Rindermann H.
*Prof. Dr. Heiner Rindermann*
*Department of Psychology, Chemnitz University of Technology*
*Wilhelm-Raabe-Str. 43, D-09107 Chemnitz, Germany*

*E-mail: heiner.rindermann@psychologie.tu-chemnitz.de*

DOI: 10.14673/HE2025341159

# Low cognitive ability estimates in developing countries: A statistical analysis of their credibility

Developing countries tend to achieve weak results in international assessments of cognitive competence. Their performance is typically about one to two standard deviations below the average scores obtained in Western countries. The results are sometimes so low that they appear difficult to believe. For example, in the World Bank dataset, Nigeria scored 262 student assessment points (SASQ, about two and a half standard deviations below the norm of 500, equivalent to an IQ of 64). In the Altinok and Diebolt dataset, Yemen scored 273 SASQ (equivalent to an IQ of 66). These results have sparked criticism, and ethical as well as political-ideological debates, culminating in motivated retractions of scientific papers. In this paper, we cross-validate the reported values by comparing different data sources and applying statistical predictions. Across test paradigms, the results for groups of countries appear consistent. However, substantial discrepancies remain for individual countries. The same holds for comparisons across test collections: aggregated country results are similar, while single-country outcomes differ. Using education, GDP per capita, and political indicators to predict test scores yields somewhat higher values for the Global South (by about 5 IQ points), with the strongest effects observed in Latin America (rising from 78 to 86 IQ points). Where large discrepancies occur between observed and predicted test results, the potential for improvement appears substantial. Cultural and evolutionary background factors proved to be stronger predictors than political variables, whereas genetic distance to South or East Africa was less predictive.

KEY WORDS: *cognitive ability; nations; intelligence; student achievement studies; developing countries.*

## 1. Introduction

Developing countries (also referred to as the Third World, emerging economies, or the Global South) achieve weak results in international studies of cognitive competence – typically one to two standard deviations below the averages obtained in Western countries. Some scores are so low that they are hard to believe. What can be their meaning? For example, in a World Bank student achievement dataset (Angrist et al., 2021), Nigeria scored 262 student assessment points (SASQ), about 2.5 standard deviations below the norm of $M$=500, $SD$=100, equivalent to an IQ of 64. In the Lim et al. (2018) combined dataset of student and psychometric intelligence tests,

Yemen scored 336 SASQ (IQ 75). The results for Yemen were even weaker in the pure student achievement study collection by Altinok and Diebolt (2024): 273 SASQ (equivalent to an IQ of 66). Yet the conversions

$$((\frac{(x-500)}{100})\times 15)+100$$

appear overly optimistic, as the UK scored above the SASQ mean in all three studies (519, 527, and 535, respectively). If the UK benchmark were set at 100 IQ (equivalent to a national mean of 98–99 including immigrants), the adjusted scores would be lower: around IQ 59–60 for Nigeria and 61–70 for Yemen. For comparison, Lynn and Becker (2019) reported slightly higher IQs for these countries – 68 for Nigeria and 72 for Yemen – using the UK (IQ 98) as benchmark. Finally, based on PISA for Development data (Programme for International Student Assessment in developing countries), education economists Lant Pritchett and Martina Viarengo (2021, p. 1) concluded that "the vast majority of children in these countries would not reach the minimum targets."

Such low results have sparked opposition, scientific critique, and ethical and political-ideological debates, at times leading to motivated retractions of scientific papers (e.g., Rindermann, 2024a). For instance, Rebecca Sear (2022, p. 1) wrote that Lynn and Becker's IQ dataset were "not fit for purpose" and "not only inaccurate but systematically biased". Or Kevin Bird (2020, p. 472) accused the Lynn data of "systematic bias". Of course, all scientists, including those from other disciplines, may condemn Lynn and Becker's data.[1] However, when using student achievement data, the results are generally similar or even worse. For example, instead of IQ 68 for Nigeria (Lynn & Becker), student achievement-based estimates yield IQs of 84 or 71 (SASQ converted to an IQ scale, benchmark UK 99 IQ points, Lim et al.) and 64 or 59 (UK=98, World Bank). Similarly, instead of IQ 72 for Yemen (Lynn & Becker), the corresponding estimates are 75 or 70 (UK=99, Lim et al.) and 63 or 57 (UK=98, World Bank).

There appear to be differences between data sources, but at least for these two countries one cannot conclude that Lynn and Becker's IQ data systematically yield lower results. Russell Warne (2023, p. 219), an intelligence researcher who compared Lynn and Becker's psychometric intelligence tests with various student assessment surveys (e.g., the World Bank collection), described their collection as "a scientifically worthwhile endeavor". It should also be noted that student achievement studies seem to enjoy a kind of protective shield against criticism, even though some results

[1] Rebecca Sear is a "demographer and anthropologist" at the "London School of Hygiene and Tropical Medicine" (21.06.2023, www.rebeccasear.org), Kevin Bird is a plant specialist. There is no evidence of training in psychometric diagnostics or student achievement studies.

are clearly dubious or even fraudulent and remain uncorrected in most collections (with values deviating sharply upward from other studies and neighboring countries, e.g., Kazakhstan in TIMSS 2007 and Cuba in LLECE 1997 and 2005–2006; see Rindermann, 2018, Appendix, pp. 2, 4).[2]

However, it cannot be denied that results in many developing countries are implausibly low. Particularly striking are the scores at the 5th percentile. Student achievement studies typically report not only mean scores but also ability levels at the 95th percentile ("intellectual classes," "smart fraction," "skilled human capital") and at the 5th percentile ("low performers"). In Ghana, students at the 5th percentile achieved an SASQ of 99, corresponding to an IQ of 36 (four standard deviations below the mean). After correcting for age differences (students being older or younger than the international mean) and extrapolating to the entire age cohort, including out-of-school adolescents, the results fall to SASQ 58 or IQ 31. These corrections are deliberately moderate, as we assumed that out-of-school youth in countries with lower achievement levels lose fewer developmental opportunities by being out of school (Rindermann, 2018, Appendix, p. 9).

What do such low results *mean*? Can they be accurate? Richard Nisbett (2009, pp. 214f.) questioned this very point:

"Let's stop and think for a moment about what an IQ of 70 might mean, if we took it seriously as an actual indicator of intelligence of sub-Saharan blacks. ... Given what we know about people with such low IQs in our society, the average African, then, might not be expected to know when to plant seeds, what the function of a chief might be, or how to calculate degrees of kinship. Obviously something is desperately wrong with these African IQ scores. They cannot possibly mean for the African population what they do for people of European culture."

Or consider 2023, when student assessment researcher Leslie Rutkowski stated: "This adjusted score [from Lynn and Becker] puts the population of Burkina Faso at just above intellectually disabled, according to the DSM-V, or what was referred to as 'mentally retarded' in earlier versions of the DSM." Richard Nisbett similarly remarked: "Obviously something is desperately wrong" – but what exactly seemed "obvious" to him, and what evidence did he provide to support this judgment? In a comparable case, Arthur Jensen (1998, pp. 367ff.) raised the same issue as Nisbett (and Rutkowski): low IQ results appear to have a different meaning for people in the

---

[2] PISA: Programme for International Student Assessment.
TIMSS: Trends in International Mathematics and Science Study.
PIRLS: Progress in International Reading Literacy Study.
LLECE: regional Latin American Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) study.

US of European descent compared with those of African descent. Based on "my own observation and testing of pupils in special classes," Jensen (1998, p. 367) wrote:

"In social and outdoor play activities ... black children with IQ below seventy seldom appeared as other than quite normal youngsters – energetic, sociable, active, motorically well coordinated, and generally indistinguishable from their age-mates in regular classes. But this was not so for as many of the white children with IQ below seventy. More of them were somehow 'different' from their white age-mates in the regular classes. They appeared less competent in social interactions with their classmates and were motorically clumsy or awkward."

For teachers, white pupils with an IQ of 70 or lower generally appeared to have intellectual disabilities, whereas black pupils with similar scores did not. Nevertheless, the validity of IQ measures – e.g., in predicting results on cognitive tasks at school that require conceptual learning and problem-solving – was similar for both groups. Jensen explained the different patterns of competence in less cognitively demanding situations, where black children tended to perform better, as reflecting different causes of low intelligence: Among white children, an IQ of 70 or below was often the result of a severe neurological impairment leading to general mental retardation, whereas among black children low IQ more often represented a normal deviation from the mean. Taken together, Nisbett's and Jensen's arguments underscore the importance of drawing on evidence from everyday life and behavior.

