Knowledge, Skills, and the Global Balance of Power:

What International Standardized Achievement Tests Tell Us about National Economic Potential and Prospects

Nicholas Eberstadt[1]
American Enterprise Institute

AEI Foreign & Defense Policy Working Paper 2025-10
May 2025

[1] The author holds the Henry Wendt Chair in Political Economy at the American Enterprise Institute (AEI). He would like to thank a number of people for their assistance with the research behind this paper: most importantly, AEI Research Assistants Evan Abramsky, Sarah Nakasone, and Peter Van Ness and AEI Interns Yuze Sui, Gayeong Song, and Hien M. Nguyen. The author alone is responsible for any remaining errors.

The research for this working paper was originally sponsored by the Office of Net Assessment in the Office of the Secretary of Defense (ONA/OSD). We gratefully acknowledge that support and thank ONA/OSD for permission to publish these results.

Abstract

This study is part of a project that seeks to take strategic demography "beyond the headcount approach," using the new information the data explosion has revealed about the human potential in national populations, with a particular focus on the strategic balance between the United States and China. In this paper we examine "knowledge capital," the economically productive knowledge and skills of national populations: how such potential differs between nations; how it affects levels of national productivity; and the determinants of that potential internationally. We explore these questions through statistical analysis of global data on academic achievement (which proxies knowledge capital), drawing also on other authoritative datasets on worldwide social and economic conditions.

Online Appendix: Accompanying Tables and Figures

The Slides may be accessed online here.

Introduction

In 2019 the world's leading authority on the testing and evaluation of educational performance all around the world, the Program on International Student Achievement (PISA)[2], announced stunning results from its latest round of global examinations of 15 year olds: China was number one, all across the board. So striking to these educators was China's recorded academic aptitude that PISA began its entire 5-volume report with the following paragraph:

> *…our PISA 2018 assessment shows that 15-year-old students in the four provinces/municipalities of China that participated in the study – Beijing, Shanghai, Jiangsu and Zhejiang – outperformed by a large margin their peers from all of the other 78 participating education systems, in mathematics and science. Moreover, the 10% most disadvantaged students in these four jurisdictions also showed better reading skills than those of the average student in OECD countries, as well as skills similar to the 10% most advantaged students in some of these countries. True, these four provinces/municipalities in eastern China are far from representing China as a whole, but the size of each of them compares to that of a typical OECD country, and their combined populations amount to over 180 million. What makes their achievement even more remarkable is that the level of income of these four Chinese regions is well below the OECD average. The quality of their schools today will feed into the strength of their economies tomorrow.*[3]
> (Emphasis added)

In math and science, China's test-takers outscored the average student from Western countries (i.e., members of the OECD—the Organization for Economic Cooperation and Development[4]) by over 100 points—the equivalent, in effect, of well over two additional grades of schooling. (In reading, the Chinese student's edge over their Western counterparts was "only" about 70 points, or a little less than two years of class-time equivalent.)

---

[2] "What is PISA," The Organisation for Economic Co-operation and Development, https://www.oecd.org/pisa/.

[3] OECD, "Preface", in *PISA 2018 Results (Volume II): Where All Students Can Succeed*, 2020, p. 3, https://www.oecd-ilibrary.org/docserver/b5fd1b8f-en.pdf?expires=1642118019&id=id&accname=guest&checksum=09B08AFB55A336C068668F43AE0AF6EE.

[4] OECD, "The OECD," 2008, https://www.oecd.org/newsroom/34011915.pdf.

These arresting figures highlight an issue arguably overdue for national security planners' consideration: namely, the role of a population's knowledge and skills (or "knowledge capital"[5]) in national economic potential, and thus ultimately national defense potential. All other things being equal, student performance today will bear on national economic performance—and thus the global balance of power—tomorrow.

This paper presents a summary of a more exhaustive examination of the evidence on this relationship. The study was undertaken as the latest phase of a continuing project on demography and the global balance of power, in which we have been harnessing the worldwide explosion of data on the characteristics of individuals who make up national populations to take strategic demography "beyond the headcount approach".

Since many millions of students from all around the world have been commonly tested over the past generation through achievement exams carefully developed, standardized and administered by trusted educational authorities[6], we have enough information at hand for statistical examinations of 1) the relationship between measured student achievement and national economic performance, and 2) the factors that "predict"[7] higher and lower levels of tested student achievement around the world.

To be clear: we fully recognize that the test scores we analyze in this report are, necessarily, imperfect proxies for the overall knowledge capital of any person—or any national population. Nevertheless, the information these tests convey is meaningful—and indeed telling.

Despite the "noise" around them, achievement test scores for knowledge and skills generate a powerful "signal". We find that they are good predictors of a country's economic productivity today—and no less important, good predictors of its per capita productivity ten years hence. High performing populations in these achievement tests also have more productive economies, even after controlling for

---

[5] Eric Hanushek and Ludger Woessmann, *The Knowledge Capital of Nations* (Cambridge: MIT Press, 2015), https://mitpress.mit.edu/books/knowledge-capital-nations.

[6] A previous study for the Office of Net Assessment by this author provides copious details on these authorities, their tests, and the data sources that compile their results. Cf. Nicholas Eberstadt, "Demographics and the Global Balance of Power: Tenth Quarterly Report for ONA/OSD Project, Submitted to the Director, Office of Net Assessment, Office of the Secretary of Defense", (unpublished paper, IHS Global Inc., July 10. 2020).

[7] Statisticians will always caution that the correlations they discover are simply associations, without predictive power. In this paper we use the term "prediction" loosely—meaning that knowing A will tell you B. That association lacks causation but conveys information.

other factors. Conversely, low performing populations on such internationally standardized knowledge and skills tend to be appreciably less productive economically, even after taking other factors into account.

National student achievement profiles vary vastly from country to country—even for tests of pupils of the same age. These differences underscore a crucial (if obvious) truth—quality of education is not the same from one nation (or even neighborhood) to the next. (As the disarming title of a key study on this topic puts it: "Schooling Ain't Learning."[8])

Generally speaking, and unsurprisingly, a country's academic achievement profile tends to track with its level of socioeconomic development. But there are also countries that seem to punch significantly above (or below) their developmental weight. Low-income countries that consistently report test student achievement results like those of high-income countries are probably not destined to be low-income for long. Vietnam is a poster child for this proposition—its students currently score on par with their counterparts in Norway.[9] Other countries—including perhaps China—may likewise be powering their way out of the realms of the lower and middle-income economies on the knowledge capital of their rising generations.

Our analysis indicates that "outliers" in academic achievement—places punching above or below their socio-economic weight—tend to cluster by geographic regions. That is to say—knowing what region of the world a country is in helps us predict its knowledge capital, as proxied by test scores, even after we have taken its developmental level into account.

We know, for example, that countries from the Latin American and Caribbean region seriously underperform in student achievement against socioeconomic peers in other parts of the world. Relatively poor knowledge capital test scores probably do not augur well for future economic trajectories for many countries from this region.

