Research

Surprisingly low results from studies on cognitive ability in developing countries: are the results credible?

Heiner Rindermann[1]

Abstract

Developing countries achieve surprisingly weak results in international cognitive competence studies. The results are about one to two standard deviations below the average norm-values obtained in Western countries. The results are so low that they are sometimes difficult to believe. For example, in the World Bank collection, Nigeria scored 262 student assessment points (SASQ, about two and a half standard deviations below the norm 500, equivalent in the IQ-scale to 64 points); in the Lim et al. collection, Yemen scored 336 SASQ (equivalent to IQ 75). These results have triggered opposition, factual-scientific criticism, but also ethical debates and political-ideological objections. We crosscheck the values here by comparing different sources of information, statistical analysis, and on-site reports. Results of the different test paradigms seem to be similar for country groups. However, there are sometimes major differences for individual countries. The same is true for comparing different test collections: results of grouped countries are similar, but single countries differ. Using education, GDP per capita and politics to predict test scores shows somewhat higher results for the Global South (5 IQ points), especially for Latin America (rising from 78 to 86 points in an IQ metric). Finally, observational studies of schools and every day life in several countries of the Global South point to problems in the scope and quality of instruction as well as of thinking. To improve outcomes and competences, it is recommended to expand education (e.g., kindergarten, extension of schooling) and better train teachers. Where there are large discrepancies between predictions and test results, the potential seems to be large.

Keywords Cognitive ability · Nations · Intelligence · Student achievement studies · Developing countries

1 Introduction

Developing countries (also called the Third World, emerging economies, or the Global South) achieve surprisingly weak results in international studies of cognitive competence. The values are about one to two standard deviations below the average norm-values obtained in Western countries. Some of the results are so poor that they are hard to believe. What can be their meaning? For example, in a World Bank student achievement collection [1], Nigeria scored 262 student assessment points (student assessment study quotient SASQ), that is about two and a half standard deviations below the norm of $M = 500$ and $SD = 100$, equivalent to 64 points on the IQ scale. Another example: In the Lim et al. [32] combined student and psychometric intelligence test collection, Yemen scored 336 SASQ (equivalent to IQ 75). However, both conversions $\left(\left(\left(\frac{(x-500)}{100}\right) \times 15\right) + 100\right)$ were overly optimistic as in both studies the UK scored above the mean 500

✉ Heiner Rindermann, heiner.rindermann@psychologie.tu-chemnitz.de | [1]Department of Psychology, Chemnitz University of Technology, Wilhelm-Raabe-Str. 43, 09107 Chemnitz, Germany.

Discover

student assessment quotient, specifically 519 SASQ in the World Bank collection, 527 SASQ in the Lim collection. If the United Kingdom (UK) were set at the Greenwich native score of 100 (which would give a mean score of 98 to 99 for the entire UK population, including immigrants), the results for Nigeria (around IQ 59/60) and Yemen (around IQ 69/70) would be considerably lower. For comparison: Lynn and Becker [35] found higher IQ scores of 68 (Nigeria) and 72 (Yemen) in their intelligence test collection for both countries, with the United Kingdom being the benchmark with an IQ of 98. Finally, based on PISA for Development studies (Programme for International Student Assessment in developing countries) the education economists Lant Pritchett and Martina Viarengo [46], p. 1) came to the estimation that "The vast majority of children in these countries would not reach the minimum targets."

Such low results have triggered opposition, factual-scientific criticism, but also ethical debates and political-ideological objections. For instance, Rebecca Sear [62], p. 1) wrote that Lynn and Becker's dataset were "not fit for purpose" and "not only inaccurate but systematically biased". Or Kevin Bird ([6], p. 472) accused the Lynn data of "systematic bias". Of course, all scientists, including those from other disciplines,[1] may condemn Lynn and Becker's data, but if they were to use data on student achievement, the results tend to be similar or even worse: Instead of intelligence test based IQ 68 (Nigeria, Lynn & Becker) they have to deal with student achievement based IQ 84 or 71 (converted, UK as 99, Lim et al.) or IQ 64 or 59 (UK as 98, World Bank) and instead of IQ 72 (Yemen, Lynn & Becker) they have to deal with IQ 75 or 70 (UK as 99, Lim et al.) or IQ 63 or 57 (UK as 98, World Bank).

There seem to be differences between data sources, but one cannot say—at least based on these two countries—that Lynn and Becker's IQ data systematically come to lower results. Thus, Russell Warne ([67], p. 219), an intelligence researcher who compared the data from Lynn and Becker's psychometric intelligence tests to various test surveys assessing students (e.g. collection from the World Bank), came with the statement "a scientifically worthwhile endeavor" to a much more nuanced position. It should also be mentioned that the student achievement studies seem to have some kind of protective shield against criticism, although there are clearly dubious to fraudulent results here, which are not corrected by most collections (deviating sharply upwards from other studies and neighboring countries: Kazakhstan in TIMSS 2007 and Cuba in LLECE 1997 and 2005–2006; see [49], Appendix, pp. 2, 4).[2]

However, it cannot be denied that results in many developing countries are implausibly low. In particular, the 5% level scores are extremely low for some countries: Student achievement studies typically provide not only mean scores but also ability scores at the 95% level ("intellectual classes", "smart fraction", "skilled human capital") and for the 5% level ("low performer"). In Ghana, at the 5 percent level, students achieve an SASQ of 99 or an IQ of 36 (four standard deviations below the mean); when corrected for age (students are older or younger than the international mean) and extrapolated to the entire age cohort inclusive adolescents out of school the results are at SASQ 58 or IQ 31. The corrections have been even attenuated because we assumed that out-of-school youth in countries with lower achievement levels lose fewer development opportunities by being out of school [49], Appendix, p. 9.

What can be the *meaning* of such low results? Can they be true at all? For example, Richard Nisbett [40] (pp. 214f.) questioned their meaning:

> "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 recently one of the student assessment researchers, Leslie Rutkowski [60]: "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 early versions of the DSM." Nisbett commented: "Obviously something is desperately wrong"—but what is "obvious" to him, and what evidence does he use to support this assessment? Arthur Jensen [25] (pp. 367ff.) mentioned the same problem as Richard Nisbett (and Leslie Rutkowski): Low IQ results seem to have a different meaning for people in the US with European descent than for people with African descent. Based on "my own observation and testing of pupils in special classes" Jensen [25] (p. 367) wrote:

---

[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.

[2] Trends in International Mathematics and Science Study (TIMSS), Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE).

"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."

While for teachers white pupils with an IQ of 70 and lower seem to be children with intellectual disabilities, black pupils were not. However, the validity of IQ measures, e.g., for results in cognitive tasks in school that call for conceptual learning and problem solving was similar for both groups. Jensen explained the differently competent behavior in cognitively less demanding situations—black children being more successful—as an effect of different causes of low intelligence: While in white children an IQ of 70 or less is caused by a severe neurological disability leading to general mental retardation, in black children the low IQ is due to a more normal deviation from the mean. What is clear from Nisbett and Jensen's argument is that evidence is needed from people's everyday world and their behavior in everyday situations.

2 Aims and outline of the study

The aim of the study is to examine whether the relatively low scores in cognitive competence tests in developing countries are true or false. To check the credibility and meaningfulness of the results we will use different methods: We crosscheck the values by comparing different sources of information, statistical analysis, and on-site reports. 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 [1, 19, 32, 35, 49] (World Bank). Third, we predict results using theoretically and empirically relevant predictor variables as education, wealth (GDP per capita) and background factors. The prediction formulas are obtained on samples without the countries to be predicted and are then applied to them. Finally, we use reports from schools on teaching conditions (e.g., [7, 38]) and descriptions of behavior in everyday life.

3 Method

3.1 Data bases

The following six plus one sources were used:

(1) *Cognitive ability (Rindermann data set)*: Data from several *student assessment studies* were combined (PISA, TIMSS, PIRLS, and others from 1991 to 2019; see [49, 55], updated March 2022).[3] The student assessment mean was assembled with data from *psychometric intelligence tests* from Lynn and Becker [35]. Adjustments were made for *age* (depending on the country, students may be older or younger than the international average, older: malus, younger: bonus) and *school attendance* (depending on the country, the percentage of young people attending school may be higher or lower than the international average, higher: bonus, lower: malus). Student data from countries with only regional data were adjusted to be (more) nationally representative (e.g., Shanghai for China). Dubious results were eliminated (Kazakhstan in TIMSS 2007 and Cuba in LLECE 1997 and 2005–2006). IQ estimates that were not directly measured but were based on neighboring countries were also corrected (malus). Great Britain was set at around 99 (Greenwich norm: British natives at IQ 100, for GB including immigrants the national mean is at around IQ 98 to 100).
  Data are given for 199 countries (including estimated but corrected values based on data from Lynn and Becker [35]).

(2) *Student achievement (human capital)*: An average of cognitive ability or cognitive human capital (the authors call it "human capital") based on student assessment studies is presented by Angrist [1]. The World Bank research group

---

[3] PIRLS: Progress in International Reading Literacy Study; study to assess the reading skills of fourth graders. For information on the other studies, see Introduction.

collected data for 163 countries and presents the results on a student rating scale with (traditionally) a mean (norm value) of 500 and a standard deviation of 100.

(3) *Cognitive human capital (learning)*: A score for cognitive ability based on student assessment studies and psychometric intelligence test data is presented by Lim et al. [32]. They call it "learning or education quality". Data are reported for 192 countries on a student assessment scale with (traditionally) a mean (norm value) of 500 and a standard deviation of 100.

(4) *Student achievement*: Gust, Hanushek and Woessmann [22] collected student assessment studies and calculated a mean score, which they call "mean achievement of students on a global scale". A second measure gives the *percentage of youth* (not only students) in a country with at least a *basic ability* level, i.e. being at SASQ 410 to 420 or higher (originally: share of children who do not reach basic skill levels, inverted by us to a positive variable standing for above basic ability level). According to PISA, pupils at this level are able to identify information and carry out routine procedures according to direct instructions in explicit situations. Data are given for 159 countries.

