] that calculates the number of factors to extract using 4 different

[5] nScree(), http://artax.karlin.mff.cuni.cz/r-help/library/nFactors/html/nScree.html

methods (optimal coordinates, acceleration factor (a non-subjective measure based on the Scree plot), parallel analysis and Kaiser's Rule (Eigen value ¿1).

I ran the analysis on both datasets. In the SPI dataset, three of the four methods indicated the number of factors to extract was 9, acceleration factor being the outlier which suggested only to extract 1. The situation was similar in the DR dataset with three methods suggesting 8 factors and acceleration factor again only suggesting 1.

Extracting the factors using maximum likelihood estimation, both Pearson correlations between scores (SPI: .98, DR: .97), loadings (SPI: 1.00, DR: 1.00) and congruence coefficients (SPI: 1.00, DR: 1.00) showed that the factors from the 1 factor analyses were very similar or identical with the first factors from the 8-9 factor analyses.

It is also worth mentioning that the percent of the variance accounted for by each factor quickly decreases from the first factor, and that the first and second factor are not even close in size. Results are shown in Table 3. The variance accounted for by the first factor was slightly lower than in the 1 factor analysis (.40 vs. .39 for SPI, and .46 vs. .43 for DR).

Proportion var. for factor N	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
SPI	0.39	0.08	0.07	0.04	0.03	0.03	0.02	0.02	0.01
DR	0.43	0.07	0.07	0.04	0.03	0.02	0.02	0.02	

Table 3: Proportion of variance accounted for each factor. Both datasets.

The relative size of the first factors compared to the second (SPI: .39 to .08, ratio = 4.9; DR: .43 to .07, ratio = 6.1) justifies calling them "large". Together with the evidence of their generalness in the previous section, this justifies calling them "large general factors". This would not have been the case if either they were not general or that the first two factors were about equally large.

11 Correlated factors

Another way to test if there is a general factor, is to use repeated factor analysis while allowing for correlated factors (oblimin rotated). This is sometimes considered the best way to extract g.[13]

The approach taken was this:

1. Determine number of factors to extract with nScree (choosing the number most criteria agree on).
2. Extract that number of factors with oblimin rotation using maximum likelihood estimation.
3. Repeat the above steps.

Using this method on cognitive data usually results in a g (general) factor at the second or third level.[34] The same result was found for both datasets with international data. The 3rd order factor correlated .97 with the first unrotated factor in both datasets indicating robust results. In the SPI dataset, there was 9 factors at the first level, 3 at the second and 1 at the third. In the DR dataset, there were 8, 2, and 1, respectively.

12 Discussion and conclusion

In the measurement of mental abilities, there has been a long standing debate on whether g from one battery of tests was the same g from another battery of tests. The issue now seems to have been settled decisively by a series of studies that compared extracted g factors from different subsets of batteries and different batteries to each other and found them to be very close to unity.[35, 36] Another method that has been used is to sample random subsets from the entire battery of tests, factor analyze the subset with and without a probe test, and then compare the g-loading of this probe test in different subsets. Studies of this kind found quite stable g-loadings.[37, 38, 39, 16, 14]

The analyses carried out in this paper suggest that the S factor is not quite like g but close. The two S factors from the two different datasets did not reach the .99 level (r = .975) even though they were extracted from a very varied set of 42 and 54 components. Perhaps the S factor is inherently more fuzzy or nebulous than g. Perhaps the exclusion of economic variables in the SPI dataset means that there is a small lack of sampling variance.

The analyses of extracting S factors from subsets of variables and comparing them either to each other (subset x subset) or the the S factor from the complete analysis (subset x whole) suggest that S can be quite reliably approximated picking a small collection of random variables. It would be interesting to redo this analysis on the cognitive data analyzed by Johnson et al.[35, 36]

It is also worth noting that mental test data and national rankings data are not strictly comparable in one way. The mental test variables are actually themselves some sort of sum of (usually unweighted) the actual test items (e.g. number of matrices solved). This is not generally the case for the national variables although some of them, e.g. Freedom House indexes, are some sort of function of lower level components. It is not clear which direction this disanalogy biases results in, if any.

Results from both datasets examined in this study were generally extremely similar despite the fact that the datasets were created by two different sets of researchers with different goals, none of which was to look for any country-level general socioeconomic factor. This indicates that the results are robust.

13 Datasets and source code

This paper is open source. Everything necessary to reproduce the PDF and results is available in the supplementary material at the journal website.

14 Acknowledgments

Thanks to Laird Shaw for help with R.

References

[1] Gregory Clark. *The Son Also Rises: Surnames and the History of Social Mobility*. Princeton University Press, 2014.

