Research Article

# National IQs and Socioeconomic Development

## Sebastian Jensen*, Emil OW Kirkegaard

*Department of General Surgery, Ulster Institute of Social Research, Bristol, United Kingdom*

## ABSTRACT

Using 47 indicators of socioeconomic development and various sources of performance on cognitive tests, we constructed the SDI (Socioeconomic Development Index) and a set of national IQs for 197 nations, the latter using no geographic imputations. Combining the various datasets reduced the estimated standard error of national IQs from 5.41 to 2.58, and a strong correlation between socioeconomic development and national IQs was observed (r=.88).

Based on the prior that Flynn effect gains do not pass measurement invariance, IQ scores should exhibit some non-negligible bias between countries. Empirical assessments of measurement invariance across nations finds that measurement invariance violations are uncommon, and are more prevalent in verbal than nonverbal tests. In most countries, national IQs show high levels of reliability and validity, and we encourage their use in the literature.

Keywords: GDP; Economic development; IQ; Flynn effect

## INTRODUCTION

Differences in economic development between countries have traditionally been quantified using GDP (Gross Domestic Product) per capita, introduced in 1937 by Simon Kuznets to capture all economic production [1]. This measurement was popularized in 1944 after the Bretton Woods conference and has become a commonly used measurement of economic development. This measurement has faced various criticisms: The most notable one being that GDP does not take into account income earned abroad, leading some economists to advocate for using GNI (Gross National Income) instead. In addition, socioeconomic development (socioeconomic development) extends beyond economic output-other variables such as mortality, educational attainment, safety, and institutional quality must be taken into consideration. Consequently, researchers developed composite indices such as the HDI (Human Development Index) and the SPI (Social

Progress Index) which use multiple indicators to construct a general index.

Both of these indexes, while useful, have their respective issues. The HDI only uses three indicators GDP, educational attainment, and life expectancy to calculate socioeconomic development, which leads to some non-negligible unreliability (ω =.93, when using GNI per capita, life expectancy, expected years of schooling, and mean years of schooling). The SPI reduces the influence of unreliability by using 50 indicators to calculate socioeconomic development, which is better, but many of these variables may suffer from non-invariance (bias) across cultures, notably indicators of sexual inequality, democracy, corruption, and freedom, which assume that current Western values are the best in a kind of "the end of history" approach [2]. While these values may be desirable or lead to higher levels of socioeconomic development, using more objective indicators of socioeconomic development (e.g. internet speed, median income) would be best to avoid the

problem of cultural bias. There is also the question of scoring: most indices of socioeconomic development use arbitrary weighting methods, like the HDI, which changed to a geometric mean method in 2011 which shifted the rank order a bit [3].

Similar to socioeconomic development, there is an issue with measuring human capital. An example of an early adopter of comparing test scores between different nations was Barbara Lerner, who compared the performance of Western Europe, the United States, and Japan in test performance and hypothesized that it was related to economic development.

Richard Lynn later collected IQ test scores from various countries, and found that national IQs and GDP per capita correlated at .82, though this dataset and other revisions of it have been extensively criticized in the literature. Some economists have made indexes of human capital based on child mortality, test scores, and educational attainment, but it could be argued that child mortality and education are a function of both human capital and socioeconomic development, making it an improper measurement.

The purpose of this study is to use state-of-the-art statistical and machine learning techniques to create the most accurate measurements of socioeconomic development and human capital that can be made. Theoretically, socioeconomic development should affect human capital due to the fact that socioeconomic development causes nations to have better nutrition and health, and societies with higher levels of human capital should create societies with higher levels of socioeconomic development. Other researchers reported strong correlations between indicators of socioeconomic development (e.g. GDP per capita) and human capital (r=.6-. 8) though these values are based on the national IQ datasets which have been unpopular in the literature.

## Data

Data on most national development indicators were sourced from the Social Progress Index [4]. When possible, averages of variables from 2018-2022 were calculated to reduce the unreliability that comes from year-to-year fluctuations. Indicators of national development that are manipulable (e.g. indexes), contain less than 100 observations, or measure a value that is sensitive to national and cultural differences (e.g. gender equality, measurements of freedom) were not considered. An exception was made for the Legatum health index, which was perceived to be of high quality. Other variables were downloaded on the internet from various sources, which have been cited in **Table 1**.

**Table 1:** Sources of variables of national differences.

| Variable | Number of Countries | Time range |
|---|---|---|
| National IQs (unweighted, psychometric) | 130 | 1945-2017 |
| National IQs (sample weighted, psychometric) | 129 | 1945-2017 |
| National IQs (quality weighted, psychometric) | 130 | 1945-2017 |
| National IQs (scholastic) | 102 | 1945-2017 |
| National IQs (composite) | 148 | 1945-2017 |
| National IQs (composite) | 81 | Varying |
| National IQs (composite) | 133 | Varying |
| National IQs (composite) | 170 | Varying |
| Recent test scores (PISA, TIMSS, PIRLS) | 39-81 | 2019-2022 |
| Test scores (Basic skills Dataset, BSD) | 126 | Varying |
| Test scores (World bank test scores, WBTS) | 174 | Varying |
| Average IQs of different countries | 7 | Varying |
| % of population in agriculture ind. | 185 | 2018-2022 |
| Caloric intake | 168 | 2018 |
| Car exports ($) | 177 | 2022 |
| Circuit exports ($) | 132 | 2022 |
| DOI Entries per country | 131 | Varying |

| | | |
|---|---|---|
| Health index | 166 | 2023 |
| Information technology exports (%) | 159 | 2018-2022 |
| Internet speed by country (mobile) | 137 | |
| Internet speed by country (broadband) | 173 | 2023-2024 |
| Median income | 159 | 2006-2021 |
| Median wealth | 161 | 2017-2021 |
| Tech exports ($) | 163 | 2018-2022 |
| Interpersonal violence | 194 | 2018-2022 |
| GNI per capita PPP adjusted | 196 | 2018-2022 |
| GNI per capita PPP adjusted | 196 | 2018-2022 |
| Child stunting | 188 | 2018-2022 |
| Intimate partner violence | 194 | 2018-2022 |
| Years lost due to infections | 194 | 2018-2022 |
| Undernourishment | 168 | 2018-2022 |
| Child mortality | 194 | 2018-2022 |
| Maternal mortality | 184 | 2018-2022 |
| Mortality due to water quality | 188 | 2018-2022 |
| Water satisfaction | 151 | 2018-2022 |
| Water sanitation | 192 | 2018-2022 |
| Water access (%) | 191 | 2018-2022 |
| Household pollution | 194 | 2018-2022 |
| Electricity usage | 194 | 2018-2022 |
| Clean fuel usage | 189 | 2018-2022 |
| Money stolen (% of pop.) | 149 | 2018-2022 |
| Percent that say it is safe to walk alone at night | 150 | 2018-2022 |
| Transportation injuries | 188 | 2018-2022 |
| Proportion with no education | 194 | 2018-2022 |
| Primary school enrollment | 167 | 2018-2022 |
| Proportion with secondary education | 176 | 2018-2022 |
| Mobile phones per person | 194 | 2018-2022 |
| Internet access | 192 | 2018-2022 |
| Mortality from ages 15 to 50 | 196 | 2018-2022 |
| Matter pollution | 188 | 2018-2022 |
| Air pollution | 194 | 2018-2022 |

| | | |
|---|---|---|
| Particulate matter exposure | 194 | 2018-2022 |
| Percent that are NEETS | 180 | 2018-2022 |
| Citable documents per capita | 195 | 2018-2022 |
| University rankings (population controlled) | 130 | 2018-2022 |
| Percent satisfied with health care | 150 | 2018-2022 |
| Percent who say they have friends and family to count on | 151 | 2018-2022 |
| Life expectancy | 193 | 2018-2022 |
| Expected years of tertiary schooling | 142 | 2018-2022 |
| GNI per capita | 191 | 2018-2022 |
| GDP per capita (composite) | 195 | 2018-2022 |

## MATERIALS AND METHODS

### Estimating Socioeconomic Development

Missing values from the socioeconomic development indicators were imputed with multiple imputation by chained equations (m=100), with a prediction threshold of r=0.4, as many indicators are highly correlated with each other. This was reduced to 0.3 in the untransformed data, as the untransformed data was less inter-correlated than the transformed data. Countries that had more than 45% of their data missing in socioeconomic indicators (Bahamas, Palestine, Kosovo, Liechtenstein, Monaco, Greenland, Nauru, Tuvalu, Palau, Saint Kitts and Kevis, San Marino, Macao, Puerto Rico, Palau, and Hong Kong) had their Social Development Index (SDI) calculated using a different method. For these countries, factor scores based on the variables that were not missing

were calculated and then their rank relative to the sample was calculated. That rank was then regressed to the mean depending on the omega reliability of the estimate, which was lowest for Greenland at .90 and highest for Saint Kitts and Kevis at .98. Due to its implausibility, the estimate for North Korea (SDI=.98, which would make it the 47th most developed country in the world), was removed from the dataset, as it's inconsistent with its very low GDP per capita ($1,500).