Conversely, we know that countries and territories from the East Asian realm—including some populations technically in Southeast Asia—tend to perform in

---

[8] Lant Pritchett, *The Rebirth of Education: Schooling Ain't Learning* (Washington: Center for Global Development, 2013), https://www.cgdev.org/publication/9781933286778-rebirth-education-schooling-aint-learning.

[9] Harry Patrinos and Noam Angrist, "Harmonized learning outcomes: transforming learning assessment data into national education policy reforms," World Bank Blogs, August 12, 2019, https://blogs.worldbank.org/opendata/harmonized-learning-outcomes-transforming-learning-assessment-data-national-education.

academic achievement tests well above the levels that would be predicted simply on the basis of their income levels and other developmental characteristics. All other things being equal, East Asian test scores tend to be a full grade level (or more) higher than would be expected for counterpart populations elsewhere at the same level of development. Just why East Asian populations systematically perform so well—and why some other regions underperform so consistently—is a question of immense import. It has given rise to a number of proposed theoretical explanations—ranging from the "cultural" to the "biological"—but thus far no firm and informed consensus has emerged as to its answer.

This brings us to the question of student achievement in the PRC today. The scores "validated" by PISA are from a select but nationally unrepresentative handful of provinces: in the main, China's richest areas. Even so: there are reasons for skepticism about the utterly stellar figures reported, as we shall see.

China may well be punching above its developmental weight in student achievement, like most of the rest of East Asia. If so, this could presage major advances in Chinese productivity in the years immediately ahead—and with those advances, a corresponding tilt in the global economic balance.

Evidence at hand, however, does not yet make such a case. Over 85 percent of China's national population—and even more of its student population—reside in regions for which Beijing does not disclose achievement scores. There is reason to think scores in these regions—and thus for China overall—could be far lower than those reported for "China B-S-J-Z" (the acronym for the amalgam of Mainland regions that participated in the most recent round of international PISA exams).

And of course it matters greatly just how much lower China's true nationwide knowledge capital indicators might actually be.

Drawing upon global relationships we estimate statistically—given the knowledge at hand—our analysis offers a central estimate for China's knowledge and skills that would place student achievement in that great country on par with places like Turkey and Mauritius—other developing societies the World Bank likewise classifies as "upper middle income economies".

That said: given currently available data, and some of the patterns we derive from them, we cannot conclusively dismiss out of hand the proposition that China's true level of nationwide student achievement could be considerably higher than this—even, potentially, reaching levels of some high-income OECD countries today.

There could be dramatically different outlooks for the next phases of the Sino-US economic competition depending on just where China's current and future knowledge and skills profile falls within that range. Attempting to determine with greater precision China's true knowledge capital contours would therefore look to be a logical, perhaps even pressing, research priority—and pursuing that question would be a natural next step in this continuing project.

Background

In our research on "demography and the global balance of power", we have been examining the influence on international security and changing global economic and defense potential of what we call "demography beyond the headcount approach". In a world where the per capita GDP between countries now can differ by a factor of 100 to 1 (or even more)[10], and in which productivity differential between countries can change surprisingly quickly, understanding the changing "demography of national economic potential" is ever more central to an informed assessment of overall national security. That demography of national economic potential takes us into the details of social, economic and other characteristics of the individuals who comprise national populations—for "human resources" are the main ingredient driving modern economic development and growth, with the role of "natural resources" in overall economic output steadily shrinking as "service sectors" and "knowledge-intense sectors" rise.

In previous research we harnessed the "data revolution"—the explosion of new information on the human resource characteristics of the individuals who comprise national populations—to demonstrate statistically the tremendous influence on both current and future national productivity levels of such human factors as health, education, and residence (urbanization): all the more so when an auspicious "business climate" facilitates "unlocking the value" of those human resources. All other things being equal, faster growth in "human resources" or "human capital" means faster growth of national economic potential—and thus defense potential.

We also demonstrated that the global balance in highly educated working age manpower is dramatically shifting: whereas the USA was the uncontested

[10] For example: for the year 2018 Singapore's per capita GDP was estimated at over $68,000, while the Central African Republic's was placed at just $623 (both measured in constant international 2011 USD); Maddison Project Database 2020, University of Groningen, https://www.rug.nl/ggdc/historicaldevelopment/maddison/releases/maddison-project-database-2020

"education superpower" at the end of the Cold War, with at least three times as many working age men and women with college degrees as the closest international competitor, it now appears that China will soon outstrip America in the sheer number of college graduates in its working age manpower pool—if indeed it has not done so already.

Education is now a strategic factor in long-term international competition, and an increasingly important one. But how to measure its impact on national economic potential? The simplest approach—the one we used in our initial research—was to proxy "education" by the sheer years of schooling a national population had experienced—for years of schooling are in principle easy to measure and there are large databases mapping out changes in such educational attainment for the world's populations over the postwar era.[11] But we know that years of education are just a crude measure of education *per se*—possibly a highly imperfect one. This is so because the *quality* of education varies so widely, both across countries and even within them. Thus while it is informative to know that China's total numbers of working age college graduates are poised to outstrip totals for the USA, a far more meaningful comparison would concern the knowledge and skills of the two respective cadres—the actual "education" that all those years of schooling were intended to impart.

The current phase of our research therefore takes us into a detailed statistical examination of the influence on national economic performance of what Stanford economist Eric Hanushek has called "the knowledge capital of nations"[12]: beginning with the analysis of that relationship as revealed by international standardized achievement tests for knowledge and skills.

What Are International Standardized Achievement Tests?

In this paper we will be analyzing results of "international standardized achievement tests", which educators use to measure learning and skills for both individuals and entire countries. Academic achievement, of course, is important in and of itself. But we can show a relationship between tested achievement and

[11] Most importantly: the Barro-Lee Educational Attainment Dataset (http://www.barrolee.com/), and the Wittgenstein Centre for Demography and Human Capital's Human Capital Explorer (http://dataexplorer.wittgensteincentre.org/wcde-v2/).

[12] Hanushek and Woessmann, *The Knowledge Capital of Nations*, https://mitpress.mit.edu/books/knowledge-capital-nations.

economic performance at the national level. We can also describe, with some statistical precision, the factors that seem to make for higher and lower national academic achievement around the globe.

Today's international standardized achievement tests trace their origins back to 19th Century "psychometric" research in Europe and North America,[13] when multiple choice exams for IQ and academic achievement were first developed. (Achievement tests are intended to assess knowledge and skills, not IQ—IQ tests examine cognitive ability.) They draw indirectly from the development of standardized college entrance examinations in the USA in the early 20th Century, and directly from early postwar efforts to produce internationally valid achievement tests offered in a multiplicity of languages.

Today's globally administered tests of knowledge and skills are overseen by two agencies, the International Education Association (IEA) and the aforementioned OECD's PISA. Governments the world over nowadays authorize IEA and PISA tests in their schools, and rely on these for evaluation of their own national educational performance. The datasets of test results that we analyze in this study, in other words, are regarded as the "gold standard" by educators and policymakers internationally—not just in Western countries, but in low-income countries on all continents, too.