[2] Emil O. W. Kirkegaard and John Fuerst. Educational attainment, income, use of social benefits, crime rate and the general socioeconomic factor among 71 immmigrant groups in denmark. *Open Differential Psychology*, 2014.

[3] Arthur Robert Jensen. *The g factor: The science of mental ability*. Praeger Westport, CT, 1998.

[4] Heiner Rindermann. The big g-factor of national cognitive ability. *European Journal of Personality*, 21(5):767--787, 2007.

[5] Richard Lynn and Tatu Vanhanen. *Intelligence: A unifying construct for the social sciences*. Ulster Institute for Social Research, 2012.

[6] International Monetary Fund. World economic outlook database, 2014. URL `http://www.imf.org/external/pubs/ft/weo/2014/01/weodata/index.aspx`.

[7] Social Progress Imperative. Social progress index 2014, 2014. URL `http://www.socialprogressimperative.org/`.

[8] Scott Stern, Amy Wares, Sarah Orzell, and Patrick O'Sullivan. *SOCIAL PROGRESS INDEX 2014: METHODOLOGICAL REPORT*. Social Progress Imperative, 2014. URL http://www.socialprogressimperative.org/publications.

[9] Democracy Ranking. Democracy ranking, 2014. URL http://democracyranking.org/.

[10] David F. J. Campbell. *The Basic Concept for the Democracy Ranking of the Quality of Democracy*. Vienna: Democracy Ranking, 2008. URL http://democracyranking.org/?page_id=590.

[11] Nathan Zhao. The minimum sample size in factor analysis, 2009. URL https://www.encorewiki.org/display/~nzhao/The+Minimum+Sample+Size+in+Factor+Analysis.

[12] Graeme D Hutcheson and Nick Sofroniou. *The multivariate social scientist: Introductory statistics using generalized linear models*. Sage, 1999.

[13] Arthur R Jensen and Li-Jen Weng. What is a good g? *Intelligence*, 18(3):231--258, 1994.

[14] Jason Major. The dependability of the general factor of intelligence: Why g is not a first principal component. 2010.

[15] Jason T Major, Wendy Johnson, and Thomas J Bouchard Jr. The dependability of the general factor of intelligence: Why small, single-factor models do not adequately represent g. *Intelligence*, 39(5):418--433, 2011.

[16] Randy G Floyd, Elizabeth I Shands, Fawziya A Rafael, Renee Bergeron, and Kevin S McGrew. The dependability of general-factor loadings: The effects of factor-extraction methods, test battery composition, test battery size, and their interactions. *Intelligence*, 37(5):453--465, 2009.

[17] Richard Lynn and Tatu Vanhanen. *IQ and the wealth of nations*. Greenwood Publishing Group, 2002.

[18] Richard Lynn and Tatu Vanhanen. *IQ and global inequality*. Washington Summit Publishers, 2006.

[19] Gerhard Meisenberg. National iq and economic outcomes. *Personality and Individual Differences*, 53(2):103--107, 2012.

[20] Heiner Rindermann and James Thompson. Cognitive capitalism the effect of cognitive ability on wealth, as mediated through scientific achievement and economic freedom. *Psychological Science*, 22(6):754--763, 2011.

[21] Gregory B Christainsen. Iq and the wealth of nations: How much reverse causality? *Intelligence*, 41(5):688--698, 2013.

[22] Nadir Altinok, Claude Diebolt, and Jean-Luc Demeulemeester. A new international database on education quality: 1965--2010. *Applied Economics*, 46(11):1212--1247, 2014.

[23] John Fuerst and Emil O. W. Kirkegaard. Do national iqs predict u.s. immigrant cognitive ability and outcomes? an analysis of the national longitudinal survey of freshman. *Open Differential Psychology*, 2014. URL http://openpsych.net/ODP/2014/04/do-national-iqs-predict-u-s-immigrant-cognitive-ability-and-outcomes-an-analysis-of-t

[24] Wikipedia. Anscombe's quartet --- wikipedia, the free encyclopedia, 2014. URL http://en.wikipedia.org/w/index.php?title=Anscombe%27s_quartet&oldid=613479365. [Online; accessed 4-August-2014].

[25] Jan te Nijenhuis, Annelies EM van Vianen, and Henk van der Flier. Score gains on g-loaded tests: No g. *Intelligence*, 35(3):283--300, 2007.

[26] Jan te Nijenhuis and Henk van der Flier. Is the flynn effect on g?: A meta-analysis. *Intelligence*, 41(6):802--807, 2013.