Variables were grouped into various categories depending on what they were measuring conceptually to compute specific scores, as displayed in **Table 2**. Principal component analysis was used to extract factor scores in all cases, so if a variable needed to be reverse coded, the algorithm would apply this correction automatically.

**Table 2:** Method used to calculate specific scores.

| Indicator name | Indicators | Cronbach's alpha | Omega total |
|---|---|---|---|
| Economic development index | GNI per capita, GDP per capita, median income, median wealth | 0.97 | 0.977 |
| Technological development index | Broadband speed, mobile internet speed, agriculture (%), mobiles per capita, internet (%), tech exports per capita ($), car exports per capita ($), circuit exports per capita ($), ICT share of GDP (%), electricity (%) | 0.865 | 0.91 |
| Educational attainment index | NEETs (%), no education (%), primary enrollment (%), secondary degree att. (%), expected yrs. of tertiary ed., unit rank controlled for pop., citable docs per cap., DOI res. per cap. | 0.917 | 0.939 |
| Index of mortality | Child mortality, maternal mortality, mortality yrs. 15-50 | 0.937 | 0.938 |

| | | | |
|---|---|---|---|
| Infrastructure development index | Infections daily, satisfaction with healthcare, health index, life expectancy, water satisfaction, water sanitation, mortality due to water qual., water use (%), child stunting, under-nourishment, caloric intake. | 0.961 | 0.973 |
| Index of pollution | Air pollution, household pollution, use of clean fuels, lead exposure | 0.762 | 0.895 |
| Safety index | % who say they had money stolen, % who say it's safe to walk alone at night, intimate partner violence, whether friends/family can be counted on, transport quality, years lost due to interpersonal violence | 0.86 | 0.904 |
| First principal component | Composite of sub-indicators | 0.961 | 0.972 |

After this, general scores of socioeconomic development were computed in different ways which involved combining several different methodological variations. This was done to produce results that are less sensitive to changes in methodology. These methodological variations include:

- Computing general development scores by iterating through each of the 48 variables and randomly selecting four independent variables that predict another random variable using restricted cubic splines. This process is repeated 5000 times per variable. Then, these predictions are averaged and the first principal component of those averages is taken. This admittedly is a very unusual method (which will be called "spline iteration" from now on), but this avoids non-linear biases and maximizes the influence of the most reliable and valid variables.

- The above method is repeated, but using a support vector machine that uses regression and a radial kernel to predict the dependent variable from four randomly selected independent variables. This method will be called the "SVM iteration" method.

- The first principal component of the indicators is extracted. This will be called the "simple component" method.

- Four components from the data are extracted, obliquely rotated, and the first principal component is extracted. This will be called the "complex component" method.

- Both of the above can be calculated by extracting the first principal component of the 47 indicators or the 7 sub-indicators.

- Applying the transformations (logarithmic, square root, reciprocal, or squared) that maximize the correlation between the variable and socioeconomic development.

16 different combinations of these methodological decisions were calculated and averaged to form the Socioeconomic Development Index (SDI). Eight possible combinations of these methods are missing, as making estimates based on the 7 sub-indicators *vs.* 47 indicators for the machine learning derived

estimates was judged as superfluous. On average, scores from these 16 methods correlated at .99, with inter-correlations ranging from .94 to .9999. This index of socioeconomic development was consistent with other measurements of development ($r=.97$ with the social progress index, $r=.98$ with HDI).

If a variable exhibited a strongly nonlinear relationship with HDI, where variance at extremes no longer predicted HDI, then values in the unpredictive range were winsorized. This was also done when one variable had large outliers (e.g. some countries produce orders of magnitude more semiconductors than others). In the case of mobile Internet speed, the maximum speed was set to 100 Mbps, as the relationship was nonlinear and the variable contained several outliers, as shown in **Figure 1**. This avoids specific variance from biasing the estimates of the general socioeconomic development, as if a country is an outlier in general development, that status should theoretically be reflected in all of its development indicators.
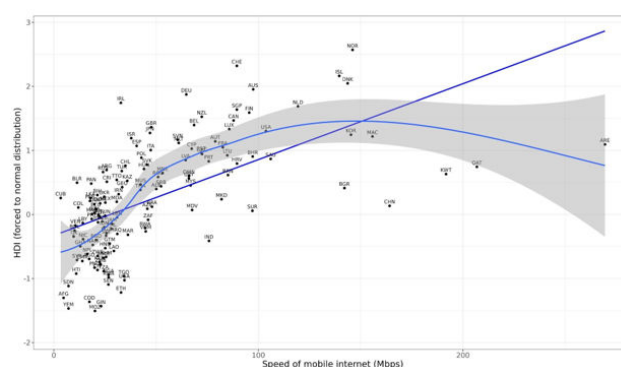


**Figure 1:** Relationship between the speed of mobile internet and HDI (forced to a normal distribution).

## Criticisms of National IQs

Sear has criticized the use of national IQs (2022), primarily the Lynn and Becker datasets for several reasons. Among these

criticisms is the use of children to estimate the average IQs of nations, as IQ scores depend on age. However, the scores on these tests are standardized by age, which makes this concern irrelevant. This can be a concern if the magnitude of group differences varies by age, but the best evidence available suggests that is usually not the case, at least not between American Blacks and Whites [5]. The same is true for Asians and Whites, where Asians score above Whites as children and adults [6]. There are exceptions, such as the Arab ~ European IQ difference, where the difference increases with age [7].

Sear also questions whether the figures that are estimated for the African countries are believable, as many of them fall in the 65 to 75 range, which is close to the conventional cutoff for intellectual disability. This ignores that not all causes and types of mental disability are the same some of them are mild and typically caused by additive genetic variance, these intellectually disabled people generally can live normal lives others are caused by severe mutations or deletions, which cause deficits in other areas of biological functioning. Arthur Jensen was initially drawn to IQ research because he noticed that black and white children in the classes for the mentally disabled behaved quite differently in the playground, the black children behaving normally, but the White being socially dysfunctional. The explanation for this pattern was that a large fraction of the white children suffered from major genetic disorders such as down's syndrome, or perinatal environmental damage (syndromic disability), while the black children were merely on the left side of their normal distribution, thus had mostly ordinary causes (familial disability). Since the syndromic causes of mental disability usually cause other deficits beyond low intelligence, this explains the large difference in the social skills of the two groups of children.

A more intuitive comparison would be differences in height between African Pygmies and those from the Dinaric Alps. On average, Pygmy men are about 153 cm tall and Dinaric men are about 186 cm tall a difference of roughly five entire standard deviations relative to the standard deviation of Dinaric male height (6.5 cm). The conventional cutoff for dwarfism in Western nations is 150 cm; within the Pygmies, roughly half of their men would fall below this cutoff, in the Dinaric Alps, only men who suffer from a genetic disorder such as achondroplasia, metatropic dysplasia, or growth hormone deficiency could be this short. The fact that Dinarics who are under 150 cm tall tend to suffer from additional complications that are not observed in Pygmies is not evidence that height measurements are biased against the latter group; merely that height differences must be understood as originating from a variety of genetic and environmental causes, which can have effects on various phenotypes.

It is doubtful that an IQ score of 70 for an African and a European means the same thing in terms of biological functioning, though these scores accurately reflect their ability to take cognitive tests, as Africans tend to score the equivalent of an IQ of 70 on scholastic tests administered by the TIMSS [8]. Whether these test scores function as biased

estimates of intelligence is debatable. Theoretically, some biases will deflate the African IQ relative to what would be expected from their true average levels of intelligence (low effort test takers, Flynn Effect related measurement variance, illiterates), and others will inflate it (use of primary/secondary school students which are less nationally representative in more uneducated countries, use of the standard deviation between groups instead of within groups, use of subtest differences instead of full scale differences).