(5) *Adult cognitive ability* (*adult competencies*): The Programme for the International Assessment of Adult Competencies (PIAAC) measured reading literacy, numeracy and problem solving among adults 16–65 years old (years 2011/12 and a further sample in 2014/15; OECD, [41, 43]). PIAAC uses a different scale (mean around 263, $SD = 18$). We transformed the scores (formula constructed with countries that have results in both data collections) to the SASQ and IQ scale. A mean is given for 33 countries.

(6) *Psychometric intelligence*: Lynn and Becker [35], internet: https://viewoniq.org) summarized results from different intelligence tests, which were carried out in different years and with different samples, to form a common mean. Data quality varies between countries, e.g. for the U.S. ($k = 58$) there are much more samples than for Argentina ($k = 5$). Measured results are given for 129 countries (in our study are no estimated values based on neighboring countries). In further steps, they also combined psychometric intelligence test results with results from student assessment studies (here not used).

(7) *Figural intelligence*: One reviewer suggested testing the stability of the results of psychometric IQ tests using fluid culture fair intelligence tests. Can similar results be found when utilizing such a culture-reduced, purely figural method? We therefore used an additional (sub)sample of psychometric intelligence tests consisting only of Raven Matrices tests. The four Raven Matrices tests CPM (Coloured Progressive Matrices, for children), SPM (Standard …, for older children and formerly adults), SPM + (SPM plus, for older teenagers and adults) and APM (Advanced …, for the above average ability levels) were combined to one common country Raven Matrices value. Results are given for 104 countries [4]. Data were missing for a few multi-ethnic societies due to weighting problems, e.g. USA and Brazil. The Raven Matrices are purely figural intelligence tests that load high on a *g* factor and measure reasoning in a culture-reduced way. However, school and education also promote this culture-reduced form of intelligence (e.g. for the CPM [45], or for the CFT, another culture-reduced test [63]).

The Rindermann, World Bank, Hanushek, OECD and Lynn datasets are improvements and extensions of previous versions by these authors and institutions. All use results from international student assessment studies as PISA, TIMSS and PIRLS and many also use data from regional studies as LLECE in Latin America or PASEC (Programme d'Analyse des Systèmes Éducatifs) and SACMEQ (Southern and Eastern Africa Consortium for Monitoring Educational Quality) from Africa.

As suggested by others, we believe that for theoretical (cognitive and substantive) and empirical (correlation, factor analysis) reasons, psychometric intelligence and student assessment studies measure fairly similar constructs (e.g., latent $r = 0.83$ [27], $r = 0.73$, [47, 50]). A (strange) problem is that in student achievement studies it is never really clear what the scale values 500 and 100 stand for—which country sample when was exactly average? However, one can afterwards standardize the values oriented to Great Britain (Greenwich) to 500 and leave the standard deviation unchanged.

3.2 Prediction formula

We used five variables to find the best formula for predicting cognitive ability at the country level. These variables are considered important determining factors, correlates and consequences of high or low cognitive abilities on a country and individual level [9, 22, 23, 26, 49]:

(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 similar basic skills (e.g., filling out an application form)—for the population 15 years and older, from Kurian [30] (p. 349f.), $N = 195$ countries. 2. Percentage of 12- to 19-year-olds for 1960–1985 (in the interval of student assessment studies; they are now adults) who have completed secondary school ($N = 117$), from Mankiw et al. [37]. 3. The mean years of schooling of individuals 25 years or older for 1990, 1995, and 2000 ($N = 107$), from Barro and Lee [2]. All authors used data from the UN or similar sources. The sum (Cronbach $a = 0.93$ in 101 common countries) is given for $N = 195$ countries. School education is a very good predictor of cognitive ability at the individual and national level and at the same time a causal factor for the development of thinking ability and the expansion of knowledge (e.g., [56, 58]) (correlations at the international level see Table 4). However, cognitive gains are not identical in different countries and with increasing length of schooling (diminishing returns).

(2) *Wealth (income, output)*: Log per capita GDP from Maddison in 1990 international dollars [36], 2010 from Bolt and van Zanden [8]. Since an income increase at a lower level arguably has much more impact on quality of life, we used the natural logarithm of GDP. This converts nonlinear, exponential increases in "currency units" to linear increases in more realistic "quality of life units".[4] The more recent Maddison dataset contains information on fewer countries ($N = 117$ versus 159), so we combined—after standardization—the 2010 and 2008 data; $N = 161$.[5] According to the theory of cognitive human capital, cognitive ability increases productivity and wealth, this at the individual as well as at the institutional and country level (e.g., [14, 18, 22, 26, 51]). At the same time, health and environmental stimulation, both of which depend on wealth, affect cognitive development (e.g., [12]). All in all, wealth can be used as a predictor of intelligence at the level on nations.

(3) *Positively valued politics* is the average of three political variables—rule of law, political freedom and democracy—for a longer period back to the 1950s ($a = 0.88$, $N = 201$; [55]). Rule of law, political freedom and democracy are positively influenced by cognitive ability and in longitudinal studies rule of law has a positive impact on ability (e.g., [44, 49]).

(4) *Religion (culture)*: Data on religions (percentage of adherents) were taken from the German Foreign Office (www. auswaertiges-amt.de/www/de/laenderinfos), a country encyclopedia (Yearbook/Jahrbuch, [24]) and the CIA World Factbook (www.cia.gov/cia/publications/factbook). Data, especially for developing countries, were often missing or had to be re-estimated or corrected. For East Asia, Confucianism was assumed to be the predominant religion. Some religions (Buddhism, Confucianism) overlap with general worldviews. An index was developed: Religions were weighted according their support for education, rationality and achievement orientation (see [49], Sect. 10.8). Data are reported for $N = 202$ countries. Religions are worldviews and they take effect through the original message (initial holy text), the exemplary of the religion founder and his role model function, the interpreted and revised doctrine and its changing understanding over time and the lived practice in the present. Several studies show a positive effect e.g. of Protestantism, Judaism and Confucianism (e.g., [5, 23, 49]).

(5) *Brain size* (cranial capacity, evolution): Data on cranial capacity are from Beals et al. [3], p. 304, their Fig. 4. A student assigned numbers to different countries based on map shading. Data for populations alive today (e.g., for Australia and America) were estimated from frequencies in the countries of origin ($N = 184$). Measurement deficits (e.g., results from brain imaging studies of cranial volume would be better data) are more likely to lead to underestimation of correlations. Brain size has a causal effect on intelligence [15, 31], but has the disadvantage of being somewhat endogenous to wealth, i.e. not being totally completely independent of wealth [34]. However, for purely statistical predictions (not for a theoretical explanation), as intended here, this is not a problem. Cranial capacity is not only a measure of brain size, but also an indicator of evolution (which is also useful for predictions) [59, 68].

---

[4] Numerically identical increases at different levels have a different meaning: A $5,000 income increase from $5000 to $10,000 means for the quality of life much more improvement than a $5000 income increase from $60,000 to $65,000. At the poorer level, the $5000 increase has an essential impact on nutrition, health, housing quality and education. At the richer level, income increase means something like a yearly update of smart phones. Therefore, in statistical analyses is usually taken the *logarithm* of wealth indicators [49] (p. 3).

[5] We also checked three other alternative measures of income and wealth, including the Penn World Tables, the Human Development Report, and Credit Suisse, but all of them were less correlated with cognitive ability, which is why we kept the Maddison measure (for the goal of optimizing the prediction).

3.3 Statistical analyses

We performed descriptive analyses (means, standard deviations, frequencies) and correlation and regression analyses. We have grouped countries into the "First World" (West-North–South-Central Europe, North America, Trans-Tasman, countries in East Asia that have been prosperous for decades or are among the old, culturally leading countries, rich oil countries), "Second World" (former or still Communist countries in Europe and East Asia, more wealthy countries in Latin America, poorer oil countries) and "Third World" resp. "Global South" (sub-Saharan Africa, South Asia, poor countries in Latin America) category system. The division is somewhat arbitrary. This is not a major problem, however, as it is more of a general question of (a) whether different data sets produce comparable results for groups of countries and (b) whether a formula obtained in a richer group of countries can also predict results in a poorer group. For this purpose, the exact composition of the country groups is less important. It doesn't matter whether, for example, Saudi Arabia is in Group 1 or 2 or Colombia in Group 2 or 3. A continuous and more scientific variable would be GDP/c. The correlations of two indicators of country development with the international collections of cognitive ability are shown in Table 4.

3.4 Qualitative descriptions

We rely on two sources here: First, descriptions of school, instruction and teachers. Information on teachers partly relates to test results (tests applied to teachers). Second, we bring descriptions of ways of thinking and acting in everyday life from other and own sources that allow inferences about cognitive abilities. But while the results of tests of cognitive ability of hundreds of thousands of students with hundreds of test items can be summarized in one single number, the descriptions of cognitions in real-life situations are much more extensive. If one were to consider such everyday situations as individual test items and require only 20 words for each description, then one would need 200 million words for 100,000 people multiplied by 100 situations.[6] If at all, such descriptions can only be provided using a few examples, which can then always be criticized for being unrepresentative. However, as Nisbett's critique of test scores and Jensen's response to such objections show, there is no way around looking at actions in everyday life. Life, not test results, is the ultimate criterion for validity.

4 Results

4.1 Correlations among major international studies and collections on cognitive ability

We compared the results from six plus one data sets on cognitive abilities with each other via correlations (see Table 1). The correlations between the different variables standing for national means of cognitive ability are high (on average $r = 0.84$ without the last column). The lowest correlations are found for the adult literacy study PIAAC with 33 countries (on average $r = 0.78$). One reason is that we only have data from a single survey here (PISA for comparison has been repeated every 3 years and 2022 since 2000), the other is the reduced range of data, e.g. no country from sub-Saharan Africa or the Arab world participated in PIAAC and only one from Latin America. The highest are found for the Rindermann data set (on average $r = 0.89$). One reason may be the larger data set (more countries and larger range), the other that several corrections have been made in this data set (for age and school participation, for regional samples of larger countries, for possible errors and fraud). The last column contains a subsample of psychometric intelligence tests, the figural culture-reduced Raven tests. The correlations are very similar.