By now, large numbers of students have been evaluated by IEA and PISA. PISA's database includes over two and a half million test results for reading, math and science for 15 year olds from representative sample surveys in over 80 countries and many more sub-regions since the first PISA wave in the year 2000. IEA's reading (PIRLS) and math-science (TIMSS) tests for 4th and 8th graders contain results for about a million and about two and a half million students from 1995 to the present, respectively. In all, these datasets on academic achievement cover almost a hundred countries with a total population of nearly three billion—over three billion if we include the regions of China tested.

Our project takes a deep dive into these IEA and PISA student achievement datasets.

[13] Adrian Wooldridge, *Measuring the Mind: Education and Psychology in England c. 1860-c.1990.* (New York: Cambridge University Press, 1995).

We also examine a dataset on student achievement prepared by the World Bank—an effort known as the Harmonized Learning Outcomes (HLO) database.[14] The HLO initiative uses some creative "shortcut" methods to agglomerate the PISA and IEA datasets by adjusting overall averages for tested countries so that these match—and then further augments country coverage by linking in regional authorities' test results from Africa and elsewhere by finding "bridge countries" used in both samples and similarly "scaling" them in.

The HLO approach is not without its critics, and the HLO dataset can only offer summary information on national results. But the HLO work has also been peer reviewed by *Nature*, where it appeared in 2021[15], and vetted internally within the World Bank by its professional staff. It has the allure of extending global coverage to over 160 countries and territories, accounting for up to 98 percent of the world's estimated population (depending on how we count China).

All these achievement tests, we should note, are prepared for students 15 years of age and under. OECD is also rolling out a knowledge and skills test for adults 16-65 (known by the acronym PIAAC[16]) but it has to date evaluated less than a tenth as many exams as PISA and is limited in coverage mainly to affluent Western democracies. We did not find this dataset suitable for analysis at this point, given our objective of detecting global trends and patterns in the relationship between achievement and national economic performance.

Achievement Test Caveats

Statistical analysis is best undertaken with an appreciation of the strengths and weaknesses of the data under consideration. This is the "what we know and how we know it" question analysts should always bear in mind in policy research. Before we dive into our findings, a few words are in order about the quantitative information so central to our report.

[14] Harmonized Learning Outcomes (HLO) Database, The World Bank, https://datacatalog.worldbank.org/search/dataset/0038001

[15] For the methodology, see Noam Angrist, Simeon Djankov, Pinelopi Goldberg and Harry Patrinos, "Measuring Human Capital Using Global Learning Data," *Nature* 592**,** 403–408, 2021, https://www.nature.com/articles/s41586-021-03323-7

[16] The Program for the International Assessment of Adult Competencies, The Organisation for Economic Co-operation and Development, https://www.oecd.org/skills/piaac/

The achievement tests in this report are educational performance instruments, carefully devised and also carefully administered (at least in principle). But as proxies for the totality of economically relevant knowledge and skills, they nevertheless have limits, for individuals and countries alike. This should hardly surprise: how could it be otherwise for an exam consisting of couple of dozen multiple choice questions administered in a single seating of just a couple of hours?

There are some voices, in academia and elsewhere, who reject achievement tests data altogether, dismissing them as meaningless, or denouncing them as racist "class-ist" tools of hegemony. But such ideologized claims are both intellectually impoverished and counter-scientific. They fail to recognize the formidable predictive power, even after controlling for other factors, that such tests possess for student outcomes at both the micro-level (individuals) and the macro-level (societies). They also typically fail to describe specific methodological shortcomings for these tests, as we do below.

To begin: there is the very real possibility that the testing of just reading, math and science skills (as achievement tests only do to this writing) may overlook other skills essential to personal and national success: for example entrepreneurship; ability to cooperate; leadership; "grit" and motivation, etc. etc.

Additional problems attend all international testing: such as the dilemma of standardization (do the same questions work equally well for German and Arabic?). There is "excluded population bias" (skewing reported results by missing or under-representing certain groups). And of course in all randomized sample surveys there is the matter of "replicability of results"—a trouble compounded in achievement test-taking by the familiar "off day" phenomenon on performance tests.

And there are further questions and problems with the IEA and PISA datasets, at least for our purposes. One of these is highly technical—the adjustment of test score results in accordance with "item response theory."[17] Two others are more straightforward:

a) Affluent OECD countries are seriously over-represented in IEA and PISA testing so far—most of the developing world and nearly all of the

[17] "Item response theory" or IRT involves standardizing test score results with additional information beyond the number of correct and incorrect answers submitted by the examinee—meaning that test takers with exactly the same multiple choice selections on an international achievement exam could be assigned different scores.

lowest income countries do not yet take part in these programs—and in consequence, genuinely global relationships between academic achievement and national economic performance are harder to estimate accurately.

b) From the standpoint of our statistical modeling, there is a mismatch between the age of the tested pupil population (4^{th} grade through age 15) and the age of the workforce for contemporary economies (almost entirely over 15 years of age). Simply put, pupils do not operate the economies –and the people who do are not tested in our datasets. A great many people across the globe stay in school past age 15—so their eventual knowledge and skills will be systematically underestimated by the tests we use. But school-age knowledge and skills can also *deteriorate*, through lack of use, and apparently these do so in some low-income settings (more on which later).

Readers with an interest can peruse more on some of these limitations to the achievement datasets in our Appendices.[18]

At the end of the day, reasonable observers should understand that test scores can only serve as an imperfect proxy of "knowledge capital". Yet by the same token, no reasonable observer ought to reject outright statistical assessments based on these test scores simply because of their identifiable imperfections.

After all: policymakers today operate in a world where they are constantly bombarded by imperfect (or positively flawed) statistical data, from public opinion polls to the latest quarterly GDP estimates to national poverty rates, and they make decisions on the basis of inherently incomplete information. The findings from achievement test datasets surely get a hearing by this same standard.

Moreover: the very limitations we have just detailed mean that the informational "signal" from achievement test scores is being muffled by statistical "noise". Consequently, the true relationship between knowledge capital and national economic performance could be *stronger* than the correspondence our statistical analysis detects. And as we shall see, the relationship between achievement scores and national productivity is highly meaningful, even after controlling for other major determinants of national economic potential.

[18] See for example Appendix A pages 4-12.