## 2. Aims and outline of the study

The aim of this study is to examine whether the relatively low scores in cognitive competence tests in developing countries are accurate or misleading. To assess the credibility and meaningfulness of these results, we employ several methods: First, we compare the results of different test paradigms (student achievement, adult achievement, psychometric intelligence tests). Second, we compare the results of different test collections (Altinok & Diebolt, 2024; Gust, Hanushek & Woessmann, 2022; Lim et al., 2018; Lynn & Becker, 2019; Rindermann, 2018; World Bank/Angrist et al., 2021). Third, we predict ability levels using theoretically and empirically relevant variables such as education, wealth (GDP per capita), and background factors. The prediction formulas are estimated on samples excluding the target countries and then applied to them.[3]

[3] The study is a revised version of Rindermann (2024a), which was retracted without scientific reasons given (Rindermann, 2025). Significant changes: 1. New data from student achievement studies. 2. Alternative measures for evolutionary background. 3. Linguistic refinement using Large Language Models. 4. Separate publication of the section on qualitative observations.

## 3. Method

### *3.1 Data bases*

The following seven sources were used:

(1) *Cognitive ability* (*Rindermann dataset*): Data from several *student assessment studies* were combined (PISA, TIMSS, PIRLS, and others from 1991 to 2019; see Rindermann, 2018; Rindermann & Carl, 2020; updated March 2022). The student assessment means were merged with data from psychometric intelligence tests (Lynn & Becker, 2019). Adjustments were made for *age* (students in some countries are older or younger than the international average; older = malus, younger = bonus) and for *school attendance* (the proportion of youth attending school may be higher or lower than the international average; higher = bonus, lower = malus). Student data from countries with only regional coverage were adjusted to be nationally representative (e.g., Shanghai for China). Dubious results were excluded (e.g., Kazakhstan in TIMSS 2007; Cuba in LLECE 1997 and 2005–2006). IQ estimates not directly measured but inferred from neighboring countries were also downgraded (malus). The UK was set at around 99 (Greenwich norm: British natives at IQ 100; including immigrants, the national mean is about IQ 98–100). Data are given for 199 countries (including estimated but corrected values based on data from Lynn & Becker, 2019). This dataset includes corrections that are not entirely free of subjective judgment but are scientifically grounded, though open to alternative interpretations (e.g., Rutkowski et al., 2024).

(2) Altinok and Diebolt (2024) compiled and estimated "learning outcomes" based on student achievement tests for 224 countries and territories from 1970 to 2020 in five-year intervals. They drew on large international surveys (TIMSS, PISA, PIRLS) as well as regional studies (e.g., SACMEQ and LLECE). However, the dataset has several problems: 1. It is unclear for which years actual measurements exist. According to their Table 3, there are 118 "observations" in reading and 104 in math for 1975. Yet I am not aware of any international comparative student achievement study from 1975. Measurements are not distinguished from imputations. 2. The method of estimation is insufficiently described, only: "We use a multiple imputation technique." (p. 705). 3. It is unclear what the scale, which (as I assume) has a theoretical mean of 500 and a standard deviation of 100, represents. While this is irrelevant for correlational studies, it matters for comparisons of mean levels. We combined the data for 2015 and 2020, yielding mean scores for 150 countries. Compared with all other datasets, the UK recorded the highest score in this collection, with 535 SASQ.

(3) *Student achievement* (*human capital*): Angrist, Djankov, Goldberg, and Patrinos (2021) present an average measure of cognitive ability – or "cognitive human capital", as the authors call it – based on student assessment studies. The World Bank research group compiled data for 163 countries and reported results on a student

rating scale with a theoretical mean (norm value) of 500 and a standard deviation of 100 (see Table 2).

(4) *Cognitive human capital* (learning): Lim et al. (2018) present a score for cognitive ability based on student assessment studies and psychometric intelligence test data. They refer to it as "learning" or "education quality." Data are reported for 192 countries on a student assessment scale (500/100).

(5) *Student achievement*: Gust, Hanushek, and Woessmann (2022) compiled student assessment studies and calculated a mean score, which they term the "mean achievement of students on a global scale." A second measure captures the percentage of youth (not only students) in a country who reach at least a basic ability level, defined as SASQ 410–420 or higher (originally reported as the share of children not reaching this level, which we inverted into a positive variable). According to PISA, pupils at this level are able to identify information and perform routine procedures following direct instructions in explicit situations. Data are available for 159 countries.

(6) *Adult cognitive ability* (*adult competencies*): The Programme for the International Assessment of Adult Competencies (PIAAC) measured reading literacy, numeracy, and problem solving among adults aged 16 to 65 (conducted in 2011/12 with an additional sample in 2014/15; OECD, 2013, 2016). PIAAC employs a different scale ($M$=263, $SD$=18). We transformed the scores – using a formula based on countries with results in both datasets – onto the SASQ and IQ scales. Data are available for 33 countries.

(7) *Psychometric intelligence*: Lynn and Becker (2019) compiled results from various intelligence tests conducted in different years and with different samples to produce a common mean. Data quality varies across countries; for example, the U.S. ($k$=58) has many more samples than Argentina ($k$=5). Measured results are reported for 129 countries (our study does not include their values estimated from neighboring countries). In subsequent steps, Lynn and Becker also combined psychometric intelligence test results with student assessment data, which are likewise not included in the present study.

The datasets by Rindermann, Altinok, the World Bank, Hanushek, the OECD, and Lynn represent improvements and extensions of earlier versions produced by these authors and institutions. All draw on results from international student assessment studies such as PISA, TIMSS, and PIRLS, and many also incorporate data from regional studies, including LLECE in Latin America and SACMEQ (Southern and Eastern Africa Consortium for Monitoring Educational Quality) in Africa.

As suggested by previous research, we assume that, for both theoretical (cognitive and substantive) and empirical (correlation, factor analysis) reasons, psychometric intelligence tests and student assessment studies measure largely similar constructs (e.g., latent $r$=.83 Kaufman et al., 2012; $r$=.73, Pokropek, Marks & Borgonovi, 2022;

Rindermann, 2024b; Rindermann & Baumeister, 2015). A notable issue is that, in student achievement studies, it is often unclear what the scale values of 500 (M) and 100 (SD) represent – i.e., which country sample and time point define the average. However, the scores can subsequently be standardized to Great Britain (Greenwich) as 500, leaving the standard deviation unchanged.

*3.2 Prediction formula*

We used five plus two additional variables to derive the optimal formula for predicting cognitive ability at the country level. These variables are widely regarded as key determinants, correlates, and consequences of high or low cognitive ability at both the national and individual levels (Ceci, 1991; Hanushek & Woessmann, 2015; Harrison, 2013; Jones, 2016; Rindermann, 2018):

(1) *Educational level (attainment) of society*: Standardized values of three measures were averaged: 1. Adult literacy rate – the ability to read and write simple sentences or perform comparable basic tasks (e.g., filling out an application form) for the population aged 15 and older (Kurian, 2001, p. 349f.; *N*=195 countries). 2. Percentage of 12- to 19-year-olds from 1960 to 1985 (corresponding to the cohorts assessed in older student studies, now adults) who completed secondary school (Mankiw, Romer, & Weil, 1992; *N*=117). 3. Mean years of schooling of individuals aged 25 and older for 1990, 1995, and 2000 (Barro & Lee, 2000; *N*=107). All sources used data from the UN or comparable organizations. The combined index (Cronbach *a*=.93 for 101 countries with common data) is reported for 195 countries. School education is a strong predictor of cognitive ability at both the individual and national level, and also a causal factor in the development of thinking ability and knowledge acquisition (Rindermann & Ceci, 2018; Ritchie & Tucker-Drob, 2018; correlations at the international level see Table 3). However, cognitive gains vary across countries and exhibit diminishing returns with longer durations of schooling.

(2) *Wealth* (*income, output*): Logarithm of per capita GDP around 2010, measured in 1990 international dollars (Maddison, 2008; 2010 from Bolt & van Zanden, 2013). Since an increase in income at lower levels arguably has a greater impact on quality of life, we used the natural logarithm of GDP. This transforms nonlinear, exponential increases in currency units into linear increases in more meaningful "quality of life" units.[4] The more recent Maddison dataset contains fewer countries (*N*=117 versus 159), so we combined the 2008 and 2010 data after standardization, yield-

[4] Numerically identical increases at different income levels have different implications for quality of life. For example, a $5,000 increase from $5,000 to $10,000 substantially improves nutrition, health, housing quality, and education, whereas the same increase from $60,000 to $65,000 has a much smaller effect, often limited to discretionary expenditures. Consequently, statistical analyses of wealth indicators typically use the natural logarithm to account for these nonlinear effects (Rindermann, 2018, p. 3).

ing *N*=161.[5] According to the theory of cognitive human capital, cognitive ability enhances productivity and wealth at the individual, institutional, and national levels (e.g., Francis & Kirkegaard, 2022; Gottfredson, 2003; Hanushek & Woessmann, 2015; Jones, 2016; Rindermann & Becker, 2018). At the same time, cognitive development is influenced by health and environmental quality, both of which depend on wealth (e.g., Eysenck & Schoenthaler, 1997). Overall, wealth can therefore serve as a predictor of intelligence at the national level.