4.2 Means in major international studies and collections on cognitive ability

Tables 2 and 3 give means, standard deviations and frequencies for grouped "First World", "Second World" and "Third World" resp. "Global South" as well as for eight sample countries. It is important to note that the different datasets have not previously been standardized together—this is particularly evident in the UK results, which vary between SASQ 489 and 527 (equivalent to $d = 0.38$ or 5.70 IQ points). However, as far as the global patterns are concerned, it can be seen at

---

6 The longest article ever published in *Intelligence* journal described and analyzed cognition in everyday life: "Everyday life as an intelligence test" by Robert Gordon [17] with more than 110 pages.

Table 1 Correlations among major international collections and studies on cognitive ability

	Student achievement + Intelligence ([49] Rindermann, 2022 updated)	Student achievement (World Bank, Angrist ; [1])	Student achievement + intelligence [Lim, 32]	Student achievement ([19])	Adult ability (PIAAC; [43])	Intelligence [35]	Intelligence Figural (Raven) [4]
Stud + IQ (R, 22)	1	0.92	0.87	0.91	0.85	0.91	0.93
Stud (WB, 21)		1	0.87	0.93	0.77	0.79	0.82
Stud + IQ (Lim, 18)			1	0.88	0.75	0.76	0.83
Stud (GHW, 22)				1	0.75	0.80	0.82
Adult (PIAAC, 16)					1	0.79	0.79
IQ (LB, 19)						1	0.96
N (countries)	199	163	192	159	33	129	104

Number of countries > 100 except for Adult PIAAC. The Rindermann data set includes estimated values. The mean of 15 correlations is $r = 0.84$ (without the partially redundant last column). All correlations are statistically significant ($p < 0.001$)

Table 2 Means in major international collections and studies on cognitive ability (in SASQ scale)

	Student achievement + Intelligence ([49] Rindermann, 2022 updated)	Student achievement (World Bank, Angrist; [1])	Student achievement + intelligence [Lim, 32]	Student achievement [19]	Achievement percentage at least basic [19]	Adult ability (PIAAC; [43])	Intelligence [35]	Intelligence figural (Raven) [4]
First World	477	509	512	491	75%	501	472	486
Second World	425	454	472	433	52%	479	427	429
Third World	320	358	408	333	16%	392	325	323
China	509	556	536	553	94%	–	533	500
UK	497	519	527	503	81%	500	489	495
USA	486	521	532	490	75%	492	472	–
Russia	479	518	537	483	76%	500	453	474
Brazil	387	414	439	394	34%	–	401	–
Indonesia	369	407	424	387	30%	392	357	360
Egypt	353	382	401	361	29%	–	410	410
South Africa	299	352	348	340	16%	–	361	–
Mean (SD)	379 (84)	417 (83)	445 (59)	398 (83)	40% (29%)	492 (32)	395 (91)	396 (90)
N	199	163	192	159	159	33	129	104

All results in a student assessment scale ($M = 500$, $SD = 100$), but not rescaled (e.g., not OECD countries or UK natives in all scales 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 3 Means in major international collections and studies on cognitive ability (in IQ scale)

	Student achievement + Intelligence ([49] Rindermann, 2022 updated)	Student achievement (World Bank, Angrist; [1])	Student achievement + intelligence [Lim, 32]	Student achievement [19]	Achievement percentage at least Basic [19]	Adult ability (PIAAC; [43])	Intelligence [35]	Intelligence Figural (Raven) [4]
First World	97	96	96	99	75%	100	96	98
Second World	89	88	87	90	52%	97	89	89
Third World	73	73	74	75	16%	84	74	73
China	101	104	101	108	94%	–	105	100
UK	100	98	99	100	81%	100	98	99
USA	98	98	100	99	75%	99	96	–
Russia	97	98	101	97	76%	100	93	96
Brazil	83	82	80	84	34%	–	85	–
Indonesia	80	81	77	83	30%	84	79	79
Egypt	78	77	72	79	29%	–	86	86
South Africa	70	72	61	76	16%	–	79	–
Mean (SD)	82 (13)	82 (13)	82 (13)	85 (12)	40% (29%)	99 (5)	84 (14)	84 (14)
N	199	163	192	159	159	33	129	104

All results in IQ scale (M = 100, SD = 15), but not rescaled (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 4 Correlations of predictor variables with major international collections and studies on cognitive ability

	Student achievement + Intelligence ([49] Rindermann, 2022 updated)	Student achievement (World Bank, Angrist; [1])	Student achievement + intelligence [Lim, 32]	Student achievement [19]	Achievement percentage at least basic [19]	Adult ability (PIAAC; [43])	Intelligence [35]	Mean correlation
Education adults	0.73	0.81	0.72	0.83	0.83	0.82	0.74	0.78
GDP/c log	0.80	0.83	0.76	0.84	0.85	0.68	0.72	0.78
Politics positive	0.50	0.64	0.55	0.65	0.68	0.57	0.53	0.59
Religion weighted	0.61	0.70	0.62	0.71	0.70	0.67	0.61	0.66
Cranial capacity	0.66	0.65	0.64	0.67	0.70	0.50	0.58	0.63
Mean correlation	0.66	0.73	0.66	0.74	0.75	0.65	0.64	.69

All correlations are statistically significant ($p < 0.01$)

Table 5 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	0.26	0.26	0.48
GDP/c log	0.28	0.27	0.35
Politics positive	− 0.04	–	0.19
Religion weighted	0.28	0.26	–
Cranial capacity	0.39	0.39	–
Explained variance R^2	74%	74%	52%
Mean IQ point difference	3.11	3.11	3.87
N (countries)	78	78	78

The cells contain standardized betas (β). "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

Table 6 Means estimated for cognitive ability based on amount of adult education, GDP/c and culture/evolution or politics

Measure	Estimated cognitive ability (using education, GDP/c, religion and cranial capacity, formula 1)		Estimated cognitive ability (using education, GDP/c and politics, formula 2)	
	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)

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

first glance that there is little difference in the results between the different data sets. "Third World" resp. "Global South" or "developing countries" have on average a mean IQ of 73 to 75. The frequently criticized Lynn and Becker dataset comes with IQ 74 to an exactly average result of all five broad data sets. According to the Gust-Hanushek-Woessmann data set only 16% of all children in those countries achieve basic cognitive skills. The results of the psychometric intelligence test do not depend on the tests chosen (i.e., the results when using all IQ tests or only the Raven tests are similar).

However, three additional remarks have to be done: First, the PIAAC data set comes for the Second and Third World to much higher results (about 10 IQ or 60 SASQ points higher than the other data sets). The PIAAC data set is with 33 countries (compared to 129 to 199 countries) very selective—especially well organized, well developed countries participate.

Second, the Lim et al. results appear in the SASQ scale much better than in the IQ scale. This is not only true for the Global South, but for all countries, e.g. for UK: IQ 99 (z = − 0.07) vs. SASQ 527 (z = + 0.27). In the IQ scale the result is slightly below the scale mean of 100, in the SASQ scale it is above. This is simply the result of my conversion from SASQ to IQ. There are three variants: (a) A simple direct conversion (IQ = (((SASQ-500)/100)*15) + 100), (b) one that sets Great Britain at around 99 for all data sets and c) one that still produces comparable standard deviations. It was chosen here variant b), but for the others a). And it must be added that the standard deviation of the Lim et al. data is much smaller: In the common country sample (N = 109) the SDs for Rindermann, World Bank, Gust and Lynn are between SASQ = 83 and 89, but for Lim et al. only 64. This causes the lower values to be higher.

	Measured cognitive ability	Estimated cognitive ability	Estimated cognitive ability
Table 7 Means measured and estimated for cognitive ability for selected countries			
Formula	Rindermann 2022 (WB, Lim, Gust, PIAAC, LB)	Formula 1 (with background)	Formula 2 (with politics)
UK (Great Britain)	100 (98, 99, 100, 100, 98)	97	98
USA	98 (98, 100, 99, 99, 96)	100	102
China	101 (104, 101, 108, –, 105)	98	85
North Korea	103 (–, 93, –, –, –)	96	84
Russia	97 (98, 101, 97, 100, 93)	92	92
Brazil	83 (82, 80, 84, –, 85)	86	85
Cuba	83 (99, 97, 105, –, 84)	86	89
Ecuador	78 (81, 76, 83, –, 78)	85	87
Cambodia	100 (87, 78, 74, –, 100)	71	79
India	74 (75, 72, 77, –, 79)	75	80
Indonesia	80 (81, 77, 83, 84, 79)	78	83
Philippines	77 (75, 70, 78, –, 92)	86	90
Egypt	78 (77, 72, 79, –, 86)	80	80
Kuwait	78 (75, 70, 78, –, 92)	85	90
Nigeria	67 (59, 71, 70, –, 68)	71	75
South Africa	70 (72, 61, 76, –, 79)	80	86
Latin America	78 (80, 78, 83, 88, 81)	86	86
Sub-Saharan Africa	68 (70, 69, 71, –, 69)	71	75

Measured Cognitive Ability: results from the Rindermann data base (2022, revised version of 2018) and in parentheses the values from 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)

Third, there are major discrepancies when looking at individual countries: e.g., Russia has in the World Bank collection 518 SASQ points, in the Gust collection 483 (a difference of 35 SASQ or $d = 0.35$). And this in data sets which are based on the same original data from various student assessment studies. Sometimes there are strange methods of combining data, such as using only TIMSS data as a benchmark or data source even when PISA is available, or not using data from all years, or using only the math scale and not reading, etc. All (good) data should be used; good data could also be weighted higher than qualitatively weaker ones. Another notable result is the relatively poor achievement of South Africa in the Rindermann dataset: SASQ 299 (HR) vs. 352 (World Bank), 348 (Lim et al.), 340 (Gust et al.) and 361 (Lynn & Becker). It should not be forgotten here that Rindermann corrects the data for the (older) age of the pupils and the fraction of youth outside the schools. Since pupils in developing countries are often older and many children are not in school, major downward corrections may take place here.