Flynn Effect related measurement invariance is concerning, as the literature overwhelmingly converges towards Flynn effects being partially caused by test bias in favour of newer cohorts [9-13]. As nations differ in the rate at which they undergo Flynn Effects, this may cause the test scores to be biased in favour of certain countries. Some of the Flynn Effect gains are still plausibly real: Brain sizes increased by about 0.7 SD between the 1930's and 70's, if this effect occurred between 1900 and 1970, then the expected increase in brain size would be 1.2 SD. Given that brain size and IQ correlate at roughly .28 and this correlation is causal from brain size to intelligence intelligence would have been expected to increase by 5 points due to this increase; assuming it is absolute and not relative brain size that is linked to IQ.

There have been some studies on whether international scholastic tests satisfy measurement invariance. There are traditionally four steps taken to test measurement invariance: Configural invariance (whether the items load on the same factors between groups), metric invariance (whether the magnitude of the factor loadings on the constructs differs between groups), scalar invariance (whether the magnitude of the intercepts of the items differs between groups), and residual invariance (whether the residual variance of the items is the same between groups) [13]. For comparing national means, scalar invariance is the most important test of measurement invariance that needs to be satisfied.

Contrary to priors, scores on cognitive tests do not exhibit large violations of measurement invariance, especially if the test involved is nonverbal. Strict measurement invariance was held within Anglo and East Asian cultural groups on the 1999 TIMSS tests, though only weak (metric, but not scalar) measurement invariance was held between the cultural groups, as shown in **Figure 2**. Their methodology is limited by the fact measurement invariance was assessed at the factor level, as groups are likely to differ in general and specific ability it would be better to assess measurement invariance at the item level.

The vast majority of the items on the 2015 PISA math and science tests passed measurement invariance in both the factor loadings and intercepts, suggesting test bias was not an issue in administration. Another study of international test bias of the PISA item data on the reading subtest found that scalar invariance was violated in most nations, with the magnitude of invariance ranging from 0.041 in Canada to 0.93 in Kyrgyzstan [14]. The presence of biased items, however, does not imply that the means are biased between groups, as the direction of the effects tends to vary at the item level [15].

|      | AUS | NZL | CAN | USA | TWN | KOR |
|------|-----|-----|-----|-----|-----|-----|
| NZL  | **Strict** |     |     |     |     |     |
| CAN  | **Strict** | **Strict** |     |     |     |     |
| USA  | **Weak** | **Strict** | **Strict** |     |     |     |
| TWN  | Weak | Weak | Weak | Weak |     |     |
| KOR  | Weak | Weak | Weak | Weak | **Strict** |     |
| JPN  | Weak | Configure | Weak | Configure | **Weak** | **Strict** |

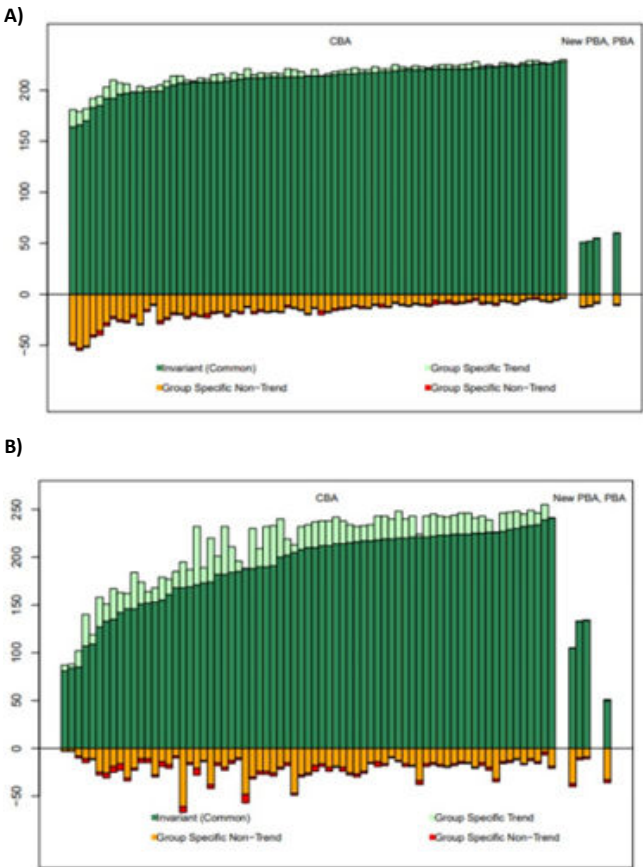**Figure 2:** Results of measurement invariance testing from Wu, et al.

The most exhaustive and recent assessment of measurement invariance between nations is an assessment that is available in the PISA 2022 technical report. They concluded that measurement invariance is a major issue for the financial literacy test, somewhat of an issue for the science and reading tests, and a minor issue for the mathematics test. **Figures 3 and 4** show the distribution of variant (orange/red/light green) and invariant (dark green) items by country and test.

**A)**

**B)**
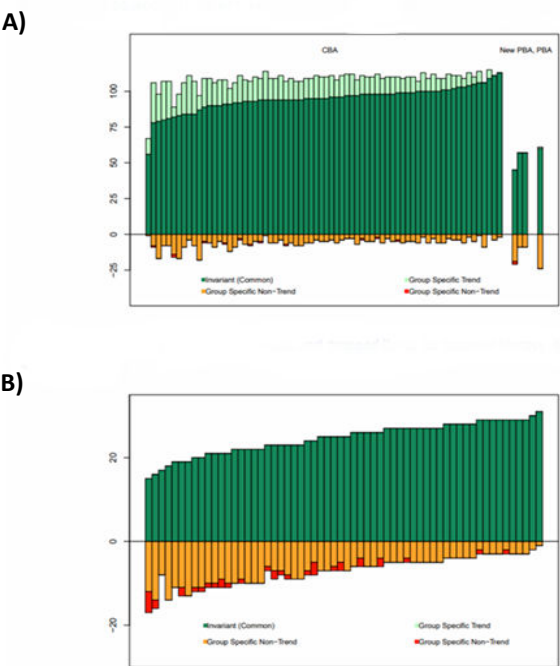
**Figure 3:** Results of the measurement invariance testing at the item level for the science and financial literacy test by country (taken from PISA, 2022). A) Frequency of invariant, variant and dropped items for science, by country/economy. B) Frequency of invariant, variant and dropped items for financial literacy, by country/economy.

**A)**

**B)**

**Figure 4:** Results of the measurement invariance testing at the item level for the mathematics and reading test by country (taken from PISA, 2022). A) Frequency of invariant, variant and dropped items for mathematics, by country/economy. B) Frequency of invariant, variant and dropped items for reading, by country/economy.

In practice, the differences between countries on PISA scores are extremely highly correlated and of roughly equal magnitude, as shown in **Table 3**. Therefore, it must be concluded that minor violations of measurement invariance on the PISA exams, and likely all scholastic tests, do not have a practically significant impact.

**Table 3:** Average score on the PISA (2022) exam by country and subtest. Taken from Recueil (2023) and Wikipedia (2024b).