(3) *Positively valued politics*: This index is the average of three political variables – rule of law, political freedom, and democracy – measured over a long period dating back to the 1950s (*a*=.88, *N*=201; Rindermann & Carl, 2020). Cognitive ability positively influences rule of law, political freedom, and democracy, and longitudinal studies indicate that rule of law, in turn, has a positive effect on cognitive ability (e.g., Oesterdiekhoff, 2014; Rindermann, 2018).

(4) *Religion* (*culture*): Data on religious affiliation (percentage of adherents) were obtained from the German Foreign Office (www.auswaertiges-amt.de), a country encyclopedia (Yearbook/Jahrbuch, 2004), and the CIA World Factbook (www.cia.gov/cia/publications/factbook). Data, particularly for developing countries, were often missing and had to be re-estimated or corrected. In East Asia, Confucianism was assumed to be the predominant religion. Some religions (e.g., Buddhism, Confucianism) overlap with broader worldviews. An index was developed in which religions were weighted according to their support for education, rationality, and achievement orientation (Rindermann, 2018, Section 10.8). Data are reported for 202 countries. Religions operate as worldviews, exerting influence through the original message (sacred founding texts), the model effect of the religion's founder, interpreted and revised doctrines and their changing understanding over time, and contemporary lived practice. Several studies report positive effects of Protestantism, Judaism, and Confucianism on cognitive outcomes (e.g., Becker & Woessmann, 2009; Harrison, 2013; Rindermann, 2018).

(5) *Brain size* (cranial capacity, evolution): Data on cranial capacity were obtained from Beals, Smith, and Dodd (1984, p. 304, Figure 4). A research assistant assigned numerical values to different countries based on map shading. For populations alive today (e.g., in Australia and the Americas), values were estimated from the frequencies in their countries of origin (*N*=184). Measurement limitations (e.g., direct brain imaging would provide more accurate data) are likely to result in an underestimation of correlations. Brain size has a causal effect on intelligence (Gignac & Bates, 2017;

[5] We also examined three alternative measures of income and wealth – the Penn World Tables, the Human Development Report, and Credit Suisse – but all were less strongly correlated with cognitive ability. For this reason, we retained the Maddison measure to optimize prediction (the aim here).

Lee et al., 2019), but it is partially endogenous to wealth, i.e., not entirely independent of economic factors (Lynn, 1990). For purely statistical predictions, as in the present study, this endogeneity is not problematic. Cranial capacity serves not only as a measure of brain size but also as an indicator of evolutionary history, which is useful for predictive purposes (Rushton & Rushton, 2004; Will et al., 2021).

(6) *Genetic distances to Africa* (*evolution*): Brain size is only an indirect indicator of evolutionary history. A better indicator is the degree of genetic distance from humans in the regions where evolution originated (South Africa or East Africa). There are two measures: Genetic distance from the cradle of *Homo sapiens* in South Africa or in Ethiopia. Cavalli-Sforza, Menozzi and Piazza (1994, pp. 75–76) calculated $F_{ST}$ genetic distances between 42 populations. Spolaore and Wacziarg (2009, pp. 480–487) used these data to calculate the genetic distances between countries at the country level and their current populations. $F_{ST}$ distance is first and foremost a measure of the genealogical relationship between populations. If one takes South or East Africa as the "cradle of humanity" as a starting point (León & Burga-León, 2015), then it represents an indicator of ancestry as evolutionary history. Upon closer examination, the data revealed some implausible results: the four countries that were genetically most similar to South Africa were Algeria, Libya, Morocco, and Tunisia, only then came Niger and other countries south of the Sahara. The population of the four North African countries' neighbor, Egypt, is genetically much more distant from the South African population (which is plausible) than that of these four countries. There is a more recent and better data set on genetic distances (Spolaore & Wacziarg, 2018): These data are based on human microsatellite variation from Pemberton, DeGiorgio and Rosenberg (2013) covering 267 worldwide populations. Again, Spolaore and Wacziarg calculated the genetic distances between countries at the country level and their current populations. I used these data to take the genetic distance from Ethiopia (also a cradle of humanity). Data are given for 194 and 173 countries. Both variables have a relevant limitation in terms of content: they reflect genetic differences among modern humans, rather than genetic differences between modern humans and earlier humans in the regions where *Homo sapiens* evolved.

*3.3 Statistical analyses*

We performed descriptive analyses (means, standard deviations, frequencies) as well as correlation and regression analyses. Countries were grouped into three categories: "First World" (Western, Northern, Southern, and Central Europe; North America; Trans-Tasman: Australia and New Zealand; East Asian countries that have been prosperous for decades; wealthy oil-producing countries), "Second World" (former or current Communist countries in Europe and East Asia, relatively wealthy countries in Latin America, less wealthy oil-producing countries), and "Third World" or "Global South" (sub-Saharan Africa, South Asia, and poorer countries in Latin America). The division is somewhat arbitrary, but this poses no major problem,

as our focus is primarily on (a) whether different datasets yield comparable results across country groups, and (b) whether a formula derived from wealthier countries can also predict outcomes in poorer countries. The exact composition of the groups is therefore less critical – for instance, it does not substantially affect results whether Saudi Arabia is assigned to Group 1 or 2, or Colombia to Group 2 or 3. A continuous and more scientific alternative would be GDP per capita.

## 4 Results

### *4.1 Correlations among major international studies and collections on cognitive ability*

We compared results from seven datasets on cognitive abilities using correlations (see Table 1). Correlations between the different variables representing national averages of cognitive ability are generally high (mean $r$=.84). The lowest correlations are observed for the adult literacy study PIAAC, which covers only 33 countries (mean $r$=.78). This is partly because PIAAC relies on a single survey (compared with PISA, which has been repeated every three years since 2000) and partly due to the restricted range of participating countries – no sub-Saharan African or Arab countries and only one Latin American country were included. The highest correlations are found in the Rindermann dataset (mean $r$=.90), likely due to its larger sample of countries and wider range of data, as well as the numerous corrections applied, including adjustments for age and school attendance, regional samples of large countries, and potential errors or fraudulent data. It also contains psychometric intelligence test data, which may explain the comparatively higher correlation with Lynn and Becker.

TABLE 1: Correlations among international collections on cognitive ability.

| | **Student Achievement + Intelligence** (Rindermann, 2022 updated) | **Student Achievement** (Altinok & Diebolt, 2024) | **Student Achievement** (World Bank; Angrist et al., 2021) | **Student Achievement + Intelligence** (Lim et al., 2018) | **Student Achievement** (Gust, Hanushek & Woessmann, 2022) | **Adult Ability** (PIAAC; OECD, 2016) | **Intelligence** (Lynn & Becker, 2019) |
|---|---|---|---|---|---|---|---|
| Stud+IQ (R, 22) | 1 | .91 | .92 | .87 | .91 | .85 | .91 |
| Stud (AD, 24) | | 1 | .91 | .91 | .93 | .75 | .77 |
| Stud (WB, 21) | | | 1 | .87 | .93 | .77 | .79 |
| Stud+IQ (Lim, 18) | | | | 1 | .88 | .75 | .76 |
| Stud (GHW, 22) | | | | | 1 | .75 | .80 |
| Adult (PIAAC, 16) | | | | | | 1 | .79 |
| IQ (LB, 19) | | | | | | | 1 |
| *N* countries (mean *r*) | 199 (.90) | 150 (.86) | 163 (.87) | 192 (.84) | 159 (.87) | 33 (.78) | 129 (.80) |

*Notes*: Number of countries > 100 except for Adult PIAAC. The Rindermann and Altinok data sets include estimated values. The mean of 21 correlations is $r$=.84. All correlations are statistically significant ($p$<.001); last row in parentheses: mean correlation of this variable.