For the difference between psychometric intelligence and student achievement tests the same is true: There are no general differences (First world mean in all collections: 97.60 IQ points, in Lynn & Becker intelligence 96; Second World: 90.20 vs. 89; Third World: 73.75 vs. 74, here excluding PIAAC with only one developing country). Again, there are sometimes major differences for individual countries (e.g., Iraq student assessment results were 355 SASQ points; IQ tests in the same scale were 429 points).

4.3 Estimated cognitive ability

We correlated the selected five predictor variables with all seven ability variables (Table 4). The highest correlations could be found for adults' educational level (on average $r = 0.78$) and logged GDP per capita ($r = 0.78$). Religion (culture) and cranial capacity (brain size, evolution) correlate on average with ability $r = 0.66$ and 0.63. For positive politics (rule of law, political freedom and democracy) the correlation is the lowest ($r = 0.59$). In comparison of the seven ability data sets the highest correlations with predictor variables are found for the Gust, Hanushek and Woessmann and for the World Bank

data (on average $r = 0.73$ to 0.75). No systematic pattern is discernible, e.g. that Lynn and Becker's intelligence variable is more strongly correlated with a biological predictor and student assessment results more strongly correlated with education; the correlations of all ability measures show the same pattern.

To generate the prediction formula, we used the ability dataset with the most countries (first column, Rindermann [49], updated 2022). We excluded countries from the Global South (Third World, developing countries). Only that makes sense, otherwise we would merely find what we put in before with a prediction formula. The aim is to examine whether one would find roughly the same high (or low) results with predictor variables as with the use of tests in developing countries.

In a first regression analysis with all five predictor variables, this led to a high variance explanation (74%, $R = 0.86$; Table 5, first result column). However, positive politics showed an implausible negative effect ($\beta = -0.04$, $r = 0.60$). Reduced to four variables gives the same amount of explained variance (74%, $R = 0.86$; Table 4, second result column). The variable with the highest (standardized) beta is cranial capacity ($\beta = 0.39$, $r = 0.59$). However, education ($r = 0.64$) and weighted religion ($r = 0.74$) showed higher correlations with national ability levels, but the predictors (education, GDP/c and weighted religion) correlate strongly with each other and thus "steal" each other's predictive power (Appendix Table 8). Finally, since there may be some (scientifically or politically motivated) reservations about the two background variables of culture (religion) and evolution (cranial capacity), we performed a regression analysis with the two variables most strongly correlated with ability, education and wealth, and with positive politics (Table 5, last column of results). The explained variance is obviously smaller (F2: 52%, $R = 0.72$ vs. F1: 74%, $R = 0.86$); now politics also shows a positive impact ($\beta = 0.19$, $r = 0.60$).

We applied the two formulas for predicting cognitive ability levels (Table 6). The results for First and Second World countries are less interesting (the prediction formulas were constructed using these country samples). Comparing results from Tables 2 and 3 (Table 2 SASQ first result column and Table 3 IQ first result column) and 5 shows no or very minor mean differences between measured and estimated data for those country groups (difference ≤ 2 IQ points). However, countries of the Global South (Third World) achieve somewhat higher results using the prediction formulas. Nevertheless, using the more predictive formula including religion (culture) and brain size (cranial capacity) the improvement is only 12 SASQ or 2 IQ points ($z = 0.12$). That is negligible. Using the less predictive but more "pc" formula (with education, GDP/c, politics), the improvement is larger (30 SASQ, 5 IQ, $z = 0.30$).

Looking in different world regions, the largest improvements based on prediction values can be observed in Latin America (rising + 8 points in IQ from IQ 78 to IQ 86; Table 7). Depending on chosen formula, sub-Saharan Africa rises by + 3 IQ points (formula 1, from IQ 67.44 to 71.57) or by + 7 IQ points (formula 2, from IQ 67.44 to 74.12; Table 7). Applying formula 2, but not the (more correct) formula 1 leads to huge decline in IQ in East Asia (formula 1: − 1 IQ points, from IQ 102 to 101; formula 2: − 10 IQ points, from IQ 102 to 92). Two causes are responsible for this: The relative poverty in China, Mongolia and above all in North Korea and the low level of democratic, liberal and constitutional development in China and North Korea.

Overachievement in test results but underachievement in prediction formula (e.g., China, North Korea, Russia) may indicate (a) biased measurement (e.g., positively selected samples or cheating), (b) potential for better 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 whenever values are predicted using regression equations, extreme values tend to be less extreme (regression toward the mean). We can see this when comparing the standard deviations (here with two decimal places). The general cognitive mean of Rindermann (Table 3, column 1) has a standard deviation of 12.52 (in IQ scale), the predicted values (Table 6, columns 2 and 4) 11.01 and 9.68. The general cognitive mean negatively correlates with the differences between predicted and measured values (predicted–measured) at $r = -0.52$ and − 0.67 ($N = 157$ countries). Countries with good measured values lose something, countries with weak values gain something. The negative correlation is larger for the worse second formula (less predictive, lower multiple correlation).

Looking at single countries there can be large differences between measured and predicted values. The estimated values for *China* are lower, especially when applying (reduced) formula 2: measured value is IQ 101 (range IQ 101–108), estimated is IQ 98 (formula 1 with culture and cranial capacity) or IQ 85 (formula 2 with politics). This can be interpreted in three ways: (1) Either the test results are wrong (e.g. because too many results come from the developed east and south coast regions, e.g. Shanghai) or (2) the development of society (especially in wealth and politics) is lagging behind the cognitive test results or, finally, (3) the formula is wrong (i.e., missing the relevant causal variables culture and evolution). Regarding interpretation (1), it is also known that many inland children living in the east and south of China do not have the right to attend school (and thus are not be tested) in their new homeland (HuKou system; [11]). In addition, results from the student achievement studies are only published for a few regions in the east and south. Therefore, the values for China from the PISA studies have always been corrected downwards in the Rindermann dataset (by SASQ − 57 or IQ − 8.55). This may also be the reason why the value for China in the Rindermann data set (IQ 101) is at the lower end of

the distribution of the values for China when comparing the data sets (IQ 101–108; Tables 2, 3. The other interpretation (2) is that the good measured cognitive results suggest a favorable future development in terms of GDP/c and politics. China showed in the last decades significantly higher economic growth than other East Asian countries, the west in general and further emerging countries. In terms of political conditions, however, such a development has not been observed since the 1990s in terms of the rule of law, political freedom, and democracy. Perhaps the theory of the emancipatory effect of education, intelligence and knowledge needs to be adjusted; at least for human rights culture is more important than cognitive factors [54]. Developments since World War II in other East Asian countries speak in favor of a cognitive-political modernization theory (e.g., Japan, Taiwan, South Korea). Let's wait and see.

North Korea has never participated in any student achievement or psychometric intelligence test study. The only data given is from International Mathematical Olympiads (IMO). The reported values are: measured "IQ" 103 (Lim et al. offer IQ 93; Table 7), estimated is IQ 96 (formula 1 with culture and cranial capacity) or IQ 84 (formula 2 with politics). It is difficult to put the positioning of the top 6 students (IMO) in a country (or a selection of these) on the scale of the student achievement and intelligence test studies (i.e., rank 4 in IMO represents which SASQ or IQ?). The ability level of the intellectual classes (percentile rank 95) was used as the benchmark for comparison and the herby estimated value was calculated down to the mean value. As always, the formula was developed using a country sample common to both datasets. Of course, a value based on 6 students is never as meaningful as a value based on representative sampling, even after corrections. North Korea has also been disqualified twice for fraud at the IMO.[7] However, it can be assumed that after political reforms there will be a favorable economic development.

Russia (measured 97, range 93–101, estimated 92 in both formulas) is a further "test-overachiever". There is no evidence of fraud or low representativeness of the samples. It is more likely that the development of society (in GDP/c and politics) that we use in our formulas lags behind its cognitive potential. A more positive development seems possible given the cognitive (and other like mineral) resources.

Underachievement in test results but overachievement in prediction formula can be found in general in developing countries; however, two caveats are necessary: The underachievement in test results is generally not large (maximum 5 IQ points) and applying formula 1 including culture and brain size it is only 2 IQ points—a small difference. The tests do not appear to severely underestimate cognitive abilities relative to what societal and general conditions would suggest. The still given underachievement in test results could point to further potential for a long-term increase in cognitive abilities (FLynn effect): With good environmental conditions and their improvement, IQ values could rise in the future. Indeed, there are signs that the Global South is catching up [39].

However, for single countries there are huge gaps (see Table 7): For Namibia, Nigeria, South Africa, Philippines and Kuwait the estimates are much better than the measured results, especially when using formula 2 without religion and brain size. The much better prediction (established in First and Second World countries) using formula 1 than formula 2 (74%, $R = 0.86$ vs. 52%, $R = 0.72$) suggests that omitting the background factors culture and biology leads to biased results.

For Cambodia, the Philippines, Kuwait and somewhat less for Egypt, there are larger differences within the measured results: The results of psychometric intelligence tests by Lynn and Becker seem to overestimate cognitive abilities—compared to the sources from the World Bank, Lim et al., Gust et al. and PIAAC (if given). For Russia, Lynn and Becker get a lower result.[8] However, sometimes Lynn and Becker's results are closer to the predicted values (e.g., for Russia, but also for Brazil, Kuwait and South Africa). It is likely that a grand average of all ability test collections would be more accurate than any single result (see "Discussion").

Two country comparisons reveal interesting conditions: *Egypt* and *Kuwait* are roughly at the same ability level despite vastly different economic conditions. With the exception of the Lynn and Becker data, ability levels are IQ 80 and below. But people in Kuwait are much richer: In our Maddison based data set in 2010 12,554 USD (Kuwait) vs. 4096 USD (Egypt), looking at a list of more current data from the World Bank, IMF and CIA the numbers are around 46,000 USD vs. 13,800 USD.[9] Wealth itself seems not be crucial for intelligence and knowledge. But in the prediction formulas which both use GDP/c Kuwait achieves considerably better results—between 5 to 10 IQ points better than Egypt. Mineral resources have a positive impact on GDP/c but it seems not on (measured) cognitive ability.