What Standardized Tests Reveal About Differences In International Achievement

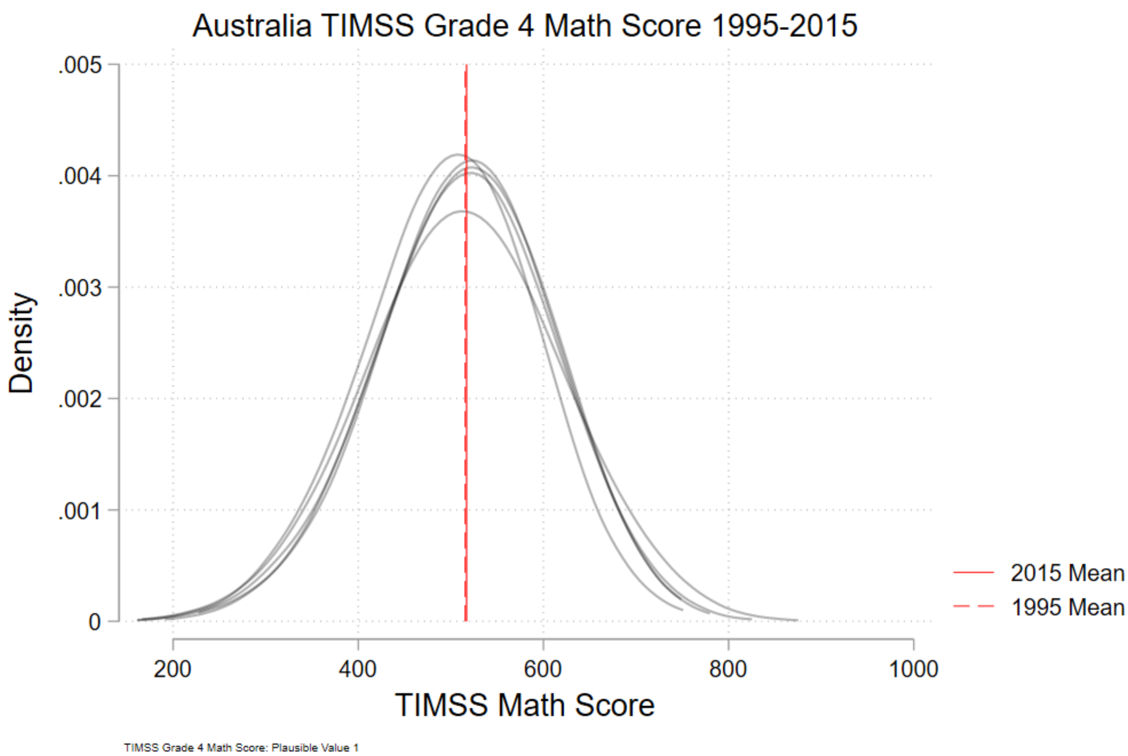
In this section we summarize our findings from a painstaking review of the OECD and IEA datasets on student achievement (PISA, PIRLS, TIMSS). Those interested in a much deeper dive into these materials can peruse Appendix B in this report.[19]

All achievement tests in this report are normed against a notional global mean of 500, set at the test's debut (scores can range from 0 to 1000). Differences in student performance are also scaled in accordance with a presumed "bell curve" of results (to statisticians, a "normal distribution"), and a 100 point differential in scores corresponds to a full "standard deviation" along that curve. In practice, a 100 point difference in scores can be taken to correspond with the impact of about a two or two and a half year differential in schooling levels.

For a sense of what an overall national profile of academic achievement performance tends to look like, we randomly select data from Australia. (See Figure 1) In this example we see that recurrent testing over a 20 year period more or less replicates original patterns of national achievement—though in all three datasets we see a drift toward higher scores over time, a tendency that could be explained by gradual improvements in educational quality. Most students score near their national mean—the national mean is therefore a main indicator of international differences in tested knowledge and skills. But outliers matter too—and from the perspective of contributions to the national economy, high performers quite possibly could be of special interest as well. Arguably of interest could be both the share of students above some given international threshold (such as say a score of 600) and the mean scores of (say) the top 10 percent of national test takers.

[19] Appendix B contains several hundred pages of tables and graphics—but this is only a portion of our statistical examination of the patterns and trends *within* these datasets.

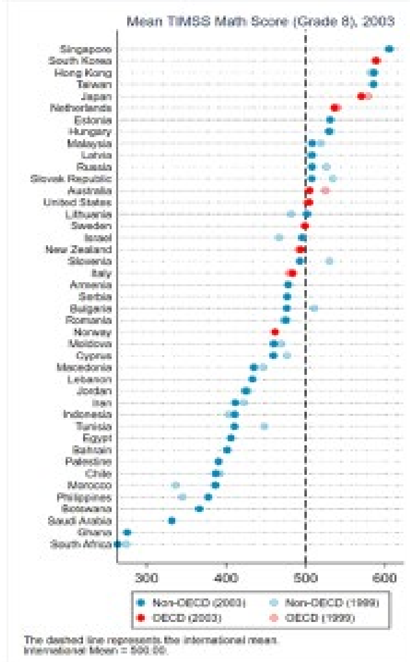
Figure 1.



The PIRLS, PISA and TIMSS datasets underscore one fact above all others: there is an enormous range in academic achievement—in tested knowledge for reading, math, and science—among countries in the contemporary world.

Figure 2 below illustrates a typical international range in mean test score performance for a typical year, this coming from TIMSS math exams for 8th graders.

Figure 2:

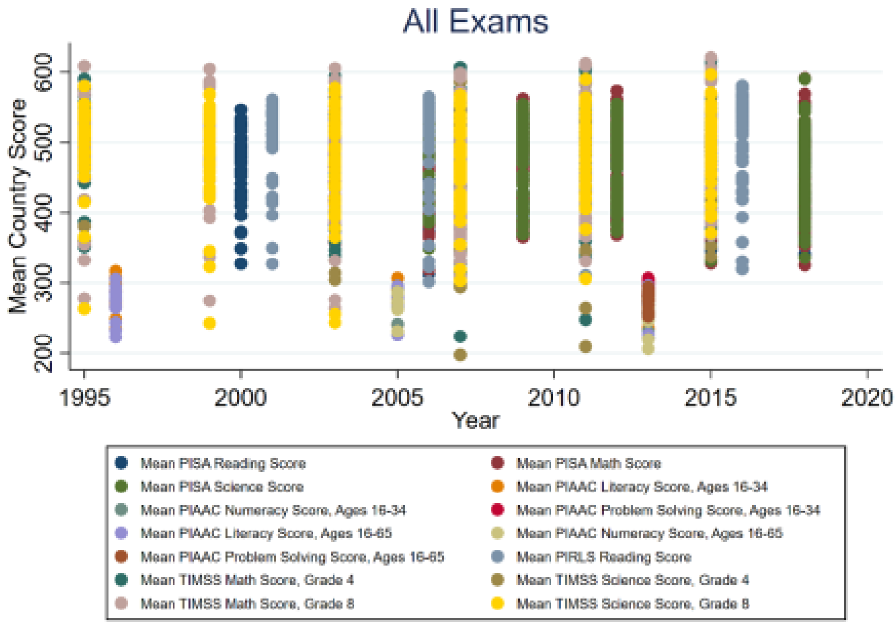


25

Ponder those international differences for a moment. In effect, a 300-point gap in mean test scores would track with a notional 6-plus years of schooling content in a contemporary Western school system. (A 400-point gap would track with a gap of eight years or more of such schooling.)