TABLE 2A: Means in international collections on cognitive ability (in SASQ scale)

.

| | **Student Achievement + Intelligence** (Rindermann, 2022 updated) | **Student Achievement** (Altinok & Diebolt, 2024) | **Student Achievement** (World Bank; Angrist et al., 2021) | **Student Achievement + Intelligence** (Lim et al., 2018) | **Student Achievement** (Gust, Hanushek & Woessmann, 2022) | **Achievement Percentage at least Basic** (Gust et al., 2022) | **Adult Ability** (PIAAC; OECD, 2016) | **Intelligence** (Lynn & Becker, 2019) |
|---|---|---|---|---|---|---|---|---|
| **First World** | 477 | 517 | 509 | 512 | 491 | 75% | 501 | 472 |
| **Second World** | 425 | 460 | 454 | 472 | 433 | 52% | 479 | 427 |
| **Third World** | 320 | 375 | 358 | 408 | 333 | 16% | 392 | 325 |
| China | 509 | 450 | 556 | 536 | 553 | 94% | – | 533 |
| UK | 497 | 535 | 519 | 527 | 503 | 81% | 500 | 489 |
| USA | 486 | 528 | 521 | 532 | 490 | 75% | 492 | 472 |
| Russia | 479 | 536 | 518 | 537 | 483 | 76% | 500 | 453 |
| Brazil | 387 | 425 | 414 | 439 | 394 | 34% | – | 401 |
| Indonesia | 369 | 393 | 407 | 424 | 387 | 30% | 392 | 357 |
| Egypt | 353 | 387 | 382 | 401 | 361 | 29% | – | 410 |
| South Africa | 299 | 373 | 352 | 348 | 340 | 16% | – | 361 |
| *Mean* (*SD*) | 379 (84) | 433 (74) | 417 (83) | 445 (59) | 398 (83) | 40% (29%) | 492 (32) | 395 (91) |
| *N* | 199 | 150 | 163 | 192 | 159 | 159 | 33 | 129 |

*Notes*: All results in a student assessment scale (theoretical *M*=500, *SD*=100), but not rescaled (e.g., not OECD countries or UK natives in all scales set at 500) resulting in different means and standard deviations. PIAAC includes only one developing country. "at least Basic": percentage of youth (inside and outside school) with at least basic PISA Level 1 attained standing for routine procedures (average for Sub-Saharan Africa with 40 countries: 6% of youth). Example countries below were selected based on population size, political relevance and availability of data.

TABLE 2B: Means in international collections on cognitive ability (in IQ scale).

| | **Student Achievement + Intelligence** (Rindermann, 2022 updated) | **Student Achievement** (Altinok & Diebolt, 2024) | **Student Achievement** (World Bank; Angrist et al., 2021) | **Student Achievement + Intelligence** (Lim et al., 2018) | **Student Achievement** (Gust, Hanushek & Woessmann, 2022) | **Achievement Percentage at least Basic** (Gust et al., 2022) | **Adult Ability** (PIAAC; OECD, 2016) | **Intelligence** (Lynn & Becker, 2019) |
|---|---|---|---|---|---|---|---|---|
| **First World** | 97 | 97 | 96 | 96 | 99 | 75% | 100 | 96 |
| **Second World** | 89 | 89 | 88 | 87 | 90 | 52% | 97 | 89 |
| **Third World** | 73 | 76 | 73 | 74 | 75 | 16% | 84 | 74 |
| China | 101 | 87 | 104 | 101 | 108 | 94% | – | 105 |
| UK | 100 | 100 | 98 | 99 | 100 | 81% | 100 | 98 |
| USA | 98 | 99 | 98 | 100 | 99 | 75% | 99 | 96 |
| Russia | 97 | 100 | 98 | 101 | 97 | 76% | 100 | 93 |
| Brazil | 83 | 84 | 82 | 80 | 84 | 34% | – | 85 |
| Indonesia | 80 | 79 | 81 | 77 | 83 | 30% | 84 | 79 |
| Egypt | 78 | 78 | 77 | 72 | 79 | 29% | – | 86 |
| South Africa | 70 | 76 | 72 | 61 | 76 | 16% | – | 79 |
| *Mean* (*SD*) | 82 (13) | 85 (11) | 82 (13) | 82 (13) | 85 (12) | 40% (29%) | 99 (5) | 84 (14) |
| *N* | 199 | 150 | 163 | 192 | 159 | 159 | 33 | 129 |

*Notes*: All results in IQ scale (*M*=100, *SD*=15), but not rescaled jointly (e.g., not OECD countries or UK natives in all scales at 100) resulting in different means and standard deviations. PIAAC includes only one developing country. "at least Basic": percentage of youth (inside and outside school) with at least basic PISA Level 1 attained standing for routine procedures (average for Sub-Saharan Africa with 40 countries: 6% of youth).

TABLE 3: Correlations of predictor variables with results of international collections on cognitive ability.

| | **Student Achievement + Intelligence** (Rindermann, 2022 updated) | **Student Achievement** (Altinok & Diebolt, 2024) | **Student Achievement** (World Bank; Angrist et al., 2021) | **Student Achievement + Intelligence** (Lim et al., 2018) | **Student Achievement** (Gust et al., 2022) | **Achievement Percentage at least Basic** (Gust et al., 2022) | **Adult Ability** (PIAAC; OECD, 2016) | **Intelligence** (Lynn & Becker, 2019) | **Mean correlation** |
|---|---|---|---|---|---|---|---|---|---|
| Education adults | .73 | .79 | .81 | .72 | .83 | .83 | .82 | .74 | .78 |
| GDP/c log | .80 | .84 | .83 | .76 | .84 | .85 | .68 | .72 | .79 |
| Politics positive | .50 | .66 | .64 | .55 | .65 | .68 | .57 | .53 | .60 |
| Religion weighted | .61 | .70 | .70 | .62 | .71 | .70 | .67 | .61 | .67 |
| Cranial capacity | .66 | .72 | .65 | .64 | .67 | .70 | .50 | .58 | .64 |
| Gen. d. South Afr. | .57 | .58 | .52 | .59 | .59 | .59 | −.33 | .57 | .46 (.57) |
| Gen. d. Ethiopia | .25 | .24 | .21 | .27 | .32 | .27 | −.29 | .31 | .20 (.27) |
| Mean correlation | .59 (.66) | .65 (.74) | .62 (.73) | .59 (.66) | .66 (.74) | .66 (.75) | .37 (.65) | .58 (.64) | .59 (.70) |

*Notes*: All correlations are statistically significant ($p<.01$). Gen. d. South Afr.: Genetic distance to South Africa (León & Burga-León, 2015); Gen. d. Ethiopia: Genetic distance to Ethiopia (Spolaore & Wacziarg, 2018). In the last row in parentheses the correlations averaged without the two genetic distance measures; in the last column in parentheses for the two genetic distance measures the correlation averaged without the PIAAC study (no data from sub-Saharan African countries).

*4.2 Means in major international studies and collections on cognitive ability*

Tables 2a and 2b present means, standard deviations, and frequencies for the grouped categories "First World," "Second World," and "Third World" (or "Global South"), as well as for eight individual sample countries. It is important to note that the different datasets have not previously been standardized together – this is particularly evident in the UK, where scores vary substantially between SASQ 489 and 535 (equivalent $d$=0.46 or 6.90 IQ points). Nevertheless, the global patterns are broadly consistent across datasets. Countries classified as "Third World," "Global South," or "developing countries" show an average IQ of 73 to 75.[6] The frequently criticized Lynn and Becker dataset reports an IQ of 74, which is exactly the mean of the five broad datasets. According to the Gust-Hanushek-Woessmann dataset, only 16% of children in these countries achieve basic cognitive skills.

However, three additional remarks should be made. First, the PIAAC dataset reports much higher scores for the Second and Third World – approximately 10 IQ points or 60 SASQ points above the other datasets (Table 2b). With only 33 countries (compared to 129–199 in other datasets), PIAAC is highly selective, including primarily well-organized and well-developed countries.

Second, the results reported by Lim et al. appear substantially higher on the SASQ scale than on the IQ scale. This pattern is not limited to the Global South but holds across all countries. For example, the UK scores IQ 99 ($z$=−0.07) versus SASQ 527 ($z$=+0.27). On the IQ scale, the result lies slightly below the scale mean of 100, whereas on the SASQ scale it is above the scale mean. This discrepancy is simply a consequence of the conversion from SASQ to IQ. Three conversion variants were considered: (a) a simple direct conversion (IQ=(((SASQ−500)/100)x15)+100); (b) one that sets Great Britain at around 99 across all datasets; and (c) one that maintains comparable standard deviations. In the present analysis, variant (b) was chosen, but for other cases variant (a) was applied. It should also be noted that the standard deviation of the Lim et al. dataset is much smaller: in the common country sample ($N$=102), $SD$s for Rindermann, Altinok, World Bank, Gust, and Lynn range between SASQ=73 and 88, whereas for Lim et al. the SD is only 65. This compression of the scale pushes lower values upward.