| Country | Maths | Country | Science | Country | Reading |
|---------|-------|---------|---------|---------|---------|
| Singapore | 575 | Singapore | 561 | Singapore | 543 |
| Macau | 552 | Japan | 547 | Ireland | 516 |
| Chinese Taipei | 547 | Macau | 543 | Japan | 516 |
| Hong Kong | 540 | Chinese Taipei | 537 | South Korea | 515 |

| | | | | | |
|---|---|---|---|---|---|
| Japan | 536 | South Korea | 528 | Chinese Taipei | 515 |
| South Korea | 527 | Estonia | 526 | Estonia | 511 |
| Estonia | 510 | Hong Kong | 520 | Macau | 510 |
| Switzerland | 508 | Canada | 515 | Canada | 507 |
| Canada | 497 | Finland | 511 | United States | 504 |
| Netherlands | 493 | Australia | 507 | New Zealand | 501 |
| Ireland | 492 | Ireland | 504 | Hong Kong | 500 |
| Belgium | 489 | New Zealand | 504 | Australia | 498 |
| Denmark | 489 | Switzerland | 503 | United Kingdom | 494 |
| United Kingdom | 489 | Slovenia | 500 | Finland | 490 |
| Poland | 489 | United Kingdom | 500 | Denmark | 489 |
| Australia | 487 | United States | 499 | Poland | 489 |
| Austria | 487 | Poland | 499 | Czech Republic | 489 |
| Czech Republic | 487 | Czech Republic | 498 | Sweden | 487 |
| Slovenia | 485 | Denmark | 494 | Switzerland | 483 |
| Finland | 484 | Latvia | 494 | Italy | 482 |
| Latvia | 483 | Sweden | 494 | Germany | 480 |
| Sweden | 482 | Germany | 492 | Austria | 480 |
| New Zealand | 479 | Austria | 491 | Belgium | 479 |
| Germany | 475 | Belgium | 491 | Norway | 477 |
| Lithuania | 475 | Netherlands | 488 | Portugal | 477 |
| France | 474 | France | 487 | Croatia | 475 |
| Spain | 473 | Hungary | 486 | Latvia | 475 |
| Hungary | 473 | Spain | 485 | Spain | 474 |
| Portugal | 472 | Lithuania | 484 | France | 474 |
| Italy | 471 | Portugal | 484 | Israel | 474 |
| Vietnam | 469 | Croatia | 483 | Hungary | 473 |
| Norway | 468 | Norway | 478 | Lithuania | 472 |
| Malta | 466 | Italy | 477 | Slovenia | 469 |
| United States | 465 | Turkey | 476 | Vietnam | 462 |
| Slovakia | 464 | Vietnam | 472 | Netherlands | 459 |
| Croatia | 463 | Malta | 466 | Turkey | 456 |
| Iceland | 459 | Israel | 465 | Chile | 448 |
| Israel | 458 | Slovakia | 462 | Slovakia | 447 |

| Country | | Country | | Country | |
|---|---|---|---|---|---|
| Turkey | 453 | Ukraine | 450 | Malta | 445 |
| Brunei | 442 | Iceland | 447 | Serbia | 440 |
| Ukraine | 441 | Serbia | 447 | Greece | 438 |
| Serbia | 440 | Brunei | 446 | Iceland | 436 |
| UAE | 431 | Chile | 444 | Uruguay | 430 |
| Greece | 430 | Greece | 441 | Brunei | 429 |
| Romania | 428 | Uruguay | 435 | Romania | 428 |
| Kazakhstan | 425 | UAE | 432 | Ukraine | 428 |
| Mongolia | 425 | Qatar | 432 | Qatar | 419 |
| Cyprus | 418 | Romania | 428 | UAE | 417 |
| Bulgaria | 417 | Kazakhstan | 423 | Costa Rica | 415 |
| Moldova | 417 | Bulgaria | 421 | Mexico | 415 |
| Qatar | 414 | Moldova | 417 | Moldova | 411 |
| Chile | 412 | Malaysia | 416 | Brazil | 410 |
| Uruguay | 409 | Mongolia | 412 | Jamaica | 410 |
| Malaysia | 409 | Cyprus | 411 | Colombia | 409 |
| Montenegro | 406 | Colombia | 411 | Peru | 408 |
| Azerbaijan | 397 | Costa Rica | 411 | Montenegro | 405 |
| Mexico | 395 | Mexico | 410 | Bulgaria | 404 |
| Thailand | 394 | Thailand | 409 | Argentina | 401 |
| Peru | 391 | Peru | 408 | Panama | 392 |
| Georgia | 390 | Argentina | 406 | Malaysia | 388 |
| North Macedonia | 389 | Brazil | 403 | Kazakhstan | 386 |
| Saudi Arabia | 389 | Jamaica | 403 | Saudi Arabia | 383 |
| Costa Rica | 385 | Montenegro | 403 | Cyprus | 381 |
| Colombia | 383 | Saudi Arabia | 390 | Thailand | 379 |
| Brazil | 379 | Panama | 388 | Mongolia | 378 |
| Argentina | 378 | Georgia | 384 | Georgia | 374 |
| Jamaica | 377 | Indonesia | 383 | Guatemala | 374 |
| Albania | 368 | Azerbaijan | 380 | Paraguay | 373 |
| Indonesia | 366 | North Macedonia | 380 | Azerbaijan | 365 |
| Palestine | 366 | Albania | 376 | El Salvador | 365 |
| Morocco | 365 | Jordan | 375 | Indonesia | 359 |
| Uzbekistan | 364 | El Salvador | 374 | North Macedonia | 359 |

| | | | | | |
|---|---|---|---|---|---|
| Jordan | 361 | Guatemala | 373 | Albania | 358 |
| Panama | 357 | Palestine | 369 | Dominican Republic | 351 |
| Kosovo | 355 | Paraguay | 368 | Palestine | 349 |
| Philippines | 355 | Morocco | 365 | Philippines | 347 |
| Guatemala | 344 | Dominican Republic | 360 | Jordan | 342 |
| El Salvador | 343 | Kosovo | 357 | Kosovo | 342 |
| Dominican Republic | 339 | Philippines | 356 | Morocco | 339 |
| Paraguay | 338 | Uzbekistan | 355 | Uzbekistan | 336 |
| Cambodia | 336 | Cambodia | 347 | Cambodia | 329 |

Some researchers have argued that the samples of Africans who took the Raven's test collected by Lynn have low levels of convergent validity and are taken from unrepresentative samples [16]. The low scores of Africans (70) on these tests cannot be blamed on selective sampling or reporting, as the average African IQ converges to an average of roughly 70 regardless of the source including sources that rely solely on results from scholastic assessments. The evidence Wicherts, et al. presented regarding IQ scores of Africans having lower levels of validity than Europeans was convincing, but not necessarily indicative of an upward or downward bias.

The expected African IQ can be estimated based on several parameters, including the average IQ of Blacks, the percentage of the difference between Blacks and Whites that is due to additive genetics, the percentage of admixture in Blacks that is European (20%), and the extent to which the environment of Sub-Saharan Africa depresses IQ scores. For example, if the between-group heritability of IQ between African Americans and White Americans is 100%, and the difference between them is 18 points, and the environment of Africa depresses IQ scores by 10 points, then the expected Sub-Saharan African IQ is 67.5 (67.5=(82-.2 × 100)/.8-10).

If the expected African IQ differs greatly from the observed one, then this difference is likely to be due to test bias or incorrect assumptions. To test whether this was the case, the expected Sub-Saharan African IQ was estimated based on a range of possible parameters. The range of the American Black IQ was assumed to be between 80-90, for the between-group heritability it was assumed to be 0-100%, and the extent to which the environment of Africa depresses Black IQs was assumed to be between 0 to 20 points. Using these parameter ranges, the expected IQ of Sub-Saharan Africa could be anywhere from 55 to 100, as shown in **Figure 5**.
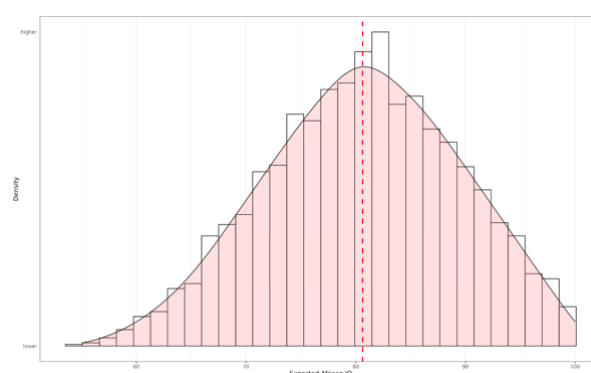


**Figure 5.** Density plot of possible Sub-Saharan African IQs according to the possible range of parameters that was chosen.

There is fairly robust evidence, from military-based randomization studies and latent modeling that education improves IQ scores, though this improvement does not translate to greater general intelligence (e.g. increases in accumulated knowledge, but not reaction time). If this conclusion is accepted, then it must be the case that differences in IQ between nations that are due to differences in educational attainment must lead to bias in favour of the more educated countries. Besides this, there is quantitative evidence summarized by Warne which indicates that unschooled populations in Central Asia do not reason about problems on IQ tests the same way Westerners do: when asked which of a set of four objects do not fit together (e.g. an axe, saw, hammer, and log), they will typically choose one of the tools, as not much can be done without three tools and no object to operate with [17].