The stunning variation at the national level can be summarily seen in the range of *average* scores reported, as shown in Figure 3 below. In all of the annual waves in these achievement tests, the highest mean score is nearly 300 points above the lowest national performer—and in some the gap is on the order of 400 points. Such disparities point to vast divergences in tested capabilities for national student populations—and they may speak to corresponding disproportions in productive capacities in the economic realm.

Figure 3:



[26]

A closer look at achievement test profiles puts some of these differences in greater relief. A comparison of Lebanon and Japan is instructive for example, for a sense of how a place the World Bank classifies as an “upper middle income economy” looks next to a “high-income economy” and OECD member. (See Figure 4). Mean scores in the two countries are separated by over 130 points—the equivalent of perhaps close to three years of class time, and these are 10th graders being examined. Perhaps no less noteworthy is that the fraction of “high performers” scoring 600 or better on this test is just over 2% in Lebanon, but over 20% in Japan.

Figure 4.

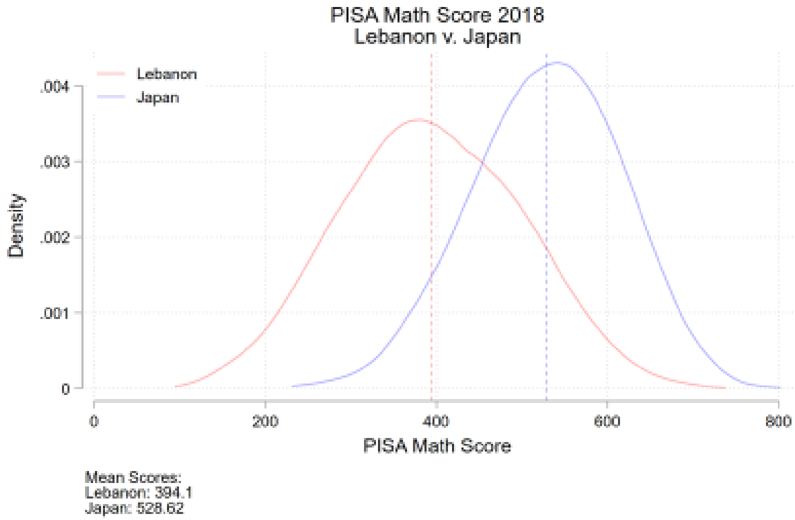


Table 4:

Jurisdiction	% Above 600
Lebanon	2.4%
Japan	21.3%

But highly significant differences in tested achievement separate some of the OECD countries. Consider newcomer Colombia in relation to longstanding OECD member Finland (long a poster child of high educational quality) in Figure 5. Here too a gap in mean test scores of about 120 points; and whereas fewer than 3% of Colombia's test-takers reach the 600-point threshold, over 30% of Finland's manage to do so.

Figure 5.

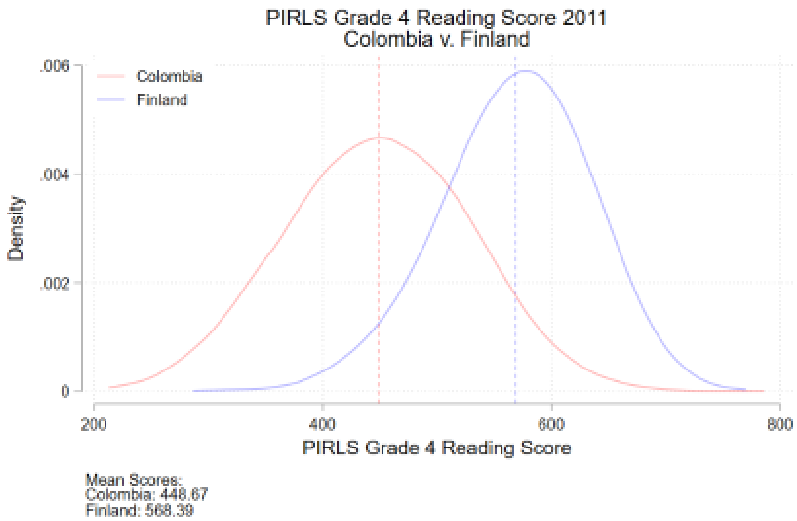
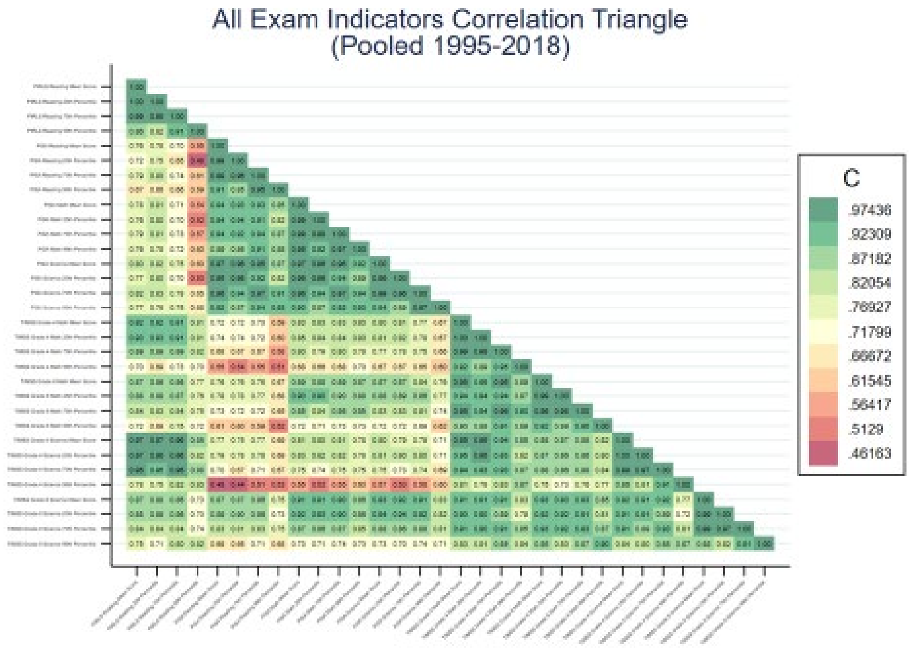


Table 5:

Jurisdiction	% Above 600
Colombia	2.4%
Finland	31.8%

The academic achievement datasets we examine in this report do not generate identical results from one test to the next—but on the whole their results for tested countries correlate closely. We can see as much from the correlation triangle below. We have prepared this as a "heat map" so that readers can tell which pairwise comparisons have higher, and lower, correlations (for statisticians, "simple *r*"). Although the shapes of distributions for test performance differ by test, year, and country, we can also see that high- and low-performers by country also track across datasets. All in all: a country that tests well (or poorly) in PIRLS will also test well (or poorly) in PISA and TIMSS. Furthermore: countries with higher mean national scores also have more high performers and fewer poor performers—while the reverse is true for countries lower with mean national scores. The consistency of these distributional relationships ("bell curve" shapes) mean that we can get almost as much information about high- and low- performers for a country's mean score as from examining those other indicators separately. (See figure 6)

Figure 6.