Third, major discrepancies appear when examining individual countries. For example, Russia scores 536 SASQ points in the Altinok dataset but only 483 in the Gust dataset – a difference of 53 SASQ points ($d$=0.53) – even though both collections are based on the same original student assessment studies. Such discrepancies often stem from questionable methods of combining data, such as relying exclu-

[6] One can also consider the country grouping used by Altinok and Diebolt (2024, their Table 2). Here, too, the weakest results are observed in developing countries, with sub-Saharan Africa at 328 SASQ, followed by South Asia (363), the Middle East (394), and Latin America (402).

sively on TIMSS as a benchmark even when PISA data are available, excluding certain survey years, or using only mathematics scores while ignoring reading. Ideally, all high-quality data should be used, with stronger data weighted more heavily than weaker sources. Another striking case is the relatively poor performance of South Africa in the Rindermann dataset: SASQ 299 compared to 373 (Altinok), 352 (World Bank), 348 (Lim et al.), 340 (Gust et al.), and 361 (Lynn & Becker). It should be emphasized, however, that Rindermann applies corrections for both the older average age of pupils and the proportion of youth outside of school in developing countries. Because pupils are often older and school enrollment lower in these countries, substantial downward adjustments may result.

The same applies to the difference between psychometric intelligence and student achievement tests: there are no systematic discrepancies. For example, the First World mean across all collections is 97.29 IQ points compared to 96 in Lynn & Becker's intelligence data; the Second World averages 89.86 versus 89; and the Third World 74.17 versus 74 (excluding PIAAC, which includes only one developing country). However, substantial differences can occur for individual countries – for instance, Iraq scores 355 SASQ points in the student assessments (data from: World Bank, Hanushek) but 429 points in intelligence tests on the same scale.

*4.3 Estimated cognitive ability*

We correlated the five predictor variables, plus two additional ones, with all eight ability measures (Table 3). The strongest associations were found for adult educational attainment (average $r$=.78) and logged GDP per capita ($r$=.79). Religion (culture) and cranial capacity (brain size, evolution) showed average correlations of $r$=.67 and .64, respectively. Positive politics (rule of law, political freedom, and democracy) had a lower correlation ($r$=.60). Genetic distance measures correlate less with ability levels than cranial capacity ($r$=.64 vs. .46/.57 and .20/.27). PIAAC data are unusable here, as there are no data available from sub-Saharan Africa. Among the eight ability datasets, the highest correlations with predictor variables were observed for the Altinok and Diebolt data and the Gust, Hanushek, and Woessmann data (average $r$=.74/.65 and . 66, respectively). No systematic pattern emerges – for example, Lynn and Becker's intelligence data are not more strongly correlated with biological predictors, nor are student assessment results more strongly correlated with education. Overall, the correlations of all ability measures follow roughly the same pattern.

The prediction formula was generated using the ability dataset with the widest country coverage (first column, Rindermann 2018, updated 2022) and the five best correlates. For this purpose, countries from the Global South (Third World, developing countries) were excluded to avoid the prediction formula simply reproducing the input values. The aim is to test whether predictor variables yield similarly high (or low) results as direct test data in developing countries.

TABLE 4: Prediction of national cognitive ability level (student achievement + intelligence; Rindermann, 2022 updated).

| **Predictors** | **5 Variables** | **Formula 1** (reduced, 4, with background) | **Formula 2** (reduced, 3, with politics) |
|---|---|---|---|
| Education adults | .26 | .26 | .48 |
| GDP/c log | .28 | .27 | .35 |
| Politics positive | −.04 | – | .19 |
| Religion weighted | .28 | .26 | – |
| Cranial capacity | .39 | .39 | – |
| Explained variance $R^2$ | 74% | 74% | 52% |
| Mean IQ point difference | 3.11 | 3.11 | 3.87 |
| *N* (countries) | 78 | 78 | 78 |

*Notes*: The cells contain standardized betas ($\beta$). "Mean IQ point difference": If one creates a formula for predicting country IQs from these predictors, then in this sample of countries the measured value differs from the estimated value by this IQ score (in absolute values). Sample excludes Global South. Using the data from Altinok results in the same $R^2$ (74.05%, $N$=78 countries, vs. 74.35%, $N$=71).

In a first regression analysis with all five predictor variables, the explained variance was high (74%, $R$=.86; Table 4, first column) and robust to the choice of ability measure. However, positive politics showed an implausible negative effect ($\beta$=−.04, $r$=.60). Reducing the model to four variables yielded the same explained variance (74%, R=.86; Table 4, second column). The predictor with the highest standardized beta was cranial capacity ($\beta$=.39, $r$=.59). However, education ($r$=.64) and weighted religion ($r$=.74) had stronger correlations with national ability levels, but these predictors (education, GDP/c, and religion) are highly intercorrelated and thus "steal" predictive power from each other (Table A1). Finally, given possible scientific or political reservations about the background variables of culture (religion) and evolution (cranial capacity), we ran a regression including only the strongest correlates – education and wealth – together with positive politics (Table 4, last column). Here, explained variance was substantially lower (Formula 2: 52%, $R$=.72 vs. Formula 1: 74%, $R$=.86), but politics now showed a positive effect ($\beta$=.19, $r$=.60).

TABLE 5: Means estimated for cognitive ability based on amount of adult education, GDP/c and culture/evolution or politics.

| | **Estimated Cognitive Ability** (using education, GDP/c, religion and cranial capacity, formula 1) | | **Estimated Cognitive Ability** (using education, GDP/c and politics, formula 2) | |
|---|---|---|---|---|
| Measure | SASQ | IQ | SASQ | IQ |
| **First World** | 477 | 97 | 474 | 96 |
| **Second World** | 427 | 89 | 429 | 89 |
| **Third World** | 332 | 75 | 350 | 78 |
| China | 488 | 98 | 401 | 85 |
| UK | 481 | 97 | 484 | 98 |
| USA | 499 | 100 | 512 | 102 |
| Russia | 446 | 92 | 446 | 92 |
| Brazil | 403 | 86 | 400 | 85 |
| Indonesia | 355 | 78 | 389 | 83 |
| Egypt | 367 | 80 | 368 | 80 |
| South Africa | 368 | 80 | 404 | 86 |
| *Mean* (*SD*) | 388 (73) | 83 (11) | 397 (65) | 85 (10) |

*Notes*: China in test results huge overachiever (ca. +60 SASQ), also Russia (ca. +30), South Africa in tests underachiever (ca. −80 SASQ). *N*=159 countries. Example countries below were selected based on population size, political relevance and availability of data.

We applied the two formulas for predicting cognitive ability levels (Table 5). The results for First- and Second-World countries are of limited interest, as the prediction formulas were derived from these samples. A comparison of Tables 2 (2a SASQ, first column; 2b IQ, first column) and 5 reveals negligible mean differences between measured and predicted values for these groups (difference≤2 IQ points). In contrast, countries of the Global South (Third World) show slightly higher predicted scores. Using the more predictive formula, which includes religion (culture) and brain size (cranial capacity), the improvement is only 12 SASQ points or 2 IQ points ($z$=0.12), which is negligible. The less predictive but more "politically correct" formula (with education, GDP per capita, and politics) yields a larger increase of 30 SASQ points or 5 IQ points ($z$=0.30).

Examining different world regions, the largest improvements based on prediction values are observed in Latin America, where IQ rises by 8 points from 78 to 86 (Table 6). Depending on the formula used, sub-Saharan Africa increases by 3 IQ points (Formula 1: from 67.44 to 71.57) or by 7 IQ points (Formula 2: from 67.44 to 74.12; Table 6). Applying Formula 2 – but not the more accurate Formula 1 – results in a substantial decline in East Asia (Formula 1: −1 IQ point, from 102 to 101; Formula 2: −10 IQ points, from 102 to 92). This discrepancy is largely attributable

to relative poverty in China, Mongolia, and especially North Korea, as well as the low level of democratic, liberal, and constitutional development in China and North Korea.

Overachievement in test results but underachievement in predicted values (e.g., China, North Korea, Russia) may indicate: (a) biased measurement (e.g., positively selected samples or cheating), (b) potential for greater societal development based on cognitive human capital, or (c) flaws in the prediction formula or in the underlying causal theories.

In general, it is important to note that when values are predicted using regression equations, extreme values tend to be regressed toward the mean. This can be seen by comparing the standard deviations (here reported to two decimal places). The measured cognitive mean from Rindermann (Table 2b, column 1) has a standard deviation of 12.52 (IQ scale), whereas the predicted values (Table 5, columns 2 and 4) have standard deviations of 11.01 and 9.68. The measured mean is negatively correlated with the differences between predicted and observed values (predicted − measured) at $r$=−.52 and −.67 ($N$=157 countries). In other words, countries with high measured values tend to decrease, while countries with low measured values tend to increase. The negative correlation is stronger for the less predictive second formula (lower multiple correlation).