Comparing *Nigeria* and *South Africa* shows as for Egypt and Kuwait similar cognitive ability levels, here around IQ 70 (Table 7). Both countries are quite average for sub-Saharan Africa. But the estimated values are much better for South

---

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

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.

Africa than for Nigeria. Also South Africans are (on average) somewhat richer than Nigerians, in our Maddison based data set in 2010 5017 USD vs. 4740 USD, looking at a list of more current data from the World Bank, IMF and CIA the numbers are around 14,672 USD (South Africa) vs. 5495 USD (Nigeria). Also in the Human Development Index (ranging from 0 to 1, the best) South Africa is better: in 2010 0.597 vs. 0.423 and in 2022 0.713 vs. 0.535 [65, 66]. South Africa probably benefits from the institutions and economy built up in the past and from a small cognitive elite that still exists. Empirical evidence in this direction is that the ability level of the intellectual classes in South Africa is particularly high (difference between 95%-level and the mean 32 IQ points) compared to the world average (difference 22 IQ points) or the average in sub-Saharan Africa (difference 27 IQ points).

Underachievement in test results but overachievement in prediction formula indicates (a) measurement problems, (b) better than expected development, e.g. due to natural resources or cognitive elites or (c) flaws in the prediction formula or in the underlying causal theories (this is especially true for the formula 2 without background variables).

4.4 Evidence from the field of education and the cognitive hermeneutics of everyday life

Education: Education is a very important determinant of cognitive ability development; attending one year of school is equivalent to an IQ gain of 3.39 points (e.g., [9, 58]). In the SASQ-scale this corresponds to 23 SAQ points. However, the more knowledge loaden a scale the larger seem to be the gains, in student achievement or crystallized scales around 30–40 SASQ [33, 42], pp. 261f.). Finally, gains are larger in OECD countries than in non-OECD countries (41 vs. 32 SASQ, OECD, [41], pp. 262), one cause could be the quality of schools and teaching. In addition to the scope of teaching, school and teacher quality play an important role [21, 49]. Especially in sub-Saharan Africa, low teacher quality impedes instruction, with detrimental effects on pupils. Using tests from TIMSS 1995, Sandefur [61], his Figs. 4 and 5) showed that teachers in Africa reach a competence level in mathematics comparable to that of eighth-grade students in Europe and East Asia. This is even underbid by reports by Isaac Mbiti [38] (p. 116):

> "Data from a variety of settings suggest that teacher subject knowledge is quite limited. In Kenya, sixth grade math teachers scored about 50 per cent on an externally administered grade appropriate math exam (...). About 40 per cent of teachers in Kenya, 20 per cent of teachers in Uganda, 5 per cent of teachers in Senegal, and 1.2 per cent of teachers in Tanzania had the "minimum knowledge needed to be effective."

Similarly Bold et al. [7], pp. 191–193, quotations compiled from their text and tables):

> "For the language subject area, we formally define 'minimum knowledge for teaching' as marking at least 80 percent of the items on the language test correctly. Only 7 percent of the language teachers meet this minimum. … A mathematics teacher is defined as having minimum knowledge for teaching if he/she scores at least 80 percent on the tasks covered in the math curriculum up to grade 4. … 15% can solve [a] difficult math story problem. … Almost one-quarter of the teachers cannot subtract double-digit numbers and one-third of the teachers cannot multiply double-digit numbers."

Just as one cannot believe the results of student achievement tests or intelligence tests, whether expressed on a 500-point SASQ scale or a 100-point IQ scale, one cannot believe these reports: adults or even teachers cannot subtract the number 19 from 56?

One reason for low competence test results is that children often do not attend classes continuously, i.e., they miss one or two days a week or do not come for a year or two, e.g., because they have to work (often for the family) or because there is no money for school and transportation to school. On top of that, teachers are often absent. Again Bold et al. [7], (p. 187, 189):

> "In each school, during a first announced visit, up to ten teachers were randomly selected from the teacher roster. At least two teaching days after the initial survey, an unannounced visit was conducted, during which the enumerators were asked to identify whether the selected teachers were in the school, and if so, if they were in class teaching. Both assessments were based on directly observing the teachers and their whereabouts. … Averaging across countries, 44 percent of teachers were absent from class, either because they were absent from school or in the school, but not in the classroom."
>
> "Moreover, even when in the classroom, teachers may not necessarily be teaching. We carried out classroom observation as part of the survey, recording a minute-by-minute snapshot of what the teacher was doing, for a randomly selected fourth-grade mathematics or language class. The percentage of the lesson lost to nonteaching activities

varied from 18 percent in Nigeria, …, to 3 percent in Uganda. … Students are taught, on average, 2 hours and 46 minutes per day, or roughly half of the scheduled time."

The same problem occurs in other countries of the Global South. E.g., in India, in unannounced visits to Indian primary schools 25% of teachers were absent and only 45% were actively engaged in teaching at the time of the visit ("Three unannounced visits were made to each of over 3700 schools."; Kremer et al. [29], p. 659). And when teachers are there, it's often not good either:

"A team of researchers who visited schools in India… found some teachers who kept schools closed or nonfunctional for weeks or months at a time, drunken teachers, and a headmaster who expected the students to perform domestic chores and babysitting. Sexual abuse of female students by male teachers is a problem in several countries. To the extent that teachers do have incentives, these incentives are often focused on exam scores. Teachers often instruct by rote, sometimes copying from textbooks onto the blackboard and having students copy from the blackboard onto notebooks or slates." (Glewwe and Kremer [16], p. 962)

For many of us, the incidents described are just as hard to imagine as the low test results—a school without teachers! Could you imagine going to a hospital and not finding a doctor there?[10] In the light of such conditions, it is no wonder that student achievement and test results are very poor.

Everyday life: Now, however, one could object that what has been described here stands for "peculiarities", institutional deficits, in the field of education, but says nothing about the ability level of people in everyday life. What we need are descriptions of thinking in daily life. We found a first hint in descriptions of teachers' abilities. According to Bold et al. [7] (pp. 191–193), only 7% of language teachers reached the minimum proficiency level, in mathematics it had been 68%. Teachers are certainly among the better educated and, on average, more intelligent adults in a society (they usually have a high school diploma and have studied, around IQ 109, [69]). A study done in Nigeria and Germany (done by a student, Luisa Falkenhayn, Rindermann et al. [57]) found a correspondence between a pattern in IQ based on Raven Matrices ($d = 1.48$) and epistemic rationality measured by everyday life examples (e.g. "Blowing wind is a phenomenon caused by chains of causes and consequences subjected to natural laws." [positive], "Only guilty persons can die because of poison." [negative] or " God is able to put money inside a person's empty pocket if that person really needs money at that moment." [negative]) ($d = 1.85$).

Before we start with descriptions from two countries, two qualifying notes in advance: (1) Certainly one can find evidence of cognitive problems in everyday life everywhere. It is a hobbyhorse for opponents of the ruling American president to find signs of stupidity or even dementia in him (e.g. in Ronald Reagan, George W. Bush or Joe Biden). At the country level, we have described such examples of cognitive problems in everyday life for Germany [52]. (2) In view of the current climate in science and society, it can be considered disrespectful to describe cognitive deficits in others. However, the telos of truth always applies as orientation in epistemic-scientific processes, not a political norm (e.g., [20, 49], pp. 211ff).

Nevertheless, it is never wrong to start with positive examples. In 1985, when I was 19 years old and traveling upstream the Rio Cauaburi (far west in the Yanomami region in Brazil) in a boat with a group of gold miners and South American Indians, I asked an Indigenous man, about 40 years old, in which direction the road from São Gabriel da Cachoeira to Cucuí ran. The Cauaburi is an endlessly meandering river, nothing can be seen of the landscape, only the forest walls to the right and left, like in a tunnel, only without a ceiling. The sky was overcast, and even if the sun were shining, it would be too vertical at midday to help with orientation. A road (BR-307) runs somewhere 50–100 km away to the west from south to north. So my question was aimed at where the west is. And I asked the Indian (in Portuguese) several times in different bends of the river, myself with the hidden compass in my hand. He always pointed his arm properly to the west. How could he know that? He must have noticed every bend in the river for days—we have been going for days—or every bend since we last saw the lower sun. Incredible!

Another day I went on a longer trip with another Indigenous man, about 18 years old, and a gold digger. The chief had said that a plane with gold had crashed into the forest and we should look for it. In itself a crazy idea, because in the jungle you can see only a few meters away and thus find nothing. Besides, you never know where you are. The terrain was always flat and there was nothing to see except leaves (from the bottom, front, right, left and back) and tree trunks. But the interesting thing was that after six hours of walking through the undergrowth, we came out exactly where we started (and it was not the same way back). How did the Indian do it?

---

[10] In fact, there are studies examining the presence of health workers. In India, for example, 40% were absent during an unannounced visit to health clinics [10] (p. 92).

Finally, Amerindians have extraordinary visual perception skills. They see every bird in the distance in the branches, and one Indian (another) found me in the Cauaburi River hiding among leaves and behind branches in the water. And this despite the fact that the Indian was sitting in a fast-moving motorboat and had seen nothing but the monotonous forest walls to the right and left of the shore for hours.

To put it in context, none of this is measured by our cognitive tests, which require mostly abstract thinking (and language and knowledge in student achievement tests and in tests of crystallized intelligence). 16 years later, in 2001, when I tried to test Yanomami Indians (in the very east of their region in Brazil) with the Raven SPM, they did not understand what was required, not even young Yanomami men with some knowledge of Portuguese and some experience of modernity gained through visits to Brazilian cities (Boa Vista). Furthermore, Indigenous peoples are not representative of Brazil or Latin America. However, these examples show how different people can be from us. Cognitive abilities seem to vary greatly from country to country and culture to culture and people to people. People are not the same, they are different.