This bias in testing that occurs due to some populations being uneducated can be tested by comparing results from psychometric testing (IQ tests) and those based on scholastic tests (e.g. PIRLS, PISA, TIMSS tests). While the quality of education varies by country, students who take scholastic tests are active in educational institutions, which should reduce the bias that results from unschooling. In terms of regional differences, scores on psychometric and scholastic tests are highly correlated regionally (r=.97); the only prominent outliers being the East Asians and Central Asians

who score about 4 to 5 points higher on psychometric tests in comparison to scholastic tests, as shown in **Table 4**. This indicates that differences in educational attainment between countries are not a practically significant source of bias when estimating the average levels of intelligence between regions, as matching populations for years of schooling does not change the average differences. Given that most scholastic tests do not show large violations of measurement invariance, it would be appropriate to conclude that the IQ tests do not show large biases against undeveloped nations.

**Table 4:** Estimated regional IQ by dataset.

| Region | BSD | WBTS | RSAS | BSAS | BQNW | BNW | BUW | SCH | PSY |
|---|---|---|---|---|---|---|---|---|---|
| Eastern Asia | 101.76 | 98.89 | 97.51 | 100.63 | 103.37 | 103.27 | 105.81 | 99.7 | 104.15 |
| Northern America | 99.18 | 100.75 | 98.76 | 99.23 | 95.55 | 95.62 | 93.84 | 99.48 | 95 |
| Western Europe | 99.16 | 99.16 | 98.12 | 98.68 | 100.23 | 99.83 | 101.68 | 98.78 | 100.58 |
| Northern Europe | 98.76 | 99.8 | 97.86 | 98.33 | 96.98 | 96.72 | 97.61 | 98.69 | 97.1 |
| Australia and New Zealand | 98.68 | 100.25 | 98.26 | 97.71 | 100.07 | 100.03 | 100.33 | 98.72 | 100.14 |
| Eastern Europe | 93.76 | 94.95 | 93.26 | 94.98 | 93.24 | 93.18 | 95.22 | 94.24 | 93.88 |
| Southern Europe | 90.8 | 91.55 | 90.01 | 90.66 | 91.6 | 91.52 | 91.93 | 90.75 | 91.68 |
| South-eastern Asia | 88.11 | 87.42 | 85.76 | 88.61 | 89.1 | 88.98 | 87.24 | 87.47 | 88.44 |
| Western Asia | 86.31 | 85.03 | 79.32 | 79.69 | 83.28 | 83.15 | 84.97 | 82.59 | 83.8 |
| Latin America/ Caribbean | 82.48 | 82.01 | 75.41 | 78.18 | 81.29 | 80.99 | 81.51 | 79.52 | 81.26 |
| Central Asia | 79.32 | 88.93 | 78.76 | 81.52 | 86.98 | 86.98 | 89.29 | 82.13 | 87.75 |
| Northern Africa | 79.19 | 78.21 | 75.51 | 72.09 | 78.21 | 78.17 | 78.27 | 76.25 | 78.22 |
| Southern Asia | 74.12 | 78.54 | 74.26 | 76.62 | 76.44 | 76.33 | 78.22 | 75.88 | 76.99 |
| Sub-Saharan Africa | 70.32 | 77.71 | 65.93 | 66.54 | 69.6 | 69.51 | 70.3 | 70.12 | 69.8 |

**Note:** BSD: Basic Skills Dataset, WBTS: World Bank Test Scores, RSAS: Rindermann's Scholastic Estimates, BSAS: Becker's Scholastic Estimates, BQNW: Becker's Quality Weighted Psychometric Estimates, BNW: Becker's sample size Weighted estimates, BUW: Becker's Unweighted Estimates, SCH: average of the scholastic estimates (BSD, WBTS, RSAS, BSAS), PSY: average of the psychometric estimates (BNW, BUW, BQNW).

It's worth mentioning that most researchers, including Becker and Rindermann, used scholastic estimates of ability derived from international tests to estimate the intelligence of nations. These data sources are immune to many of the biases that plague the estimates that are based on convenience samples: They tend to test about a thousand students per country, the samples are roughly representative of the student body of the country, and the same test is administered to all countries at roughly the same time. Within individuals, scores on IQ tests and scholastic ability tests correlate positively and differences in IQ between nations correlate highly with scholastic estimates, such as those from the basic skills dataset (r=.82), as shown in **Figure 6**. If it is the case that these scholastic estimates correspond closely with the psychometric ones between nations, then that suggests that the psychometric data is not of low quality.
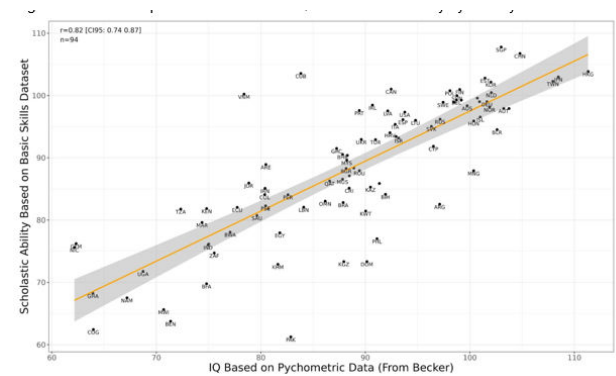


**Figure 6:** Relationship between measured IQ and scholastic ability by country.

The relationship between IQ based on psychometric data and scholastic estimates also holds within regions, although the relationship attenuated (r=.41, weighted by sample size), as shown in **Table 5**. This indicates that this correlation is not a function of regions being assigned systematically lower or higher values by the data sources, rather that nations differ in ability, and these differences are reflected in test performance.

**Table 5:** Correlation between Becker's unweighted estimates of IQ and the world bank test score results by region. World bank test scores were used over the basic skills dataset because the world bank dataset measured more nations.

| Region | Correlation | Sample size |
| --- | --- | --- |
| Central Asia | 0.97 | 4 |
| Sub-Saharan Africa | 0.7 | 22 |
| Eastern Europe | 0.66 | 8 |
| Eastern Asia | 0.6 | 5 |
| Western Asia | 0.58 | 14 |
| Southern Europe | 0.44 | 9 |
| South-eastern Asia | 0.39 | 8 |
| Latin America/Caribbean | 0.29 | 15 |
| Southern Asia | 0.23 | 6 |
| Northern Europe | 0.07 | 10 |
| Northern Africa | -0.41 | 3 |
| Western Europe | -0.49 | 6 |

When correlations between indicators of development between the psychometric and scholastic estimates are contrasted, they are typically similar in magnitude, as shown in **Figure 7**. The correlation between the correlations derived from both variables is .91, further evidencing that scholastic and psychometric tests are measuring a similar construct across countries.
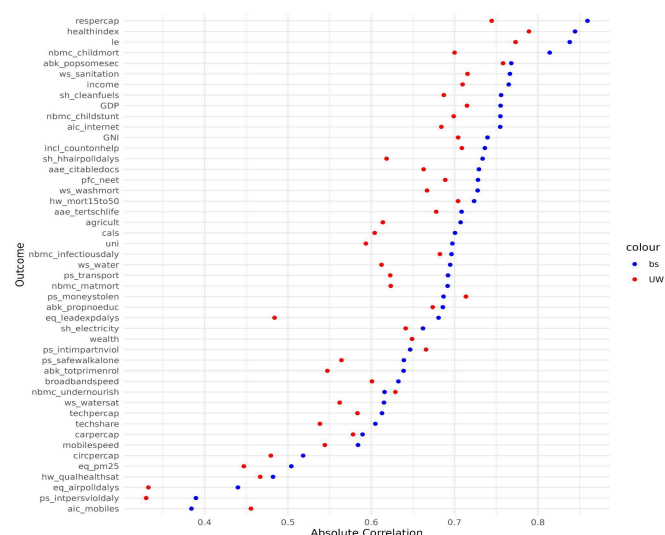


**Figure 7:** Absolute correlations between indicators of development and average IQs based on Becker's unweighted estimates (red) and estimates of average ability from the basic skills dataset (blue). National variables were transformed and their missing data was imputed.