Examining individual countries reveals substantial differences between measured and predicted values. For *China*, estimated values are lower, particularly when applying the reduced Formula 2: the measured IQ is 101 (range 101–108), whereas the predicted IQ is 98 (Formula 1, including culture and cranial capacity) or 85 (Formula 2, including politics). This discrepancy can be interpreted in three ways: (1) the test results are biased, for example because most data come from the developed eastern and southern coastal regions such as Shanghai; (2) societal development – particularly in wealth and political institutions – lags behind cognitive test results; or (3) the prediction formula is incomplete, missing relevant causal variables such as culture and evolution.

Regarding interpretation (1), many inland children in eastern and southern China are not entitled to attend school in their new residence under the Hukou system (Chen & Feng, 2013) and are therefore not tested. Moreover, student achievement studies report results for only a few eastern and southern regions. Consequently, PISA-based values for China have consistently been corrected downward in the Rindermann dataset (SASQ −57, IQ −8.55).

The second interpretation is that the high measured cognitive scores may indicate favorable future development in terms of GDP per capita and political institutions. Over the past decades, China has experienced significantly higher economic growth than other East Asian countries, the West in general, and other emerging nations. In contrast, political development – measured by rule of law, political freedom,

and democracy – has not advanced markedly since the 1990s. This suggests that the theory of the emancipatory effects of education, intelligence, and knowledge may require adjustment; for instance, with regard to human rights, cultural factors appear to be more important than cognitive ones (Rindermann & Carl, 2018). Developments in other East Asian countries since World War II (e.g., Japan, Taiwan, South Korea) support a cognitive-political modernization theory. Time will tell.

*North Korea* has never participated in any student achievement or psychometric intelligence study. The only available data come from the International Mathematical Olympiads (IMO). Reported values are: measured "IQ" 103 (Lim et al. report IQ 93; Table 6), and estimated IQ 96 (Formula 1, including culture and cranial capacity) or 84 (Formula 2, including politics). It is difficult to map the ranking of the top six IMO students to the scales of student achievement or intelligence test studies (e.g., what SASQ or IQ corresponds to a 4th-place IMO ranking?). The ability level of the intellectual classes (95th percentile) was used as a benchmark for comparison, and the estimated value was extrapolated to the mean. As always, the formula was developed using a country sample common to both datasets. Of course, a value based on only six students is far less meaningful than one derived from a representative sample, even after corrections. North Korea has also been disqualified twice for fraud at the IMO.[7] Nevertheless, given the cognitive resources, it can be assumed that political reforms could lead to favorable economic development.

*Russia* (measured IQ 97, range 93–101; estimated IQ 92 in both formulas) is another "test overachiever." There is no evidence of fraud or low representativeness in the samples. It is more likely that societal development (in GDP per capita and political institutions), which is used in our formulas, lags behind the country's cognitive potential. A more positive development appears possible given its cognitive (and other, e.g., mineral) resources.

*Underachievement* in test results but overachievement in the prediction formula is generally observed in developing countries. However, two caveats should be noted: the underachievement in test scores is typically small (maximum 5 IQ points), and when applying Formula 1, which includes culture and brain size, the difference is only 2 IQ points – a negligible gap. The tests do not appear to severely underestimate cognitive abilities relative to what societal and general conditions would suggest. The remaining underachievement may indicate additional potential for long-term increases in cognitive abilities (FLynn effect): with improved environmental conditions, IQ values could rise in the future. Indeed, there are signs that the Global South is catching up (Meisenberg & Lynn, 2023).

[7] https://en.wikipedia.org/wiki/International_Mathematical_Olympiad#Penalties (21.09.2023)

TABLE 6: Means measured and estimated for cognitive ability for selected countries

| | **Measured Cognitive Ability** | **Estimated Cognitive Ability** | **Estimated Cognitive Ability** |
|---|---|---|---|
| Formula | Rindermann 2022 (AD, WB, Lim, Gust, PIAAC, LB) | Formula 1 (with background) | Formula 2 (with politics) |
| UK (Great Britain) | 100 (100, 98, 99, 100, 100, 98) | 97 | 98 |
| USA | 98 (99, 98, 100, 99, 99, 96) | 100 | 102 |
| China | 101 (87, 104, 101, 108, –, 105) | 98 | 85 |
| North Korea | 103 (–, –, 93, –, –, –) | 96 | 84 |
| Russia | 97 (100, 98, 101, 97, 100, 93) | 92 | 92 |
| Brazil | 83 (84, 82, 80, 84, –, 85) | 86 | 85 |
| Cuba | 83 (88, 99, 97, 105, –, 84) | 86 | 89 |
| Ecuador | 78 (84, 81, 76, 83, –, 78) | 85 | 87 |
| Cambodia | 100 (73, 87, 78, 74, –, 100) | 71 | 79 |
| India | 74 (76, 75, 72, 77, –, 79) | 75 | 80 |
| Indonesia | 80 (79, 81, 77, 83, 84, 79) | 78 | 83 |
| Philippines | 77 (72, 75, 70, 78, –, 92) | 86 | 90 |
| Egypt | 78 (78, 77, 72, 79, –, 86) | 80 | 80 |
| Kuwait | 78 (78, 75, 70, 78, –, 92) | 85 | 90 |
| Nigeria | 67 (67, 59, 71, 70, –, 68) | 71 | 75 |
| South Africa | 70 (76, 72, 61, 76, –, 79) | 80 | 86 |
| *Latin America* | 78 (83, 80, 78, 83, 88, 81) | 86 | 86 |
| *Sub-Saharan Africa* | 68 (72, 70, 69, 71, –, 69) | 71 | 75 |

*Notes*: Measured Cognitive Ability: results from the Rindermann data base (2022, revised version of 2018) and in parentheses the values from Altinok & Diebolt, (AD), the World Bank (WB), Lim et al., Gust et al., PIAAC (few countries) and Lynn & Becker (LB); Formula 1: using education, GDP/c, religion and cranial capacity; formula 2: using education, GDP/c and politics. Example countries were selected based on population size, political relevance, availability of data and given qualitative information (see text). China achieves an astonishingly low value in the Altinok & Diebolt (AD) dataset. The data tables were checked, and the reported value for both 2015 and 2020 is 450 SASQ.

However, for individual countries there are substantial gaps (see Table 6). For Namibia, Nigeria, South Africa, the Philippines, and Kuwait, the estimates are much higher than the measured results, especially when using Formula 2, which omits culture and brain size. The superior predictive accuracy of Formula 1 compared to Formula 2 (74%, $R$=.86 vs. 52%, $R$=.72), established in First- and Second-World countries, suggests that omitting background factors such as culture and biology leads to biased results.

For Cambodia, the Philippines, Kuwait, and, to a lesser extent, Egypt, there are larger discrepancies within the measured results. The psychometric intelligence tests

by Lynn and Becker tend to overestimate cognitive abilities compared to data from Altinok and Diebolt, the World Bank, Lim et al., Gust et al., and PIAAC (where available). For Russia, Lynn and Becker report a lower value.[8] However, in some cases, their results are closer to the predicted values (e.g., Russia, as well as Brazil, Kuwait, and South Africa). It is likely that a grand average of all ability test collections would provide a more accurate estimate than any single collection (see Discussion).

Two country comparisons reveal interesting patterns: *Egypt* and *Kuwait* are at roughly the same cognitive ability level despite vastly different economic conditions. With the exception of the Lynn and Becker data, their IQs are 80 or below. However, people in Kuwait are much wealthier: according to our Maddison-based dataset, GDP per capita in 2010 was USD 12,554 in Kuwait versus USD 4,096 in Egypt; more recent data from the World Bank, IMF, and CIA suggest figures of approximately USD 46,000 versus USD 13,800.[9] Wealth itself does not appear to be a decisive factor for intelligence and knowledge. Nevertheless, in the prediction formulas, which include GDP per capita, Kuwait achieves considerably higher results – between 5 and 10 IQ points above Egypt. Mineral resources contribute positively to GDP per capita, but do not seem to affect measured cognitive ability.