A scientifically (for questions of education and intelligence) interesting country is *Cuba*. Cuba never participated in the large international student assessment studies TIMSS, PISA and PIRLS. However, there exist results from the regional Latin American Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) study. In the LLECE scale, the 1997 average of all participating countries is 259 points ("LLQ"), the range without Cuba is 230 (Honduras) to 271 LLQ (Argentina). But Cuba reached 350 LLQ, almost three standard deviations above the Latin American mean including Cuba ($d = +2.77$) or two standard deviations above the next best achieving country Argentina ($d = +2.40$)! In 2006 the pattern was similar: Cuba the best ($d = +2.54$ higher than the Latin American mean including Cuba). Who can believe that? Cuba and socialist countries in general are notorious for data massaging [28]. According to the mentioned study, a more detailed analysis of Cuban economic statistics revealed numerous omissions and contradictions. Often, e.g., in the area of education (alphabetization) and infant mortality, terms are used differently than is customary internationally. On these two indicators, Cuba seems to be the leader in Latin America or would even be better than the U.S.! According to the authors, no Cuban would believe in official statistics, all figures were fake. The only amazing thing were that media and others abroad would believe it.

In our Table 7, Cuba has according to the Rindermann data set an IQ of 83, in Lynn and Becker 84, but in the student achievement collections of the World Bank, Lim et al. and Gust et al., there are values between 97 and 105! These collections have used the LLECE results, which I have deleted. According to the Gust et al. data set, Cubans (SASQ 532 and IQ 105) are the fifth smartest people in the world, smarter than, for example, Japanese, Koreans, Finns, Dutch, and North Americans!

I was several times in the 1990es in Cuba. This provided an opportunity for observations relevant to cognitive ability. One example: Many clocks were broken in Cuba, e.g. in private rooms, but also in train stations and on the streets. That in itself stands for economic development, but it gave rise to a relevant observation: Out of a sudden idea without any major ulterior motive (I hadn't done any research into education and intelligence at the time), I came up with a riddle or brain teaser:

"Imagine", I pointed to my analogue wristwatch with dial and hands, "a clock is broken. It no longer works. It shows the same time all day long. How many times a day does it show the correct time?" (originally asked in Spanish)

This question relates to daily life in Cuba because, as mentioned earlier, many clocks did not work. Of a sample of about 20 people, 75% said, "this clock never shows the correct time", 15% said "once" and 10% "twice". A Cuban teacher with better command of Spanish than me came in her survey to a similar distribution. I was perplexed and did not understand. Why can't they understand such an easy task? When I asked in Cuba tourists from Spain, Italy or Germany, they always came up with the answer. But they also often began to suspect ulterior motives, were uncertain, what trick was behind this question? For example, it was asked whether it related to east–west or west–east air travel and other complex and verbose considerations.

At that time, I was not familiar with cognitive research. In Munich, I spoke to Professor Oerter, a renowned professor of developmental psychology and interested in cross-cultural research. He said that according to Piaget's theory, individuals who have problems with the clock question have not developed the last or penultimate cognitive developmental stage that enables abstract-formal thinking and cognitive perspective taking.

I then asked a teacher in Germany to ask her pupils the same question (April 2005, German Hauptschul teacher, vocational track school). In 7th grade, 12 out of 26 pupils (46%) knew the correct answer, in 8th grade 15 out of 27 pupils (56%) and in 10th grade 4 out of 5 pupils (80%). Thus, the Cubans surveyed would correspond to 6th or 5th vocational graders

in vocational schools in Germany. If it is not an unrepresentative sample, this would correspond to an IQ by Greenwich standards of about 75–70 in Cuba.

Of course, my sample could not be representative. However, I also made a questionnaire study in Cuba [48]. 75 persons participated, the majority between 18 and 26 years old. About 30% were (university) students or worked in skilled occupations, and about 20% described themselves as white. The questionnaires were distributed by Cubans (in September 1997 in the region of Santiago de Cuba), and the individuals received $1 in appreciation (the minimum monthly wage was about $20). In the answers, the verbal and mathematical aspects are of interest here:

On the one hand, a fairly liberal approach to Spanish spelling rules was noticeable: r-l, x-c-s-z, g-j, v-b and y-ll were swapped. E.g., "cobral" instead of "cobrar", "extraño" instead of "estraño", "cojer" instead of "coger", "cayar" instead of "callar". "H" at the beginning of the word and "s" at the end of the word are often omitted (e.g., "Abana" instead of "Habana", "todo" instead of "todos"). Punctuation marks were rarely used. Capitalization was missing. The respondents simply wrote the way they pronounce the words.

On the other hand, contradictory answers to duplicate questions and implausible information (contradictions between different questions) were often obvious. For example, the number of intimate relationships with tourists in the last 8 years and in the last 4 weeks was asked. Although the women and men were interviewed in a place with no tourists, most gave relatively high 4-week numbers (e.g. 10 in the last 4 weeks) with relatively low total numbers (e.g. 20 in the last 8 years). Even if "time off" (home stay, off-season, long-term relationship, started dating only a few weeks or months ago) is included, 20 would be too low. In the case of five persons, the stated age and date of birth contradicted each other (a younger age stated than can be calculated from the reported date of birth).

Therefore, there is clear evidence of educational problems and cognitive deficits. Language was not mastered correctly, contradictions between statements were not recognized (or were considered irrelevant) and there were problems with arithmetic. Finally, they did not seem to care what potential readers of their responses thought about them or did not think about it at all (no perspective-taking). The chosen sample was certainly not representative, there were no old people (however, on average they are less educated than young people), and the middle class was underrepresented. Nevertheless, supporting evidence is lacking for a country that appears to be at the top of both officially reported formal education and student achievement levels, rather the evidence suggests weak levels. Cognitive ability levels at around SASQ 525 or IQ 100, as reported by the World Bank, Lim et al. and Gust et al., are certainly too high. Since the samples described here appear to be somewhat below average (the middle class was underrepresented), an IQ of 75 to 70 is probably not correct either. Scores between IQ 83 and IQ 89, as measured or estimated for Cuba (Table 7), seem more appropriate.

A second country in Latin America that we want to take a closer look at is *Ecuador*. Unlike Cuba, there are no exceptional results for Ecuador, its results are in line with the Latin American average, and the values of the different studies do not differ much (Table 7; e.g., Ecuador in Rindermann data set IQ 78, Latin America IQ 78). The by the prediction formulas estimated values, which are better on average for Latin America, are also better for Ecuador and are again close to the regional mean (Ecuador measured 78, predicted 85 and 87, Latin America measured 78, predicted 86 and 86; Table 7). We also have data (Raven SPM, PIRLS-Reading, TIMSS-Mathematics, 161 9 to 14 year old pupils in 2009) from on own study with Indigenous Ecuadorian children. The mean score here was IQ 71, in the remote highlands it was IQ 65, near Quito IQ 76, and in the Amazon lowlands IQ 74 [53]. The result near Quito (IQ 76) was close to the national average (IQ 78, Table 7), that from the highlands not. As in Cuba with the observational data, our sample was not representative (too few people from the middle class). However, regardless of whether you look at the national IQ or student achievement measurements, the national estimates (Table 7), as well as the data from the Indigenous sample, they are all low, not as low as in Africa, but still weak. Unlike in the case of Africa, however, there have never been any protests from the scientific camp or from certain political circles against studies on cognitive abilities and their results in Latin America. Latin American scientists have even conducted intelligence testing studies themselves and no doubts have been raised (e.g., [13]).

However, are these values credible? What does an IQ around 78 mean for everyday life? As for African countries and India, we also have information on teacher absenteeism in Ecuador. During unannounced visits, 14% of the teachers were absent [10] (p. 92). The special thing about Ecuador was that many teachers were absent on both visits—thus they were so called "ghost workers", persons who never teach. This certainly has adverse consequences for the development of children.

In 2020 and 2021, I was in Ecuador three times (staying two weeks each time), I visited the two big cities (Quito and Guayaquil), was in the sparsely populated highlands, in the Amazon lowlands, in the cloud forest and on the dry coast. Many experiences of interacting with local people and coping with the demands of everyday life were completely

unremarkable, some of them excellent. For example, a corona test center in Quito worked particularly quickly, and the result was received on the same day (compared to one to two days in Germany). However, there were also different experiences:

In a hotel in Baños de Agua Santa there was a note at the reception: "I Come back in 15 min." ("Regreso en 15 minutos. Gracias.") Of course, this information does not help. The only thing that was certain was that someone would come back between 0 and 15 min (if the statement were true). A change of perspective was missing, to put yourself in the position of the hotel guest, who does not know when the hotel employee left the reception. Like many occurrences, this one is ambiguous: someone particularly clever could put down such a note and then stay away for hours. But that is rather unlikely, it was a well-managed hotel.

Most of the time, as a traveler you only have limited contact with the locals. A brief conversation while shopping, at the hotel reception, at the hairdresser. That was it. Only those travelers who have relatives or friends in the destination country have a deeper insight. However, there will be a selection effect here—who do scientists have as acquaintances? An exception could be a longer taxi ride: In February 2021 we made two longer trips with a taxi driver, first an afternoon in Baños de Agua Santa, then a full day from Baños to Puyo including various excursions. Here one has contact with a fairly average person from the country, not an academic, but also not a person from the economic underclass—after all, he has a car, can drive it, pays for fuel and repairs, manages income and expenses. And as a passenger by his side, there was plenty of time to chat with him, more than a whole working day. Examples from the communication (originally in Spanish):

- As far as politicians and their evaluation were concerned, he only expressed sympathy and antipathy. There was no statement about their political positions and the like. The fact that one had already lost twice in elections (Guillermo Lasso) was the only attribute he mentioned.
- He claimed in the car while driving (not before deciding to go with him), the car had air conditioning. I was sitting next to him during this. But his car did not have air conditioning. How can he assert something that the communication partner can refute in the simplest way, something that is so obviously wrong?
- He claimed the rain would stop in 15 min, he said this every half hour.
- He claimed the day before that we would drive to the Rio Napo. I checked that with the help of a map. In my opinion, that was way too far. Then at 3 p.m. on the day of the ride, he said it was too far. But this is not a completely clear example. He could have said it the day before, just to get me as a customer, and not because he could not cognitively assess time and distance appropriately.
- He called a (traffic) policeman "comisario". However, a "comisario" is no regular traffic cop, but either a senior police officer or someone who investigates crime.
- He called a red banana "hierba", herb or grass.
- He did not buckle up with the seat belt, only in cities with possible police contact. In his opinion, that car passengers do not have to wear seat belts (for safety reasons), but only in the event of a check (to avoid a penalty).
- He parked the car in the middle of the path. As a result, he obstructed others coming after him and then had to move his car. Moreover, he made this mistake twice in a row.
- He did not know a single bird by name, although he made many trips with tourists to rainforest areas.
- At the end of the day, he complained that there had been no lunch. But at the same time he had never driven us to an inn or everywhere we asked there was nothing (it was during the corona period).
- He first called crocodiles "caimán" and later "lagarto" (lizard, also caiman in Latin America). The term "lagarto" for a crocodile is therefore not wrong, but misleading. However, it becomes incorrect when "caimán" was always used before and the taxi driver agreed with my (checking) inquiry whether it was a herbivorous iguana ("iguana"). Finally, the term "lagarto" in Latin America for a crocodile is also a sign of collective misjudgment. It represents an overextension of the term. A dolphin is also not a fish and an earthworm is not a snake.