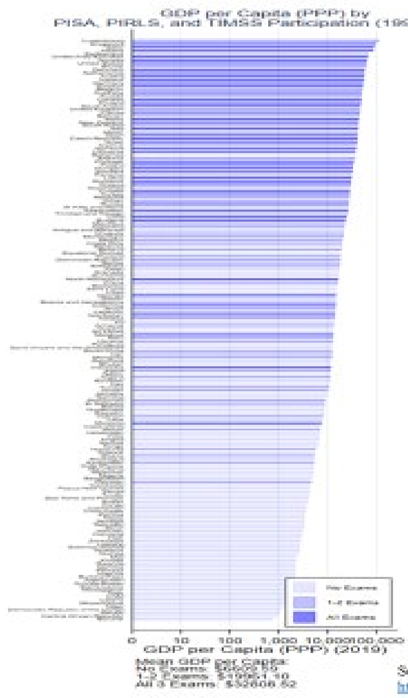


Although we looked at the three achievement datasets' results separately in our own statistical analysis, a creative initiative carried out under the aegis of the World Bank pools these sets together—and also adds in regional test results from countries outside the IEA and OECD testing purview, but with testing regimens that purportedly follow the same methodology. The Harmonized Learning Outcome (HLO) database proposes to aggregate all these national scores by establishing, so to speak, "exchange rates" between the various datasets: calibrating up or down the reported overall mean scores from one dataset to concord with the reported scores of the other.

As noted earlier: the HLO approach to pooling international test score results from across datasets is controversial in some quarters—some leading authorities criticize the approach. On the other hand, that approach has passed muster at one of the leading peer review science journals, and through internal vetting at the World Bank as well. And the promise of such an augmented dataset lies in the fact that it covers almost the whole world, affording near global coverage for the statistical analysis in this report. (See Figures 7a and 7b)

As may be seen in Figures 7a and 7b, there is a huge income gap between the countries currently omitted from any PIRLS/PISA/TIMSS dataset and those included in at least one of them: per capita income in the former is on average barely a third of the latter, and scarcely a fifth of countries included in all three. And without belaboring the point: those omitted countries account for over three and a half billion of the world's population (not counting China, which is only partly included in the PISA database.). Extending the country coverage through regional testing efforts "bridged" to these three main global datasets, as HLO does, makes it possible at least in principle to obtain a truly global sample of countries for our global statistical analysis.

Figure 7a.

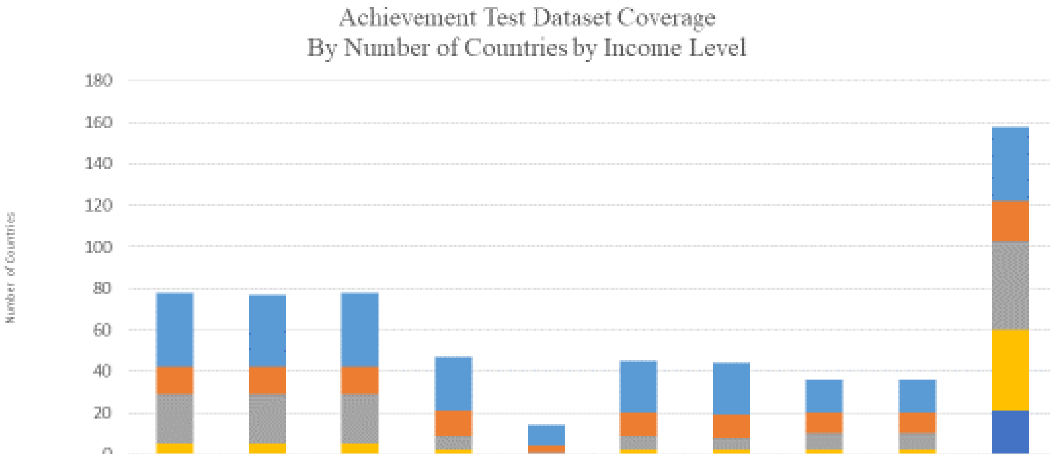


Economic Performance Range in the PIRLS/PISA/TIMSS Sample:

Mean GDP per Capita
No Exams: $6609.59
1-2 Exams: $19961.10
All 3 Exams: $32808.52

Source: World Bank, "GDP per Capita, PPP (Constant 2017 International $)," accessed July 16, 2021, https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD.

Figure 7b.



	PISA Math 2018	PISA Reading 2018	PISA Science 2018	PIRLS 2016	ePIRLS 2016	TIMMS Grade 4 Math 2015	TIMMS Grade 4 Science 2015	TIMMS Grade 8 Math 2015	TIMMS Grade 8 Science 2015	HLO
OECD	36	35	36	26	10	25	25	16	16	36
High-Income	13	13	13	12	3	11	11	10	10	19
Upper-Middle Income	24	24	24	7	1	7	6	8	8	43
Lower-Middle Income	5	5	5	2	0	2	2	2	2	39
Low-Income	0	0	0	0	0	0	0	0	0	21

The HLO approach can only "norm" mean scores—but since we have already seen that the relationship between mean scores and high performers is fairly steady, the loss here for our analysis is not major. In the following sections of the report we use HLO data for our evaluations of the global relationship between tested achievement and economic performance, along with results from the IEA and OECD datasets.

Knowledge Capital, Achievement Tests, And National Economic Performance

In this section we summarize the results of our statistical modeling of the relationship between tested academic achievement and national economic performance. This part of our study was extensive and detailed; we only present a few highlights and conclusions here. (Those interested in reviewing some of the extensive results of our modeling can peruse Appendix C at the end of this report.[20])

In brief: we found international standardized test scores to be a powerful predictor of national productivity levels, both today and ten years into the future. Tested

[20] Appendix C is voluminous but contains only a fraction of the multivariate regression models we tested in preparation for this section of the report.

skills and knowledge make an *independent* contribution to national economic performance: that is to say, the relationship remains statistically meaningful after we control for other major determinants of economic potential, including health, urbanization, and "business climate" (the quality of a country's institutions and policies).

The statistically revealed global relationships between test scores and national economic performance in our modeling were always strongest in the HLO dataset—as one might expect, given the much greater completeness of global coverage in HLO than the IEA and OECD datasets.

In our results test scores were *not* superior predictors of national economic performance vis-a-vis mean years of national schooling. This noteworthy finding, we suspect, may speak more directly to the limitations of test scores as a measure of national knowledge capital than to the true underlying relationship between knowledge capital and national economic performance.

That said: we found that a country's years of education *plus* achievement scores were impressive predictors of a country's per capita GDP. Taken together, years of schooling *in combination with* the quality of education as proxied by achievement tests performed very well in our models for estimating national economic performance. This modeled result comports with the commonsensical proposition that national economic potential is strongly affected by *both* the amount of schooling *and* the quality of schooling its population receives.