Comparing *Nigeria* and *South Africa* reveals a pattern similar to that of Egypt and Kuwait: both countries have comparable cognitive ability levels (around IQ 70, Table 6). They are fairly average for sub-Saharan Africa. However, the estimated (predicted) values are considerably higher for South Africa than for Nigeria. South Africans are also somewhat wealthier: according to our Maddison-based dataset, GDP per capita in 2010 was USD 5,017 in South Africa versus USD 4,740 in Nigeria; more recent data from the World Bank, IMF, and CIA suggest figures of approximately USD 14,672 versus USD 5,495. South Africa also scores higher on the Human Development Index (ranging from 0 to 1, with 1 being the best): 0.597 vs. 0.423 in 2010 and 0.713 vs. 0.535 in 2022 (UNDP, 2010, 2022). South Africa likely benefits from historically established institutions and economic structures, as well as from a small cognitive elite that still exists, both of which are remnants of European influence. Empirical evidence supporting this is that the ability level of the intellectual classes in South Africa is particularly high: the difference between the 95$^{th}$ percentile and the mean is 32 IQ points, compared to 22 IQ points globally and 27 IQ points on average in sub-Saharan Africa.

Underachievement in test results but overachievement in the prediction formula

---

[8] According to an email from David Becker, 7 July 2023, there was a mistake regarding Russia. The correct psychometric result for Russia is 96.15, not 93.18 IQ points.

[9] https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(PPP)_per_capita, retrieved July 1, 2023.

may indicate: (a) measurement issues, (b) better-than-expected development, for example due to natural resources or cognitive elites, or (c) flaws in the prediction formula or the underlying causal theories (particularly for formula 2, which excludes background variables).

## 5. Discussion

Low results in cognitive competence studies in the Global South – or, more pointedly, poor outcomes in intelligence test studies conducted in developing countries – have repeatedly been met with skepticism. At times, these studies have even faced openly hostile polemics. However, the empirical evidence suggests that the test results are not fundamentally flawed.

The tests stem from the tradition of intelligence and student achievement assessments. Data were collected either in single regional studies or in large-scale surveys coordinated by international organizations. The researchers involved represent diverse fields, including psychology, economics, political science, health sciences, educational research, and mathematics. Despite this methodological and disciplinary diversity, the results across paradigms and research traditions are remarkably similar at the level of country groups. Correlations between different collections of cognitive ability data are high to very high (*r*s between .75 to .92; Table 1). Mean scores for Third World countries cluster around IQ 73–76 or SASQ 320–408 (Table 2).

A clear challenge in such studies is the differing calibration of scales. For example, values for England (UK) range between SASQ 489 and 535 (a difference of $d$=0.46 SDs), or between IQ 98 and 100 ($d$=0.13 SDs). However, such discrepancies can be standardized: not only the reference point (England) can be aligned, but also the standard deviations. As the results on the IQ scale and for the benchmark country England indicate, this has been accomplished to some extent. Regardless of the method, test collection, or paradigm chosen, developing countries consistently show weak results.

Nevertheless, there are sometimes substantial differences for individual countries (e.g., Iraq: student assessment results were 355 SASQ points, whereas IQ tests on the same scale yielded 429 points). The same holds true when comparing different test collections: while grouped country results are similar, individual countries often diverge (e.g., for Iraq, Gust et al. reported 378 SASQ, whereas Angrist et al. found 331 SASQ). Using education, GDP per capita, and politics to predict test scores produces somewhat higher estimates for the Global South (about +30 SASQ or +5 IQ points), particularly for Latin America (+8 IQ points, from 78 to 86) and sub-Saharan Africa (+3 to +7 IQ points, depending on the chosen formula). The very fact that PIRLS and TIMSS have developed special, easier tests (so-called prePIRLS) for some

developing countries to better capture performance at the lower end of the ability spectrum – or have administered tests designed for the 4th grade in the 6th grade, or for the 8th grade in the 9th grade (e.g., PIRLS 2011 in Botswana, Morocco, Honduras, and Kuwait; TIMSS 2011 in Yemen and South Africa) – illustrates the magnitude of these ability gaps.

Problems in studies of individual countries should not be swept under the carpet. In the case of Cuba (LLECE) and Kazakhstan (PIRLS and TIMSS), there are suspicions of fraud, as the reported results appear implausibly high. For some countries (e.g., Cambodia and Namibia), measured and predicted values (based on education, GDP/c, etc.) deviate considerably. Many countries have only participated in a single survey or regional study. It would be advisable for these countries to participate regularly in major international student assessment studies (e.g., PIRLS, PISA, TIMSS). This is especially true for India (only two regions took part in PISA 2009), China (data are available only for eastern regions), and, above all, sub-Saharan Africa. To improve outcomes and competencies, expanding educational provision (e.g., more years of kindergarten, extended schooling) and improving teacher training are recommended. Where large discrepancies exist between predictions and test results, the potential for improvement appears particularly high.

Despite all possible criticism of seemingly implausible results from cognitive ability studies, one should not forget that even with "run-of-the-mill" data produced by large international organizations with hundreds or thousands of employees, unusual outcomes occur. Results on wealth (GDP per capita at purchasing power parity, PPP) can also appear odd for individual countries. For example, according to the International Monetary Fund (IMF), Ireland is the richest country in the world, Guyana is richer than France, and Japan is poorer than Andorra. Moreover, there are large discrepancies between different sources for the same variable. For Ireland, for instance, the IMF reports $145,196 per capita, the World Bank $126,905, and the CIA $102,500. Although these figures are not drawn from exactly the same year (2021–2023), one would nevertheless expect the results to be more consistent.[10]

Weaknesses in national and international data collections are more the rule than the exception – for example, in statistics on unemployment, accidents, education, and so on (Hu, 2024). For this very reason, numerous robustness checks have been conducted with cognitive ability data, which have confirmed their usefulness (e.g., Lynn & Meisenberg, 2010). The data on intelligence and knowledge at the country level are apparently not so bad after all. Only those who dislike the findings as a

10 Source: https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(PPP)_per_capita (16 August 2023). When accessed on September 9, 2025, the data for Ireland were much more consistent across the three sources (ranging from 106,000 to 109,000 USD). Apparently, according to two of the three sources, Ireland has suddenly become "impoverished."

whole – especially for political reasons – are unlikely ever to be convinced.

Even without collecting new data – whether through test studies or observational studies – the existing data could be further improved:

A) One approach would be to first average the results from all collections of cognitive ability studies (with or without prior standardization). This would likely yield more accurate overall values.

B) A more sophisticated approach would be to estimate each country's ability levels based on other variables (e.g., using the best prediction formula, formula 1) and, where appropriate, by reference to ethnically and culturally similar countries – often neighboring nations, such as estimating Belgium based on France and the Netherlands (and to some extent Germany), or the United Kingdom based on Canada, Australia, and New Zealand (and partly Ireland and the Netherlands). The less reliable the measured data (e.g., in Cuba, North Korea, or Afghanistan), the greater the uncertainty, the more the measured values should be replaced by weighted estimates. Student achievement studies such as TIMSS and PISA employ a similar approach for estimating individual student achievement, for example, by predicting mathematics scores from reading and science scores and using additional biographical information such as sex, migration status, parents' education, and the scores of other students in the same grade and school.

In this way, the existing results could be leveraged to derive even more meaningful country-level averages.

**Data availability statement:** The dataset used in this study will be made publicly available on ResearchGate.

**Conflict of Interest:** The author declares that there is no conflict of interest.

## References

Altinok, N., & Diebolt, C. (2024). Cliometrics of learning-adjusted years of schooling: Evidence from a new dataset. *Cliometrica*, *18*(3), 691–764.

Angrist, N., Djankov, S., Goldberg, P. K., & Patrinos, H. A. (2021). Measuring human capital using global learning data. *Nature, 592,* 403–408.

Barro, R. J., & Lee, J.-W. (2000). *Barro-Lee data set. International data on educational attainment: updates and implications.* Boston: Harvard University.

Beals, K. L., Smith, C. L., & Dodd, S. M. (1984). Brain size, cranial morphology, climate, and time machines. *Current Anthropology*, *25*, 301–330.

Becker, S. O., & Woessmann, L. (2009). Was Weber wrong? A human capital theory of Protestant economic history. *Quarterly Journal of Economics*, *124*, 531–596.

Bird, K. A. (2021). No support for the hereditarian hypothesis of the Black–White achievement gap using polygenic scores and tests for divergent selection. *American Journal of Physical Anthropology*, *175*(2), 465–476.

Bolt, J., & van Zanden, J. L. (2013). *The first update of the Maddison Project. Re-estimating growth before 1820.* Groningen: Maddison-Project Working Paper WP-4. Data retrieved from www.ggdc.net/maddison/maddison-project/data/mpd_2013-01.xlsx.

Cavalli-Sforza, L. L., Menozzi, P., & Piazza, A. (1994). *The history and geography of human genes.* Princeton: Princeton University Press.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27*(5), 703–722.

Chen, Y., & Feng, S. (2013). Access to public schools and the education of migrant children in China. *China Economic Review*, *26*, 75–88.