In addition to misjudgments and misstatements, there was no discernible learning and no discernible reflection by him on his own speaking and thinking, no reflection on how what he said might be received by the other (perspective-taking). Intelligence and conscientiousness can also play a role together here, influencing each other.

Further examples:

- In Guayaquil, on a Sunday, I could not withdraw money from several ATMs (no money came out), but it was debited from my accounts. It took several hours of time (of me and bank employees in Germany, otherwise in Europe, Ecuador or the USA) to undo this. This never happened before in any other country.
- In general, it is often too noisy: On the bus, the movies and music are too loud. Also, the movies shown on the bus for all passengers, including children, are often too violent (e.g., John Wick). (Both represent too much sensory stimulation or low sensitivity). Buses, trucks, cars and motorcycles are sometimes extremely noisy. There is a lot of honking and often you can hear car alarms. Commercials and election advertisements from trucks blasting the streets with loudspeakers are very loud. Metal blinds of the stores that go up or down early in the morning, in the evening, and at night are very noisy. People speak loudly. Houses do not have soundproof windows, even on roads, only single-pane windows. Engine noises and voices penetrate uninsulated into apartments. None of this seems to be a problem for them. However, research shows a negative impact of noise on cognition [64].
- Finally, this was the first time in my life that I have witnessed a racist insult: In the La Marín bus station in Quito, a "half-black" man, about 25 years old, said to a darker black girl (also mixed race) about 5 years old, "Your mom is black and ugly, black as oil." How can an adult say such a thing to a child?

All this can also be compared to swimming skills. In three different public swimming pools near Quito, no one could be observed who could swim professionally (for example, breaststroke or crawl). It must be added, however, that admission to an outdoor pool costs $4 per adult and $3 per child (ages 2–12; 2021 Ilaló near Quito); there is also no family pass. Even in direct comparison with Germany (without taking into account differences in average income), this is more expensive.[11] Hardly anyone can afford to go swimming; moreover, there seem to be no professional swimming lessons.

5 Discussion

Low results in cognitive competence studies in the Global South, or, to put it more pointedly, poor results in developing countries in intelligence test studies, have repeatedly met with reservations. Sometimes there was openly hostile polemics against these studies. However, empirical evidence suggests that the test results are not that wrong.

The tests come from the tradition of intelligence or student achievement tests. The surveys were conducted in single regional studies or coordinated by worldwide organizations in large-scale surveys. The scientists involved come from the fields of psychology, economics, political science, health sciences, educational research and mathematics. Despite this wide diversity and breadth of methods and disciplines, the results of the different paradigms and research camps appear to be similar for the country groups. Correlations between different cognitive ability test collections are high to very high (*r*s between 0.75 and 0.92; Table 1). The mean scores for Third World countries hoover around IQ 73–75 or SASQ 320–408 (Tables 2, 3). An obvious problem with such studies is the different calibration of the scales. For example, the values for England (UK) are between SASQ 489 and 527 (a difference of $d = 0.38$ SDs), in IQ between IQ 98 and 100 ($d = 0.13$ SDs). But this can be easily standardized, not only the reference point England can be equated, but also the standard deviations. As can be seen from the results of the IQ scale and the benchmark country England, this has (to a certain extent) been done. No matter which method one chooses, which test collection or which test paradigm, developing countries achieve weak results.

However, there are sometimes major differences for individual countries (e.g., Iraq student assessment results were 355 SASQ points; IQ tests in the same scale were 429 points). The same is true for comparing different test collections: Results of grouped countries are similar, but single countries differ (e.g., for Iraq, Gust et al. found 378 SASQ, but Angrist et al. found 331 SASQ). Using education, GDP per capita and politics to predict test scores shows somewhat higher results for the Global South (30 SASQ or 5 IQ points), especially for Latin America (rising 8 IQ points from 78 to 86 points in an IQ metric) and sub-Saharan Africa (+ 3 to + 7 IQ points, depending on chosen formula). The very fact that PIRLS and TIMSS have either developed special (easier) tests (called prePIRLS) for some developing countries in order to better differentiate in the lower ability range, or that they have used tests for the 4th grade in the 6th grade or for the 8th grade in the

---

9th grade (e.g., in PIRLS 2011 for Botswana, Morocco, Honduras and Kuwait, in TIMSS 2011 in Yemen and South Africa), speaks for larger ability gaps.

One could now criticize all of this in that only artificial tests were ever used. Maybe there are just different test-taking skills and the real intelligence that shows in mastery of everyday life might indicate a worldwide similarity in cognitive abilities. A test result is not reality, just a description of reality.

But also observational studies—initially at schools that are causally important for the development of cognitive abilities—in many countries of the Global South point to problems in the scope and quality of teaching. Observations in everyday life (spelling, written language, argumentation, behavior) indicate clearly recognizable weaknesses (or how to put it?) in the cognitive area. For example, failure to take perspectives, not evaluating politicians according to political criteria, making obviously false claims.

All in all, it appears that the test results are credible. While the results are often surprising, the available empirical evidence suggests that they are broadly accurate. One reason for doubting the data may be that Western scientists too often work with WEIRD (western, educated, industrialized, rich and democratic) samples and too rarely spend time on the ground in developing countries. It is simply necessary to talk to the local people, for a long time and in their language. However, this will always only be possible on a random basis and test studies with hundreds to thousands of participants will always be broader and more representative. Certainly, it will always be possible, especially for general skeptics of the cognitive ability paradigm, to criticize details of the observations. A criticism that is certainly correct is that in casual observations it is more the unusual that stands out and is worth reporting and not what one also knows from home. So I shared here in the descriptions my experience in Ecuador with a receptionist in Baños de Agua Santa, but not the ones in hotels in Mindo (nothing special) and Riobamba (shop-talk about professional mirrorless cameras). One have to systematize observational studies: Train several people in observation and then send them to, for example, four countries with the greatest possible cognitive difference. The descriptions above can only provide a first indication of what type of behavior should be observed.

Problems in the studies on individual countries should also not be swept under the carpet. In the case of Cuba (LLECE study) or Kazakhstan (PIRL and TIMS study), there is a suspicion of fraud (transmitted results are far too good). And there are considerable deviations between measured and predicted values (based on education, GDP/c etc.) for individual countries (e.g., Cambodia and Namibia). Many countries have only participated in single surveys or regional studies. It would be good for these countries to participate regularly in the major international student assessment studies (e.g., PIRLS, PISA or TIMSS). In order to improve outcomes and competencies, it is recommended to expand education provision (e.g., more children more years in kindergarten, extension of schooling) and improve teacher training. Where there are large discrepancies between predictions and test results, the potential seems to be particularly large.

Despite all possible criticism of implausible-sounding results from cognitive ability studies, one should not forget that even with "run-of-the-mill data" produced by large international organizations with hundreds to thousands of employees, unusual results occur: Results on wealth (GDP/c at purchasing power parity/ppp) are also odd and strange for single countries. e.g., according to the International Monetary Fund (IMF), Ireland is the richest country in the world, Guyana is richer than France, Japan poorer than Andorra etc. And there are large differences between different sources for the same variable, e.g., for Ireland, according to the IMF, \$145,196 per capita; according to the World Bank, \$126,905; and according to the CIA, \$102,500. The data are not exactly from the same year (2021–2023), but even then the results should be more similar.[12] The data on intelligence and knowledge per country are apparently not that bad. Only those who do not like it as a whole, especially for political reasons, will probably never be convinced.

Without collecting new data, neither through test studies nor through observational studies, one could already further improve the given data:

A. One could first average the results of all collections of cognitive ability studies (with or without standardization beforehand). This would probably lead to more accurate values overall.
B. A bit more sophisticated: For each country, one could estimate ability levels based on other variables (e.g., using best prediction formula 1) and also estimate results using ethnically and culturally similar countries, often neighboring countries such as Belgium by means of France and Holland (and somewhat Germany), but also the United Kingdom by means of Canada, Australia, and New Zealand (and somewhat Ireland and the Netherlands). The less good the measured data are, the more uncertain the values are (e.g., in Cuba, North Korea, or Afghanistan), the more given

---

measured values should be replaced by weighted estimates. Student achievement studies such as TIMSS and PISA take a similar approach to obtaining individual student data (i.e., estimating mathematics based on reading and science scores, in addition using biographic information such as sex, migration status, parents' education, scores of other students in the same grade and school).

In this way, the results already given could be used to arrive at even more meaningful country averages.

Author contributions I wrote the paper and made all analyses.

Data availability If permitted, data based on surveys and compilations by others and by us will be published.

Declarations

Competing interests The authors declare that there is no competing interests.

Appendix

See Table 8.