Before reporting more on our results, a few words are in order here about the general method in the statistical analysis that derived them. In this and subsequent sections of the report, we use a simple econometric technique known as "OLS multivariate analysis" (also called "regression models") to examine the relationship between test scores and other variables. In this section, we model the influence of national test score results and a number of other national factors ("independent variables") on per capita GDP (our "dependent variable" here).[21] In addition to information from the various international test score datasets already mentioned, we used well recognized international datasets to provide information on the following factors: health; educational attainment; urbanization levels; "business climate", and per capita GDP. We use the same datasets and indicators for these variables[22] as in previous phases of this ONA project; this continuity affords us

---

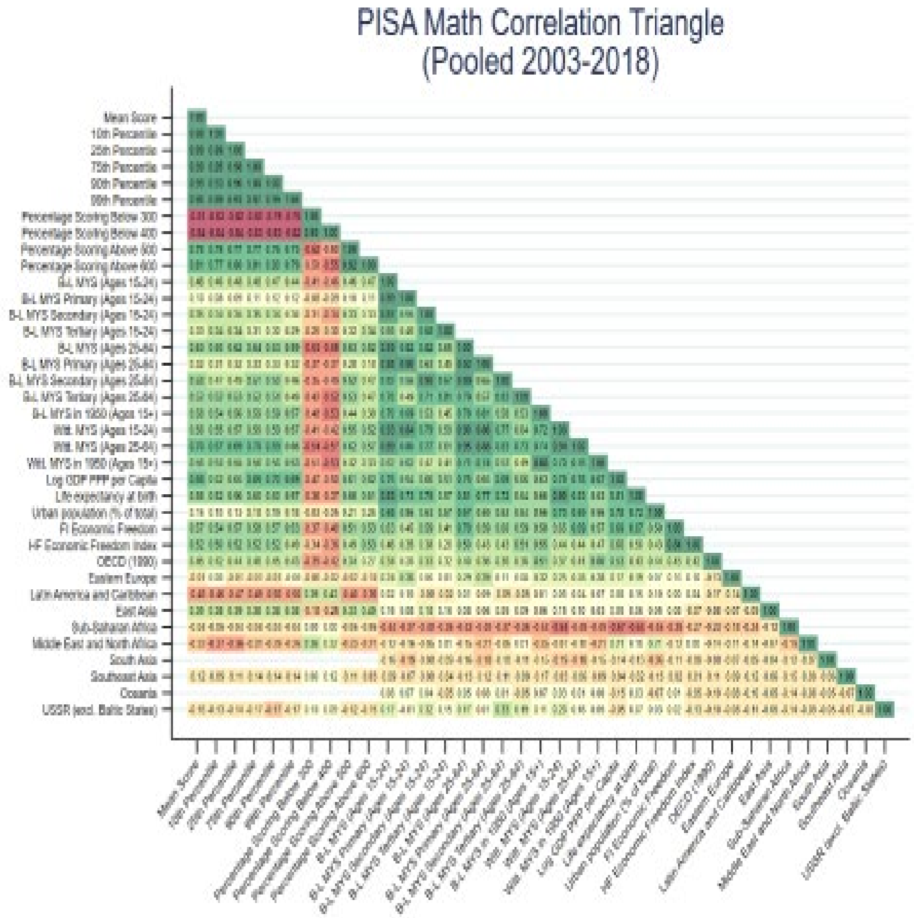
[21] Unless otherwise specified, the models we discuss are "random effects" models with pooled data.
[22] See Appendix A Page 13 for more detail.

with a measure of familiarity with sources and model outcomes that helps us put this new work in better perspective. (In some models we also added so-called "dummy variables" to indicate the region of the world in which the country in question was located, since there were indications this could matter too, at least for some of the questions in this report.)

The varied, myriad associations between the numerous indicators we worked with in the many hundreds of statistical models we devised would be tedious in the extreme to run through here. Instead we can give a sense of the whole with a "heat map" correlation triangle representing the associations between test score indicators and our socio-economic and regional indicators for PISA math scores. (See Figure 8)

Figure 8 contains over 660 pairwise correlations between indicators we examined in this report—and please note that this is only one of nine such "maps" we prepared from the information in the PIRLS, PISA, and TIMSS datasets. But the general relationships we show in the "map" below also come through in all the others.

Figure 8.



In Figure 8 (and the other “maps”) the relationship between tested student achievement and per capita GDP is strong, with a correlation coefficient close to 0.7—and not only with regard to mean scores but also for high performers. That said: tested achievement is only one of the “human factors” pertinent to national productivity with a strong correlation—life expectancy, mean years of schooling, urbanization, and “economic freedom” also generate strong correlation coefficients (i.e., 0.5 or higher) with per capita GDP. In fact, the correlation with per capita GDP is a bit higher for life expectancy and for urbanization than for test scores—while the correlations with per capita GDP for educational attainment and for test scores are roughly comparable.

No less important: the correlation between these other “human factors” and tested achievement scores are almost uniformly strong, as well. Such cross-correlations (what statisticians call “collinearity”) complicate the task of teasing out the independent contribution to national economic performance for knowledge capital as proxied by our academic achievement test scores.

Using PIRLS, PISA and TIMSS data, our models found a statistically meaningful association between tested knowledge and skills—this too when we control for health (life expectancy at birth), urbanization (percentage population in urban locations), and "business climate" (quality of institutions and policies, sometimes called indices of "economic freedom"). But we are mindful that these academic achievement indicators are *less* powerful predictors of national economic performance than the obvious alternative of educational attainment: i.e., sheer years of schooling for a national population.

In previous statistical examinations of the relationship between human resources and national economic performance in earlier phases of this project, we found a robust and remarkably stable correspondence between a country's mean years of schooling (MYS) and its per capita productivity. Using just four indicators—life expectancy; mean years of schooling; urbanization ratio; and "business climate"—we consistently generated models accounting for 80%-85% of international differences in GPD per capita across countries and over time: typically, these models showed each additional year of MYS tracked with an increase in per capital GDP of 9%-12% (after controlling for other factors).

Those enduring results for the MYS-GDP per capita relationship were replicated in our statistical models for this project. But test score indicators from all three datasets on student achievement provided notably less predictive power with respect to per capita GDP.

On the whole our "four factor" models lost about *15 to 20 percentage points* of predictive power (R-squared value) when we swapped in a test score indicator in the place of a MYS indicator. And just as with models that relied upon years of school, the predictive power (R-squared) of ten-year-lag models (models using say year 2000 data to predict results for 2010) was almost the same as zero-lag models (say, year 2000 data predicting year 2000 results).

Earlier we outlined a variety of reasons that test score indicators from the three main student achievement datasets might have limitations for predicting national economic performance—even if the true underlying relationship between knowledge capital and productivity actually were actually stronger and more determinative than between productivity and years of schooling (as we are inclined to suspect).