Eysenck, H. J., & Schoenthaler, S. J. (1997). Raising IQ level by vitamin and mineral supplementation. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Intelligence, heredity and environment* (pp. 363−392). Cambridge: Cambridge University Press.

Francis, G., & Kirkegaard, E. O. W. (2022). National intelligence and economic growth: A Bayesian update. *Mankind Quarterly*, *63*(1), 9–78.

Gignac, G. E., & Bates, T. C. (2017). Brain volume and intelligence. *Intelligence, 64*, 18–29.

Gottfredson, L. S. (2003). *g*, jobs and life. In H. Nyborg (Ed.), *The scientific study of general intelligence. Tribute to Arthur R. Jensen* (pp. 293−342). Oxford: Pergamon.

Gust, S., Hanushek, E., & Woessmann, L. (2022). *Global universal basic skills: Current deficits and implications for world development* (Nr. w30566). National Bureau of Economic Research.

Hanushek, E. A., & Woessmann, L. (2015). *The knowledge capital of nations.* Cambridge, MA: MIT Press.

Harrison, L. E. (2013). *Jews, Confucians, and Protestants: Cultural capital and the end of multiculturalism.* Lanham: Rowman & Littlefield.

Hu, M. (2024, July 3). National IQ papers must be retracted: Why Kevin Bird and Rebecca Sear don't get it. Retrieved from *Human Varieties* website: https://humanvarieties.org/2024/07/03/national-iq-papers-must-be-retracted-why-kevin-bird-and-rebecca-sear-dont-get-it

Jahrbuch (2004). *Jahrbuch*. [*Yearbook*.] München: DTV.

Jensen, A. R. (1998). *The* g *factor. The science of mental ability*. Westport: Praeger.

Jones, G. (2016). *Hive mind: How your nation's IQ matters so much more than your own.* Stanford: Stanford University Press.

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive *g* and academic achievement *g* one and the same *g*? An exploration on the Woodcock-Johnson and Kaufman tests. *Intelligence*, *40*(2), 123–138.

Kurian, G. T. (2001). *The illustrated book of world rankings*. Armonk: Sharpe.

Lee, J. J., McGue, M., Iacono, W. G., Michael, A. M., & Chabris, C. F. (2019). The causal influence of brain size on human intelligence: Evidence from within-family phenotypic associations and GWAS modeling. *Intelligence, 75*, 48–58.

León, F. R., & Burga-León, A. (2015). How geography influences complex cognitive ability. *Intelligence*, *50*, 221–227.

Lim, S. S., Updike, R. L., Kaldjian, A. S., Barber, R. M., Cowling, K., York, H., … Murray, C. J. L. (2018). Measuring human capital: A systematic analysis of 195 countries and territories, 1990-2016. *The Lancet, 392*, 1217–1234.

Lynn, R. (1990). The role of nutrition in secular increases in intelligence. *Personality and Individual Differences, 11*, 273–285.

Lynn, R., & Becker, D. (2019). *The intelligence of nations*. London: Ulster Institute for Social Research.

Lynn, R., & Meisenberg, G. (2010). National IQs calculated and validated for 108 nations. *Intelligence, 38*, 353–360.

Maddison, A. (2008). *Statistics on world population, GDP and per capita GDP, 1–2008 AD*. Retrieved from www.ggdc.net/maddison/Historical_Statistics/horizontal-file_02-2010.xls.

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, *107*, 407–437.

Meisenberg, G. & Lynn, R. (2023). Ongoing trends of human intelligence. *Intelligence*, *96*, 101708.

Nisbett, R. E. (2009). *Intelligence and how to get it. Why schools and cultures count*. New York: Norton.

OECD (2013). *OECD skills outlook 2013. First results from the survey of adult skills (PIAAC)*. Paris: OECD.

OECD (2016). *Skills matter: Further results from the survey of adult skills (PIAAC)*. Paris: OECD.

Oesterdiekhoff, G. W. (2014). Evolution of law and justice from ancient to modern times. *Journal on European History of Law*, *5*(1), 54–64.

Pemberton, T. J., DeGiorgio, M., & Rosenberg, N. A. (2013). Population structure in a comprehensive genomic data set on human microsatellite variation. *G3 (Bethesda), 3*(5), 891–907. https://doi.org/10.1534/g3.113.005728

Pokropek, A., Marks, G. N., & Borgonovi, F. (2022). How much do students' scores in PISA reflect general intelligence and how much do they reflect specific abilities? *Journal of Educational Psychology, 114*(5), 1121–1135.

Pritchett, L., & Viarengo, M. (2021). *Learning outcomes in developing countries: Four hard lessons from PISA-D*. Research on Improving Systems of Education (RISE).

Rindermann, H. (2018). *Cognitive capitalism: Human capital and the wellbeing of nations*. Cambridge: Cambridge University Press.

Rindermann, H. (2024a). Surprisingly low results from studies on cognitive ability in developing countries: Are the results credible? *Discover Education, 3*(1), 55. [Retracted without consent. Original here: www.researchgate.net/publication/380754429_Surprisingly_low_results_from_studies_on_cognitive_ability_in_developing_countries_are_the_results_credible]

Rindermann, H. (2024b, December 2). International student assessments measure cognitive ability. *OSF Preprints*. https://doi.org/10.31219/osf.io/9stby_v2

Rindermann, H. (2025, 24. June). The one-sided and unsubstantiated retraction of *Discover Education*'s most successful article (by editor Prof. Kerry Kennedy). *Substack (Mind & Image, H. Rindermann)*. https://mindimage.substack.com/p/the-one-sided-and-unsubstantiated

Rindermann, H., & Baumeister, A. E. E. (2015). Validating the interpretations of PISA and TIMSS tasks: A rating study. *International Journal of Testing, 15*, 1–22.

Rindermann, H., & Becker, D. (2018). FLynn-effect and economic growth: Do national increases in intelligence lead to increases in GDP? *Intelligence, 69*, 87–93.

Rindermann, H., & Carl, N. (2018). Human rights: Why countries differ. *Comparative Sociology*, *17*, 29–69.

Rindermann, H., & Carl, N. (2020). The good country index, cognitive ability and culture. *Comparative Sociology*, *19*, 39–68.

Rindermann, H., & Ceci, S. J. (2018). Parents' education is more important than their wealth in shaping their children's intelligence: Results of 19 samples in seven countries at different developmental levels. *Journal for the Education of the Gifted*, *41*, 298–326.

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science*, *29*, 1358–1369.
Rushton, J. Ph., & Rushton, E. W. (2004). Progressive changes in brain size and musculoskeletal traits in seven hominoid populations. *Human Evolution*, *19*(3), 173-196.
Rutkowski, L. (2023). Review of: „Facing the Facts About Test Score Gaps“ from Gregory Connor. *Qeios*. https://doi.org/10.32388/PF3D9Y
Rutkowski, L., Rutkowski, D., & Thompson, G. (2024). What are we measuring in international assessments? Learning? Probably. Intelligence? Not likely. *Learning and Individual Differences*, *110*, 102421.
Sear, R. (2022). *'National IQ' datasets do not provide accurate, unbiased or comparable measures of cognitive ability worldwide* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/26vfb
Spolaore, E., & Wacziarg, R. (2009). The diffusion of development. *Quarterly Journal of Economics*, *124*(2), 469–529. https://doi.org/10.1162/qjec.2009.124.2.469
Spolaore, E., & Wacziarg, R. (2018). Ancestry and development: New evidence. *Journal of Applied Econometrics, 33*(5), 748–762.
UNDP (United Nations Development Programme) (2010). *Human development report 2010*. New York: United Nations.
UNDP (United Nations Development Programme) (2022). *Human development report 2021/22*. New York: United Nations.
Warne, R. T. (2023). National mean IQ estimates: Validity, data quality, and recommendations. *Evolutionary Psychological Science, 9*, 197–223.
Will, M., Krapp, M., Stock, J. T., & Manica, A. (2021). Different environmental variables predict body and brain size evolution in *Homo*. *Nature Communications*, *12*(1), 4116.

## Appendix

Table A1: Correlations among the predictors in the regression analyses presented in Table 4

| | Education adults | GDP/c log | Politics positive | Religion weighted | Cranial capacity |
|---|---|---|---|---|---|
| Column 1 (all variables) | | | | | |
| Cognitive ability (R, 2022) | .64 | .57 | .60 | .74 | .59 |
| Education adults | 1 | .45 | .55 | .58 | .28 |
| GDP/c log | | 1 | .69 | .62 | .05 |
| Politics positive | | | 1 | .82 | .16 |
| Religion weighted | | | | 1 | .43 |

*Notes*. *N*=78 countries for all variables. Correlations and number of observations for all three regressions presented in Table 4 are identical (only the number of predictors vary).