Sear also noted that there was no formal search strategy or exclusion strategy carried out by Becker and Lynn-this is a fair criticism, but keep in mind that search strategies are easy to falsify and that flexibility is necessary to estimate national intelligence. In some cases, unweighted means are more accurate than sample size weighted means when the sample sizes of the studies are large, when the sample sizes are small, it would be better to weigh by the sample size. For countries that have a large amount of data (e.g. South Africa) adding psychiatric, foreign, or rural samples to the dataset would be unnecessary. In other countries that have no data available, low quality samples would be better than none. In most nations, the scholastic data is of higher quality than the psychometric data, but if the psychometric data is of high quality, then it may be wise to weigh it more highly for that specific nation.
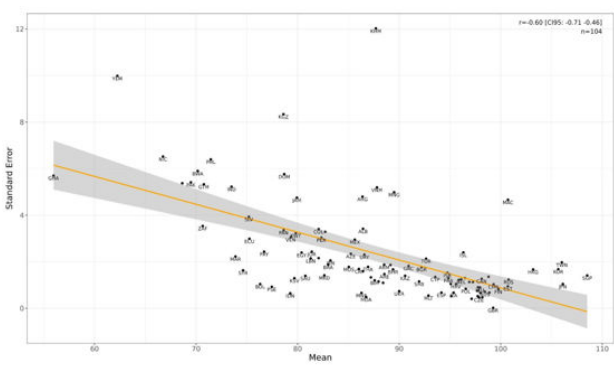
## National IQ Standard Errors

Sear's focal criticism of the national IQ datasets, particularly Lynn's and Becker's, was that the quality of data is not equally distributed across regions. This is an inevitability, given less developed countries have lower data quality, thus the criticism is not specific to intelligence measurements [18]. Many countries in Becker's dataset were estimated using small samples this is true, but a small sample is still better than none, and even a sample of 20 can provide a reasonably precise estimate of a population mean, as the standard error will be only 3.4 IQ points. The true standard error of national IQ estimates is even higher than this, as the various proxies for national intelligence that were collected only correlated at .87 on average, implying an average standard error of 5.41 ($5.41=\text{sqrt}(1-0.87) \times 15$). This large standard error indicates that the error variance is due to heterogeneity between samples, not random sampling error. Restricting to the earlier set of datasets that had no overlapping data (recent TIMSS/PIRLS/PISA results, Rindermann's SAS estimates, and Becker's quality weighted psychometric estimates) resulted in the same average correlation (.87). In any case, many other national datasets were based on small samples, when nothing else was available, and they were not excessively criticized for this reason [19].

Warne argued in a reply to Sear that the quality of Becker's data does not vary by regional group or average level of national IQ, based on the fact that Becker's quality assessments of the data do not vary by the average IQ of the sample. This is incorrect, as high levels of sample quality in certain regions may be indicative of fraud. Empirically, Becker's quality weighted estimates of intelligence have roughly the same correlation with SDI (.81) as his unweighted estimates (.83). Based on priors, it should be the case that higher quality samples should result in more accurate estimates of intelligence; because they don't, the alternative hypothesis that the higher quality samples are more likely to be fraudulent must be considered.

The hypothesis that lower IQ nations have more imprecisely estimated means by collecting estimates of national intelligence that were based on different data (recent TIMSS/PIRLS/PISA assessments, Becker's psychometric estimates weighted by quality, Rindermann's estimates of scholastic ability) and estimating the means and the standard errors, where the standard deviation of the sample averages divided by the square root of the number of samples.

Standard errors and means are correlated negatively between countries (r=-0.60, p<.001), meaning that estimates made of lower IQ countries were less accurate, as shown in **Figure 8**. On average, a country's estimated IQ has a standard error of 2.33, though this figure varies substantially by country: from 0.41 in Denmark to 12 in Cambodia.



**Figure 8:** Plot of standard errors and means of national IQ estimates.

This is not due to intelligent countries having data from more samples; the negative relationship between the mean and the standard error holds after controlling for the number of samples used to estimate intelligence, as shown in **Table 6**.

**Table 6:** Regression models that predict the standard errors of the estimates.

| Parameter | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| Estimated mean IQ | -0.12 (0.016)*** | - | -0.089 (0.018)*** |
| Number of samples | - | -0.49 (0.079)*** | -0.26 (0.086)** |
| $R^2$ | 0.36 | 0.28 | 0.41 |

**Note:** *-> p< .05, **->p < .01, ***-> p< .001.

## Estimating National Intelligence Averages

To compute the intelligence of nations, measured IQ and achievement test results are used. While these are not perfect measurements of intelligence, IQ scores are predictive of socially important outcomes and show low levels of bias between groups in contrast to personality measurements which are confounded by reference group effects [20].

Multiple sources of data were consulted, including psychometric estimates (Becker unweighted, Becker sample-weighted, Becker quality-weighted), scholastic estimates (World Bank test scores, basic skills dataset, PISA 2022 results, Becker scholastic estimates, Rindermann scholastic estimates), and composite estimates (Lynn 2012, Lynn 2002, Becker composite, Rindermann composite). If a dataset included geographic imputations, the imputations were removed.

Rindermann included estimates that were based on performance in the mathematics olympiad for North Korea, Belarus, Brunei, Cambodia, Mauritania, Tajikistan, and Turkmenistan; these were kept, though this was most relevant for Turkmenistan, which has no measured data.

Samples were normed in a fashion that placed the UK at a mean of 99.26, which is roughly what the UK's average psychometric IQ is compared to British Whites. In one case

where a UK sample was not available, the average of Americans was used as an anchor instead.

It was tested whether some samples were of higher quality than others, and statistical analysis suggested that this was the case (which is available in the supplement), though subjective indicators of quality (e.g. how new the data is, how much data the indicators are based on) was also taken into consideration. Concretely speaking, Lynn's and Becker's composite estimates were given lower weights due to the fact that they are based on older data and provide little incremental validity. An overall average was computed using nested means:

- **Nest 1:** Lynn's estimates, Becker's composite estimates, Becker's scholastic estimates, and recent TIMSS math results.

- **Nest 2**: average of nest 1, recent TIMSS science results, average of Becker's psychometric estimates, recent PIRLS results, World Bank test scores

- **Nest 3:** average of nest 2, recent PISA results, and Rindermann's scholastic estimates

- **Nest 4:** average of nest 3, basic skills dataset, Rindermann's IQ estimates

Another method was tested where random effects meta-analytic means were calculated for each country. Sample sizes were assigned based on the perceived quality of each dataset:

- N=10 → TIMSS math, Becker psychometric averages.
- N=20 → Becker composite, TIMSS science, Lynn estimates, Becker's scholastic estimates.
- N=40 → PIRLS results, PISA results, WB test scores, Rindermann SAS estimates.
- N=80 → Rindermann IQ estimates and basic skills dataset.

Samples that displayed unusual heterogeneity or extreme means in either direction were manually reviewed, where the sources were consulted and a subjective best estimate was given. Most countries that had suspiciously large amounts of variance in estimates were undeveloped countries, though there were notable exceptions like Vietnam and China. In the case of Vietnam, Becker included estimates of the IQ of rural Vietnamese who scored an IQ of 78 in his dataset; their performance on the PISA tests suggests that the true national IQ is somewhere between 95 and 100. In China, the differences in estimates between datasets is due to a debate over how the PISA samples should be weighted relative to the rest of China. The World Bank estimated its human capital to

be the IQ equivalent of 90, while the basic skills dataset estimated its human capital to be the IQ equivalent of 107 both agreed that the PISA results were not representative, but differed in the extent to which this biased the overall average. Using the China family panel study, regional differences in cognitive ability were calculated, and it was determined that China's recent PISA results are biased because they come from more intelligent provinces like Shanghai (IQ=107) and Beijing (IQ=108), and that if the results were weighted relative to the whole population, they are indicative of an IQ of roughly 99. The scores from the IQ samples are also inflated by the fact that they come from educated and Eastern samples, when this bias is corrected for, the results imply an average of roughly 102 for the whole country.

In total, 42 countries had their national IQs estimated based on a manual review, and the estimates correlated at .97 with the estimates that would have been made otherwise and were 1.9 IQ points higher (p<.001, two-sided paired t-test) on average. In most cases, the manual revisions were unnecessary, as shown in **Table 7**.