[27] Jan te Nijenhuis. The flynn effect, group differences, and g-loadings. *Personality and Individual Differences*, 55(3):224--228, 2013.

[28] Jan te Nijenhuis, Birthe Jongeneel-Grimen, and Emil O. W. Kirkegaard. Are headstart gains on the g factor?: A meta-analysis. *Intelligence*, In Press, 2014.

[29] John E Hunter and Frank L Schmidt. *Methods of meta-analysis: Correcting error and bias in research findings*. Sage, 3rd edition, 2014.

[30] Martin Voracek. National intelligence and suicide rate: an ecological study of 85 countries. *Personality and Individual Differences*, 37(3):543--553, 2004.

[31] Donald I Templer, Heather Joy Connelly, David Lester, Hiroko Arikawa, and Leah Mancuso. Relationship of iq to suicide and homicide rate: An international perspective 1. *Psychological reports*, 100(1):108--112, 2007.

[32] Lars Andersson, Peter Allebeck, J-E Gustafsson, and D Gunnell. Association of iq scores and school achievement with suicide in a 40-year follow-up of a swedish cohort. *Acta Psychiatrica Scandinavica*, 118(2):99--105, 2008.

[33] T Chandola, IJ Deary, D Blane, and GD Batty. Childhood iq in relation to obesity and weight gain in adult life: the national child development (1958) study. *International journal of obesity*, 30(9):1422--1432, 2006.

[34] John Bissell Carroll. *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press, 1993.

[35] Wendy Johnson, Thomas J Bouchard Jr, Robert F Krueger, Matt McGue, and Irving I Gottesman. Just one g: consistent results from three test batteries. *Intelligence*, 32(1):95--107, 2004.

[36] Wendy Johnson, Jan te Nijenhuis, and Thomas J Bouchard Jr. Still just 1 g: Consistent results from five test batteries. *Intelligence*, 36(1):81--95, 2008.

[37] Robert L Thorndike. Stability of factor loadings. *Personality and Individual Differences*, 8(4): 585--586, 1987.

[38] Malcolm James Ree and James A Earles. The stability of g across different methods of estimation. *Intelligence*, 15(3):271--278, 1991.

[39] Charlie L Reeve and Nikki Blacksmith. Equivalency and reliability of vectors of g-loadings across different methods of estimation and sample sizes. *Personality and Individual Differences*, 47(8): 968--972, 2009.

Appendix

Loadings for first factor in Social Progress Index		
Component	*Loading*	*r x IQ*
Undernourishment (% of pop.)	-0.71	-0.57
Depth of food deficit (calories/undernourished person)	-0.69	-0.55
Maternal mortality rate (deaths/100,000 live births)	-0.84	-0.77
Stillbirth rate (deaths/1,000 live births)	-0.86	-0.75
Child mortality rate (deaths/1,000 live births)	-0.90	-0.80
Deaths from infectious diseases (deaths/100,000)	-0.82	-0.82
Access to piped water (% of pop.)	0.88	0.77
Rural vs. urban access to improved water source (absolute difference between % of pop.)	-0.70	-0.60
Access to improved sanitation facilities (% of pop.)	0.86	0.74
Availability of affordable housing (% satisfied)	0.25	0.20
Access to electricity (% of pop.)	0.84	0.76
Quality of electricity supply (1=low; 7=high)	0.79	0.74
Indoor air pollution attributable deaths (deaths/100,000)	-0.74	-0.57
Homicide rate (1= <2/100,000; 5= >20/100,000	-0.64	-0.71
Level of violent crime (1=low; 5=high)	-0.57	-0.57
Perceived criminality (1=low; 5=high)	-0.54	-0.46
Political terror (1=low; 5=high)	-0.61	-0.48
Traffic deaths (deaths/100,000)	-0.66	-0.61
Adult literacy rate (% of pop. aged 15+)	0.82	0.74
Primary school enrollment (% of children)	0.58	0.58
Lower secondary school enrollment (% of children)	0.82	0.67
Upper secondary school enrollment (% of children)	0.88	0.75
Gender parity in secondary enrollment (girls/boys)	0.56	0.40
Mobile telephone subscriptions (subscriptions/100 people)	0.68	0.55
Internet users (% of pop.)	0.92	0.81
Press Freedom Index (0=most free; 100=least free)	-0.46	-0.27
Life expectancy (years)	0.89	0.84
Non-communicable disease deaths between the ages of 30 and 70 (probability of dying)	-0.72	-0.60
Obesity rate (% of pop.)	0.65	0.44
Outdoor air pollution attributable deaths (deaths/100,000)	0.19	0.32
Suicide rate (deaths/100,000)	0.19	0.29
Greenhouse gas emissions (CO2 equivalents per GDP)	-0.36	-0.29
Water withdrawals as a percent of resources	0.26	0.18
Biodiversity and habitat (0=no protection; 100=high protection)	0.25	0.19
Political rights (1=full rights; 7=no rights)	-0.63	-0.48
Freedom of speech (0=low; 2=high)	0.33	0.19
Freedom of assembly/association (0=low; 2=high)	0.45	0.27
Freedom of movement (0=low; 4=high)	0.38	0.27
Private property rights (0=none; 100=full)	0.72	0.60
Freedom over life choices (% satisfied)	0.31	0.16
Freedom of religion (1=low; 4=high)	0.14	-0.04
Modern slavery, human trafficking and child marriage (1=low; 100=high)	-0.61	-0.50
Satisfied demand for contraception (% of women)	0.79	0.73
Corruption (0=high; 100=low)	0.76	0.62
Women treated with respect (0=low; 100=high)	0.15	0.18
Tolerance for immigrants (0=low; 100=high)	0.15	-0.05
Tolerance for homosexuals (0=low; 100=high)	0.68	0.55
Discrimination and violence against minorities (0=low; 10=high)	-0.59	-0.42
Religious tolerance (1=low; 4=high)	0.12	0.00
Community safety net (0=low; 100=high)	0.68	0.51
Years of tertiary schooling	0.65	0.67
Women's average years in school	0.87	0.75
Inequality in the attainment of education (0=low; 1=high)	-0.80	-0.73
Number of globally ranked universities (0=none; 5= >50)	0.61	0.61