Table 8 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)	0.64	0.57	0.60	0.74	0.59
Education adults	1	0.45	0.55	0.58	0.28
GDP/c log		1	0.69	0.62	0.05
Politics positive			1	0.82	0.16
Religion weighted				1	0.43

$N = 78$ 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)

References

1. Angrist N, Djankov S, Goldberg PK, Patrinos HA. Measuring human capital using global learning data. Nature. 2021;592:403–8.
2. Barro RJ, Lee J-W. Barro-Lee data set. International data on educational attainment: updates and implications. Boston: Harvard University; 2000.
3. Beals KL, Smith CL, Dodd SM. Brain size, cranial morphology, climate, and time machines. Curr Anthropol. 1984;25:301–30.
4. Becker D. National intelligence levels based on Raven tests CPM, SPM, SPM+ and APM. 2024.
5. Becker SO, Woessmann L. Was Weber wrong? A human capital theory of Protestant economic history. Quart J Econ. 2009;124:531–96.
6. Bird KA. No support for the hereditarian hypothesis of the Black-White achievement gap using polygenic scores and tests for divergent selection. Am J Phys Anthropol. 2021;175(2):465–76.
7. Bold T, Filmer D, Martin G, Molina E, Stacy B, Rockmore C, Svensson J, Wane W. Enrollment without learning: teacher effort, knowledge, and skill in primary schools in Africa. J Econ Perspect. 2017;31:185–204.
8. Bolt J, van Zanden JL. 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. 2013.
9. Ceci SJ. How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. Dev Psychol. 1991;27(5):703–22.
10. Chaudhury N, Hammer JS, Kremer M, Muralidharan K, Rogers FH. Missing in action: teacher and health worker absence in developing countries. J Econ Perspect. 2006;20(1):91–116.
11. Chen Y, Feng S. Access to public schools and the education of migrant children in China. China Econ Rev. 2013;26:75–88.
12. Eysenck HJ, Schoenthaler SJ. Raising IQ level by vitamin and mineral supplementation. In: Sternberg RJ, Grigorenko EL, editors. Intelligence, heredity and environment. Cambridge: Cambridge University Press; 1997. p. 363–92.
13. Flores-Mendoza C, Ardila R, Rosas R, Lucio ME, Gallegos M, Reátegui Colareta N. Intelligence measurement and school performance in Latin America: a report of the study of Latin American intelligence project. Cham: Springer; 2018.
14. Francis G, Kirkegaard EOW. National intelligence and economic growth: a Bayesian update. Mank Q. 2022;63(1):9–78.
15. Gignac GE, Bates TC. Brain volume and intelligence. Intelligence. 2017;64:18–29.
16. Glewwe P, Kremer M. Schools, teachers, and education outcomes in developing countries. In: Hanushek EA, Welch F, editors. Handbook of the economics of education. Amsterdam: North-Holland; 2006. p. 945–1017.
17. Gordon RA. Everyday life as an intelligence test: effects of intelligence and intelligence context. Intelligence. 1997;24(1):203–320.
18. Gottfredson LS. Jobs and life. In: Nyborg H, editor. The scientific study of general intelligence. Tribute to Arthur R. Jensen. Oxford: Pergamon; 2003. p. 293–342.
19. Gust S, Hanushek E, Woessmann L. Global universal basic skills: current deficits and implications for world development (Nr. w30566). National Bureau of Economic Research. 2022.
20. Haidt J. Why universities must choose one telos: truth or social justice. Retrieved from Heterodox: The Blog website: https://heterodoxacademy.org/blog/one-telos-truth-or-social-justice-2. 2016.
21. Hanushek EA. The economic value of higher teacher quality. Econ Educ Rev. 2011;30:466–79.
22. Hanushek EA, Woessmann L. The knowledge capital of nations. Cambridge: MIT Press; 2015.
23. Harrison LE. Jews, confucians, and protestants: cultural capital and the end of multiculturalism. Lanham: Rowman & Littlefield; 2013.
24. Jahrbuch. Jahrbuch. [Yearbook.] München: DTV. 2004.
25. Jensen AR. The g factor. The science of mental ability. Westport: Praeger; 1998.
26. Jones G. Hive mind: how your nation's IQ matters so much more than your own. Stanford: Stanford University Press; 2016.
27. Kaufman SB, Reynolds MR, Liu X, Kaufman AS, McGrew KS. Are cognitive *g* and academic achievement *g* one and the same *g*? An exploration on the Woodcock-Johnson and Kaufman tests. Intelligence. 2012;40(2):123–38.
28. Krämer W, Leciejewski K. Statistic under socialism. Official data as a mix of reality and ideology and their reception by the media. AStA Wirtschafts Sozialstatistis Archiv. 2021;15(2):73–91.
29. Kremer M, Chaudhury N, Rogers FH, Muralidharan K, Hammer J. Teacher absence in India: a snapshot. J Eur Econ Assoc. 2005;3(2–3):658–67.
30. Kurian GT. The illustrated book of world rankings. Armonk: Sharpe; 2001.
31. Lee JJ, McGue M, Iacono WG, Michael AM, Chabris CF. The causal influence of brain size on human intelligence: evidence from within-family phenotypic associations and GWAS modeling. Intelligence. 2019;75:48–58.
32. Lim SS, Updike RL, Kaldjian AS, Barber RM, Cowling K, York H, et al. Measuring human capital: a systematic analysis of 195 countries and territories, 1990–2016. The Lancet. 2018;392:1217–34.
33. Luyten H, Merrell Ch, Tymms P. The contribution of schooling to learning gains of pupils in years 1 to 6. Sch Eff Sch Improv. 2017;28(3):374–405.
34. Lynn R. The role of nutrition in secular increases in intelligence. Personality Individ Differ. 1990;11:273–85.
35. Lynn R, Becker D. The intelligence of nations. London: Ulster Institute for Social Research; 2019.
36. Maddison A. 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. 2008.
37. Mankiw NG, Romer D, Weil DN. A contribution to the empirics of economic growth. Quart J Econ. 1992;107:407–37.

38. Mbiti IM. The need for accountability in education in developing countries. J Econ Perspect. 2016;30(3):109–32.
39. Meisenberg G, Lynn R. Ongoing trends of human intelligence. Intelligence. 2023;96: 101708.
40. Nisbett RE. Intelligence and how to get it. Why schools and cultures count. New York: Norton; 2009.
41. OECD. OECD skills outlook 2013. First results from the survey of adult skills (PIAAC). Paris: OECD; 2013.
42. OECD. PISA 2012 results: What students know and can do: student performance in mathematics, reading and science, vol. I. Paris: OECD; 2013.
43. OECD. Skills matter: further results from the survey of adult skills (PIAAC). Paris: OECD; 2016.
44. Oesterdiekhoff GW. Die Entwicklung der Menschheit von der Kindheitsphase zur Erwachsenenreife. [The development of humankind from childhood to adulthood.]. Berlin: Springer; 2013.
45. Ombredane A, Robaye F, Plumail H. Résultats d'une application répétée du matrix-couleur à une population de Noirs Congolais. Bull Centre d'Études Res Polytech. 1956;6:129–47.
46. Pritchett L, Viarengo M. Learning outcomes in developing countries: Four hard lessons from PISA-D. Research on Improving Systems of Education (RISE). 2021.
47. Pokropek A, Marks GN, Borgonovi F. How much do students' scores in PISA reflect general intelligence and how much do they reflect specific abilities? J Educ Psychol. 2022;114(5):1121–35.
48. Rindermann H. Sexual relationships between tourists and locals in Cuba: forms and experiences in the perspective of Cubans. Z Sexualforsc. 1999;12(2):159–77.
49. Rindermann H. Cognitive capitalism: human capital and the wellbeing of nations. Cambridge: Cambridge University Press; 2018.
50. Rindermann H, Baumeister AEE. Validating the interpretations of PISA and TIMSS tasks: a rating study. Int J Test. 2015;15:1–22.
51. Rindermann H, Becker D. FLynn-effect and economic growth: do national increases in intelligence lead to increases in GDP? Intelligence. 2018;69:87–93.
52. Rindermann H, Becker D. The future of intelligence: a prediction of the FLynn effect based on past student assessment studies until the year 2100. Personality Individ Differ. 2023;206: 112110.
53. Rindermann H, Carl N. Indigenous Ecuadorian children: parental education, parental wealth and children's cognitive ability level. Learn Individ Differ. 2017;54:202–9.
54. Rindermann H, Carl N. Human rights: why countries differ. Comp Sociol. 2018;17:29–69.
55. Rindermann H, Carl N. The good country index, cognitive ability and culture. Comp Sociol. 2020;19:39–68.
56. Rindermann H, Ceci SJ. 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. J Educ Gifted. 2018;41:298–326.
57. Rindermann H, Falkenhayn L, Baumeister AEE. Cognitive ability and epistemic rationality: a study in Nigeria and Germany. Intelligence. 2014;47:23–33.
58. Ritchie SJ, Tucker-Drob EM. How much does education improve intelligence? A meta-analysis. Psychol Sci. 2018;29:1358–69.
59. Rushton JPh, Rushton EW. Progressive changes in brain size and musculo-skeletal traits in seven hominoid populations. Hum Evol. 2004;19(3):173–96.
60. Rutkowski L. Review of: "Facing the Facts About Test Score Gaps" from Gregory Connor. Qeios. 2023. https://doi.org/10.32388/PF3D9Y.
61. Sandefur J. Internationally comparable mathematics scores for fourteen African countries. Econ Educ Rev. 2018;62:267–86.
62. Sear R. '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. 2022.
63. Stelzl I, Merz F, Remer H, Ehlers Th. The effect of schooling on the development of fluid and cristallized intelligence: a quasi-experimental study. Intelligence. 1995;21:279–96.
64. Thompson R, Smith RB, Bou Karim Y, Shen C, Drummond K, Teng C, Toledano MB. Noise pollution and human cognition: an updated systematic review and meta-analysis of recent evidence. Environ Int. 2022;158: 106905.
65. UNDP (United Nations Development Programme). Human development report 2010. New York: United Nations; 2010.
66. UNDP (United Nations Development Programme). Human development report 2021/22. New York: United Nations; 2022.
67. Warne RT. National mean IQ estimates: validity, data quality, and recommendations. Evol Psychol Sci. 2023;9:197–223.
68. Will M, Krapp M, Stock JT, Manica A. Different environmental variables predict body and brain size evolution in Homo. Nat Commun. 2021;12(1):4116.
69. Wolfram T. (Not just) Intelligence stratifies the occupational hierarchy: ranking 360 professions by IQ and non-cognitive traits. Intelligence. 2023;98: 101755.