**Table 7:** Average IQ by country, by method.

| Country | Mathematical estimate | Manual (final) estimate |
|---|---|---|
| Afghanistan | 74.8 | 75.7 |
| Cambodia | 83.09 | 84.1 |
| Canada | 100.22 | 100.88 |
| China | 101.03 | 100.2 |
| Cuba | 90.64 | 87.9 |
| Dominica | 68.96 | 75.84 |
| Dominican Republic | 77.07 | 82.41 |
| Ecuador | 80.5 | 82.04 |
| Egypt | 79.56 | 81.26 |
| El Salvador | 77.14 | 79.87 |
| Equatorial Guinea | 61.56 | 69.67 |
| Estonia | 101.14 | 101.86 |
| Finland | 100.62 | 100.86 |
| Gambia | 62.83 | 63.7 |
| Guatemala | 75.46 | 78.78 |
| Haiti | 71.89 | 72.74 |
| Honduras | 74.57 | 79.3 |
| Hong Kong SAR China | 103.54 | 106.02 |
| Iraq | 84.62 | 82.27 |
| Ireland | 98.02 | 99.1 |

| | | |
|---|---|---|
| Jamaica | 77.18 | 79.82 |
| Japan | 103.96 | 105.9 |
| North Korea | | 87.9 |
| South Korea | 104 | 103.84 |
| Kuwait | 79.51 | 84.26 |
| Kyrgyzstan | 77.29 | 80.51 |
| Laos | 84.23 | 84.77 |
| Macao SAR China | 102.62 | 103.9 |
| Marshall Islands | 80.45 | 86.5 |
| Mongolia | 89.66 | 93.37 |
| Nepal | 73.01 | 76.98 |
| Netherlands | 99.58 | 100.08 |
| Nicaragua | 74.39 | 77.95 |
| Pakistan | 73.42 | 70.86 |
| Papua New Guinea | 79.37 | 71.77 |
| Romania | 89.14 | 87.34 |
| Samoa | 81.91 | 88 |
| Singapore | 106.37 | 108.7 |
| Taiwan | 103.34 | 105.23 |
| Uzbekistan | 83.88 | 83.95 |
| Vietnam | 93.63 | 98.52 |
| Zambia | 70.52 | 77 |

# RESULTS

Measurements of national intelligence and socioeconomic development correlated at .88 between countries (n=197). Average IQs and SDI have been plotted in **Figures 9 and 10**. The average IQ of the world is 85.3 when weighted by population size.
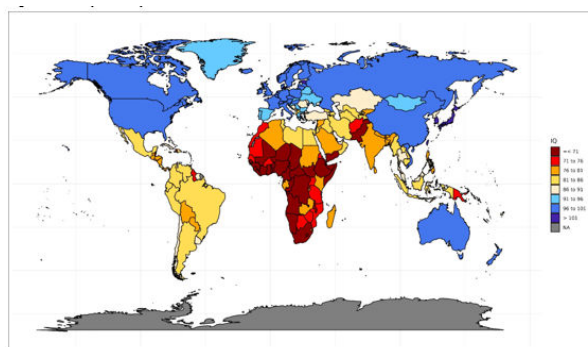


**Figure 9:** IQ by country.
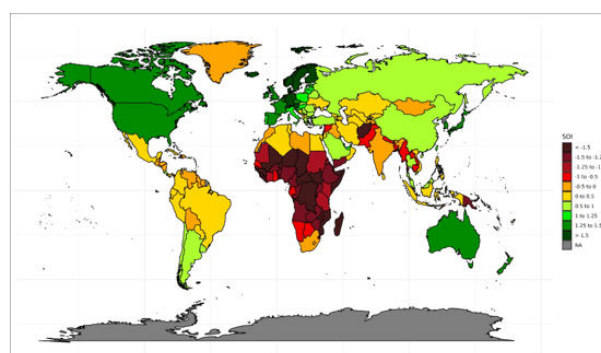


**Figure 10:** Map of socioeconomic development around the world.

Heterogeneity was observed in the correlation between SDI and national IQ according to the Breusch-Pagan test (p=.0012), with lower IQ nations showing more variance in the relationship between intelligence and socioeconomic development. The non-linear relationship between the two variables marginally passed significance testing (F=2.54, p=.04)

The relationship between SDI and average IQ has been plotted in **Figure 11**.
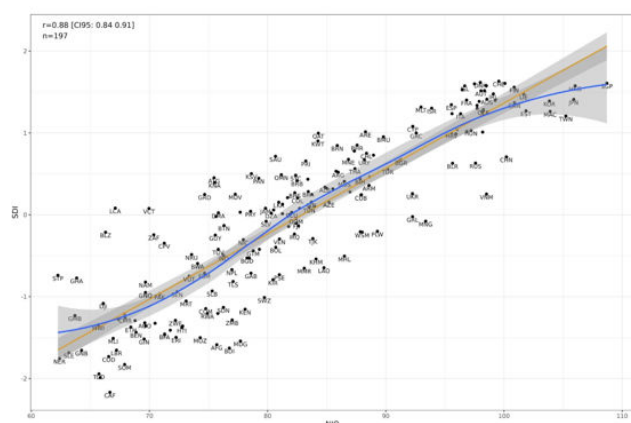


**Figure 11:** Relationship between national IQs and the socioeconomic development index.

Despite the strong levels of agreement between the measurements of socioeconomic development, there were still some large outliers in the relationship. Many Middle Eastern countries, China, and Turkey all rose over 20 ranks in our measurement of socioeconomic development relative to the Social Progress Index, as shown in **Figure 12**.
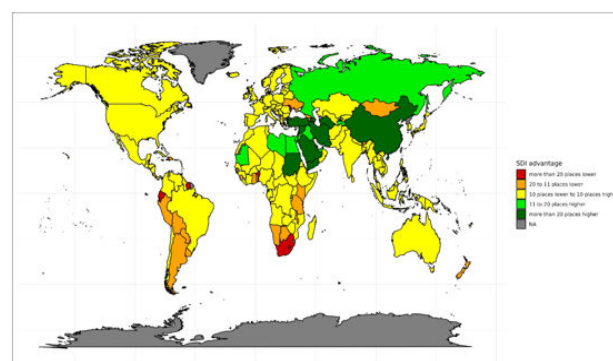


**Figure 12:** Difference in ranks between Social Development Index (SDI) and Social Progress Index (SPI). Green colour corresponds to higher relative rank, redder colour to lower.

Average IQs and SDIs have been displayed in **Table 8**, with average IQs ranging from 70.8 in Sub-Saharan Africa to 100.8 in Eastern Asia. Regional differences in intelligence and socioeconomic development highly correlate (r=.96), as shown in **Figures 13 and 14**.

**Table 8:** Average IQ and SDI by region.

| Region | Average IQ | Average SDI |
| --- | --- | --- |
| Eastern Asia | 100.79 | 1.07 |
| Western Europe | 98.61 | 1.51 |
| Northern Europe | 98.46 | 1.41 |
| Australia and New Zealand | 98.34 | 1.33 |
| Northern America | 95.17 | 0.9 |
| Eastern Europe | 94.16 | 0.75 |
| Southern Europe | 90.69 | 0.94 |
| South-eastern Asia | 87.1 | 0.13 |
| Western Asia | 84.21 | 0.36 |
| Polynesia | 83.9 | -0.26 |
| Central Asia | 83.64 | 0.09 |
| Micronesia | 80.94 | -0.54 |
| Latin America / Caribbean | 80.18 | 0.09 |
| Northern Africa | 79.79 | -0.19 |
| Southern Asia | 77.37 | -0.42 |
| Melanesia | 75.72 | -0.8 |
| Sub-Saharan Africa | 70.76 | -1.22 |
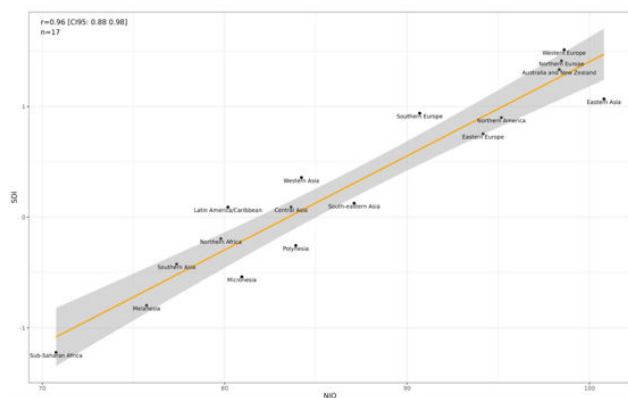
**Figure 13:** Plot of average IQs and SDIs by region.
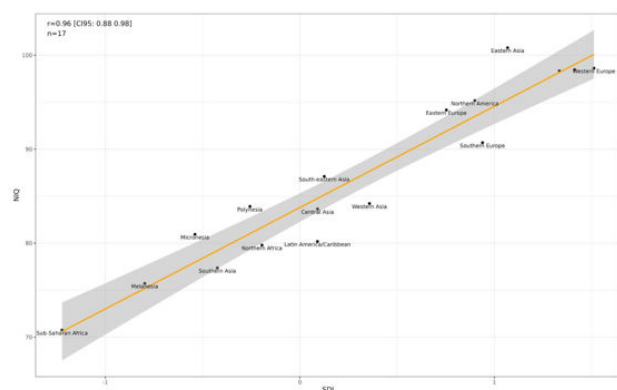