Table 4: Social Progress Index dataset loadings from PCA.

Loadings for first factor in Democracy Ranking dataset		
Component	*Loading*	*r x IQ*
Political rights (aggregated scores): Freedom House	0.67	0.43
Civil liberties (aggregated scores): Freedom House	0.74	0.53
Global Gender Gap Report	0.56	0.39
Press Freedom: Freedom House	0.67	0.44
Corruption Perceptions Index (CPI): Transparency International (TI)	0.79	0.64
Change(s) of the head of government (last 13 years, peaceful)	0.23	0.20
Political party change(s) of the head of government (last 13 years, peaceful)	0.39	0.30
GDP per capita, PPP (constant 2005 international $)	0.87	0.72
GDP per capita, PPP (current international $)	0.87	0.72
Central government debt, total (% of GDP)	0.56	0.51
Inflation, consumer prices (annual %)	0.58	0.49
Unemployment, total (% of total labor force)	0.55	0.55
Unemployment, youth total (% of total labor force ages 15-24)	0.59	0.51
CO2 emissions (kg per 2005 PPP $ of GDP)	-0.17	-0.26
CO2 emissions (metric tons per capita)	-0.59	-0.49
GDP per unit of energy use (constant 2005 PPP $ per kg of oil equivalent)	-0.48	-0.50
Electric power consumption (kWh per capita)	-0.70	-0.58
Electricity production from hydroelectric sources (% of total)	-0.53	-0.54
Labor force, female (% of total labor force)	0.11	0.02
Unemployment, female (% of female labor force)	0.53	0.47
Primary education, pupils (% female)	0.31	0.15
School enrollment, secondary, female (% gross)	0.87	0.76
School enrollment, secondary, female (% net)	0.87	0.73
School enrollment, tertiary, female (% gross)	0.78	0.66
Life expectancy at birth, female (years)	0.86	0.85
Life expectancy at birth, total (years)	0.86	0.84
Health expenditure per capita, PPP (constant 2005 international $)	0.82	0.67
Health expenditure, public (% of GDP)	0.66	0.47
Health expenditure, private (% of GDP)	-0.21	-0.28
Hospital beds (per 1,000 people)	0.52	0.58
Physicians (per 1,000 people)	0.68	0.64
Mortality rate, infant (per 1,000 live births)	0.86	0.82
Mortality rate, under-5 (per 1,000 live births)	0.83	0.81
School enrollment, secondary (% gross)	0.89	0.78
School enrollment, secondary (% net)	0.84	0.74
School enrollment, tertiary (% gross)	0.90	0.83
Pupil-teacher ratio, primary	0.59	0.52
Telephone lines (per 100 people)	0.88	0.81
Internet users (per 100 people)	0.92	0.83
Mobile cellular subscriptions (per 100 people)	0.71	0.61
Research and development expenditure (% of GDP)	0.61	0.52
Scientific and technical journal articles (per 1,000 people)	0.78	0.66

Table 5: Democracy Ranking dataset loadings from PCA.