**Figure 14:** Plot of average IQs and SDIs by region (axes inverted).

The analysis that related the standard errors and the means of national IQs was repeated for the dataset that included all national IQ datasets. We found a negative correlation between standard errors and means (spearman's rho=-.63, p<.001), meaning that countries with higher IQs had their estimates more precisely taken, as shown in **Figure 15**. This negative correlation also held for socioeconomic development, where more developed countries had lower standard errors (rho=-.65, p<.001).
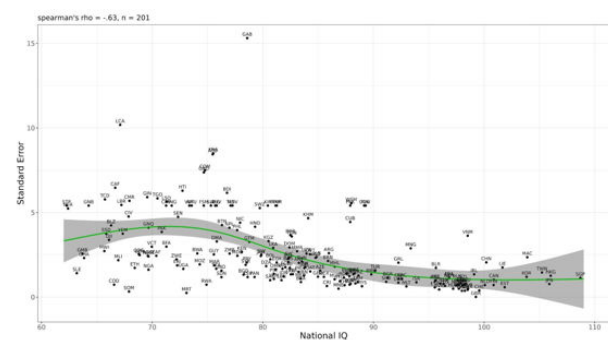


**Figure 15:** Relationship between standard errors of national IQs and estimated national IQ.

## DISCUSSION

We were able to replicate prior literature that found that measurements of socioeconomic development are correlated with measurements of human capital, though our correlation is higher than the ones found prior (r=.88). This is probably because our measurements of human capital and socioeconomic development are of higher quality than the ones that preceded it-the measurement of socioeconomic development is based on 47 variables and advanced statistical techniques were used to calculate the averages; the national IQ measurement is a composite of other datasets, which causes the error to decrease.

The large magnitude of the correlation is a function of the relationship being bidirectional: Increases in intelligence have been observed as countries have become more economically developed, and the deficiency in IQ of certain undeveloped nations (e.g. Africa) clearly cannot be attributed to genetic causes, therefore it would be reasonable to conclude that socioeconomic development causes intelligence. On the other hand, intelligence is the most robust and strong predictor of economic growth, and causality from intelligence to socioeconomic development can be proven with the use of historical variables such as age heaping and cranial capacity.

Our measurement of socioeconomic development, the SDI, correlates highly with the HDI and the SPI (r=.98 and .97, respectively), indicating that it has high levels of external validity. The SDI estimates the development of authoritarian countries such as Iran, Saudi Arabia, Singapore, Turkey, Turkmenistan, and Russia to be higher than the SPI, probably because it does not base its estimates of socioeconomic development on cultural values or political indexes.

The national IQ estimates were shown to have non-negligible inaccuracy a standard error of roughly 5.41 IQ points. We have estimated that the composite measurement (SE of 2.6) has 50% less error than the average dataset that measures proxies for national intelligence. Most of the estimates made of individual countries are accurate, though a few have very high standard errors (Gabon, Cambodia, Cuba, Saint Lucia, and Haiti) or are based on dubious estimation methods (Turkmenistan was estimated using mathematical olympiad performance, North Korea was estimated using North Korean refugees and it was difficult to judge how to correct for Flynn Effects). We also found that more intelligent and developed countries tended to have more precisely estimated national IQs, even after controlling for the fact that intelligent and developed countries are more likely to be represented in these datasets.

The research on whether scholastic test scores between nations pass measurement invariance suggests that measurement invariance between countries is usually tenable, with nonverbal tests (e.g. mathematics) showing more invariance than verbal (e.g. reading) ones. As these nonverbal and verbal tests have differences of roughly the same magnitude across countries, the violations of measurement invariance are not likely to be a practically significant source of bias when assessing differences in IQ

between countries. Some studies have suggested that matrix reasoning does not test intelligence equally between Europeans and Sub-Saharan Africans the research is not definitive enough to make inferences, unfortunately.

## CONCLUSION

Some groups that are genetically highly similar still differ greatly in IQ: South Koreans score 16 points higher than North Korean refugees on cognitive tests, and African Americans score 11-14 points higher than Africans. This sets a rough upper limit on how much Flynn Effects can bias estimates of intelligence between nations. The magnitude of the observed differences between nations is much larger than this, with scores ranging from 108.7 in Singapore to 62.26 in Sao Tome. Because of that, it would be rational to conclude that the disparities in test scores between countries are largely due to true differences in ability instead of test bias.

## ACKNOWLEDGEMENT

## REFERENCES

1.  Angrist N, Djankov S, Goldberg PK, Patrinos HA (2021) Measuring human capital using global learning data. Nature. 592(7854):403-408.

2.  Asil M, Brown GT (2016) Comparing OECD PISA reading in English to other languages: Identifying potential sources of non-invariance. Int J Test. 16(1):71-93.

3.  Bakhiet SF, Dutton E, Ashaer KY, Essa YA, Blahmar TA, et al. (2018) Understanding the Simber Effect: Why is the age-dependent increase in children's cognitive ability smaller in Arab countries than in Britain?. Pers Individ Dif. 122:38-42.

4.  Beaujean AA, Osterlind SJ (2008) Using item response theory to assess the Flynn effect in the National Longitudinal Study of Youth 79 Children and Young Adults data. Intelligence. 36(5):455-463.

5.  Beaujean AA, Sheng Y (2010) Examining the Flynn effect in the general social survey vocabulary test using item response theory. Pers Individ Dif. 48(3):294-298.

6.  National Academies of Sciences, Engineering, and Medicine (2015) Mental disorders and disabilities among low-income children.

7.  Cardoza SL (2006) Differential item functioning in Asian-Americans on the Stanford-Binet standardization fifth edition verbal subtests.

8.  Carlsson M, Dahl GB, Öckert B, Rooth DO (2015) The effect of schooling on cognitive skills. Rev Econ Stat. 97(3):533-547.

9.  Jensen S, Kirkegaard EO (2024) National IQs and Socioeconomic Development.

10. Cox SR, Ritchie SJ, Fawns-Ritchie C, Tucker-Drob EM, Deary IJ (2019) Structural brain imaging correlates of general intelligence in UK Biobank. Intelligence. 76:101376.

11. Crede M, Bashshur M, Niehorster S (2010) Reference group effects in the measurement of personality and attitudes. J Pers Assess. 92(5):390-399.

12. De la Cruz JT, De los Santos Pérez L, Tejada FC (2022) Relationship among dental fluorosis, intellectual quotient and academic performance. Fluoride. 55(1):41-48.

13. DeCarli C, Maillard P, Pase MP, Beiser AS, Kojis D, et al. (2024) Trends in intracranial and cerebral volumes of Framingham heart study participants born 1930 to 1970. JAMA Neurol. 81(5):471-480.

14. Francis G, Kirkegaard EO (2022) National intelligence and economic growth: a Bayesian update. Mank Q. 63(1).

15. Gust S, Hanushek EA, Woessmann L (2024) Global universal basic skills: Current deficits and implications for world development. J Dev Econ. 166:103205.

16. Iliescu D, Ilie A, Ispas D, Dobrean A, Clinciu AI (2016) Sex differences in intelligence: A multi-measure approach using nationally representative samples from Romania. Intelligence. 58:54-61.

17. Jensen AR (1970) A theory of primary and secondary familial mental retardation.

18. Kamin LJ (2006) African IQ and mental retardation. S Afr J Psychol. 36(1):1-9.

19. Karwowski M, Milerski B (2021) Intensive schooling and cognitive ability: A case of Polish educational reform. Pers Individ Dif. 183:111121.

20. Kirkegaard EO, Karlin A (2020) National intelligence is more important for explaining country well-being than time preference and other measured non-cognitive traits. 61(2).