One of those reasons is the limited country coverage of the PIRLS, PISA and TIMSS datasets. Although the cumulative membership in these datasets continues to expand, they nevertheless omit countries and territories comprising over three fifths of the world's population—virtually all of them developing societies, with the lowest income regions particularly underrepresented. If the lower half of the global income spectrum is largely missing, then the modeling of global relationships between income and any other variable is probably going to be error-prone (statistically "biased") and somewhat misleading.

This is why we also decided to rely on the World Bank HLO dataset for modeling international test score/national productivity relationships. Notwithstanding technical criticisms raised about its agglomeration of achievement test results from unrelated international datasets—a potential "apples to oranges" problem—the allure of HLO's much greater global coverage was compelling for our purposes in this project.

We found the HLO database offered much more information about the global relationship between academic achievement and national economic potential. There was still a gap in predictive power when the HLO scores in question were for "secondary school students" (such as PISA tests)—but that gap was cut to roughly 10 percentage points. For "primary school student" scores (such as PIRLS and TIMSS 4th grade reading, math, science tests) the disparity in overall predictive power was all but eliminated—including for models predicting national economic performance ten years into the future.[23]

Our "four factor" models can account for up to 85% of the differences in GDP per capita across the world when we use HLO test scores instead of years of adult schooling; these models are very nearly as powerful in predicting per capita GDP ten years into the future as for current levels of per capita GDP.

We compare the performance of two selected "four factor models" below in predicting per capita GDP—the only difference between two is that one swaps out educational attainment as measured by mean years of schooling, and instead replaces that indicator with all HLO scores (reading, math and science; primary and secondary students). (See Tables 1a and 1b).

[23] Just why the HLO dataset's achievement scores for primary students should predict GDP/capita so much better than the corresponding scores for secondary school students is a question we cannot yet answer—but it a reality that we recognize, and flag here.

Table 1.

a) Four Factor Model With MYS (zero lag)

Barro-Lee Mean Years of School	
Barro-Lee Mean Years of School (25-64)	0.102***
	(18.88)
Life expectancy at birth, total (years)	0.0373***
	(20.05)
Urban population (% of total population)	0.0204***
	(29.36)
Economic Freedom of the World Index	0.118***
	(7.91)
Constant	3.881***
	(38.71)
No of Observations	1843
R-Squared	0.862
F Statistics	2879.3

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

b) Four Factor Model With All Means HLO Scores (zero lag)

HLO Mean Score	
HLO Mean Scores (Math, Reading, an~)	0.00373***
	(11.41)
Life expectancy at birth, total (years)	0.0378***
	(10.26)
Urban population (% of total population)	0.0204***
	(17.14)
Economic Freedom of the World Index	0.147***
	(5.18)
Constant	2.896***
	(15.11)
No of Observations	704
R-Squared	0.832
F Statistics	862.7

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

In these selected but illustrative models, a one-year increase in mean years of schooling for a country's working age (25-64) population tracks with an increase in its PPP-adjusted per capita GDP (in 2017 international dollars) of roughly 10 percent, while a 100 point increase in mean scores would correspond with a 46 percent increase in per capita GDP. Both of these can be considered powerful influences on a country's economic performance.

But as Figure 8 shows (among many other things), MYS and HLO primary test scores display a strong positive correlation: in other words, countries where schooling levels for working age people are high are likely to be places where test scores *are also high*. This means our educational attainment indicators may be "capturing" the contribution that should be attributed to student achievement—an indicator with which it is tightly associated in real life, but one that is functionally separate from it.

We can parse the interplay of educational quality and educational quantity in national economic performance by making a "five factor model" for GDP/per capita that brings both of them in (along with life expectancy, urbanization ratios, and an indicator of "business climate". We illustrate results for one such model below in Table 2.

Table 2.

HLO Mean Score and Barro-Lee Mean Years of School

HLO Mean Scores (Math, Reading, an~)	0.00232*** (6.01)
Barro-Lee Mean Years of School (25-64)	0.0674*** (6.45)
Life expectancy at birth, total (years)	0.0358*** (9.61)
Urban population (% of total population)	0.0188*** (15.52)
Economic Freedom of the World Index	0.120*** (4.14)
Constant	3.345*** (15.81)
No of Observations	658
R-Squared	0.840
F Statistics	683.4

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

To go by coefficients from Tables 1 and 2, MYS would appear to account for about two thirds of the total "education effect" in worldwide national productivity differences, and HLO academic achievement scores about one third. Thus our statistical analysis indicates *both* years in school and quality of schooling appear to matter to add to a country's productive potential.[24]

Coefficients in Tables 1 and 2 also suggest, in very rough terms, the contribution of our "four factors" to overall national economic productivity differences in the world today. As we have noted in earlier reports, the ratio of per capita GDP between highest and lowest productivity countries would be on the order of 100:1 (say, Switzerland vs. Burundi).

As a very rough benchmark—assuming a range of 8 years in MYS for adult populations around the world from lowest to highest, and a range of 350 points from lowest to highest for the mean national test scores for students—our illustrative modeling would suggest that "education", including both years of schooling and quality of education, account for more than a quarter but less than a third[25] of the total productivity gap between nations today, with other "human factors" such as health, urbanization, "business climate" accounting for most of the rest. In the arithmetic of modern economic development, any factor that accounts for that much of a country's overall economic growth can most certainly be regarded as a "big deal".

Note that we get analogously powerful predictive results from a model that assumes the economic impact of quantity and quality of schooling are "multiplicative" rather than additive, as in the previous model—self-reinforcing so to speak (see Table 3).

This model generates powerful results—the educational factor plus the other "human factors" in it predict well over 80 percent of intercountry differences in productivity levels, and the results are also highly meaningful from the standpoint

[24] In these models, the estimated influence on national economic performance from an additional year of school-time can be compared with the corresponding increase in test scores required for the same impact. This particular model implies about 30 extra test points are equivalent to an additional year of schooling. Existing research suggests an equivalence more on the order of 40-50 points in mean test point to each year of schooling. Our estimate above is not far from that range, and other models we developed generate results yet closer to it.

[25] About 29 percent in this example.

of “statistical significance” (the odds our modeled coefficients are not just random numbers).

But one feature is particularly noteworthy. Once again, *both* quantity *and* quality of education matter crucially in modeled national economic performance. In this model, however, improvements in quality and quantity of education have “cascading” upward effects: all else equal, for example, the estimated level of per capita productivity for a country with 14 mean years of schooling and mean test scores of 600 would be over 10 times greater than one with just 4 years of school and mean scores of 200.

Such an upward bending arc would look rather like what economists call “increasing returns to scale”.[26] Unlike the classic, more familiar economic concept of diminishing returns to scale (when additional land, capital or labor generate less than proportionate increases in production) “increasing returns to scale” refers to situations where additional factors of production result in disproportionate *increases* in output.

Such situations are no longer unfamiliar, either in real life or in theory. Indeed the concept of “increasing returns is to scale” figures centrally in some models of economic growth, such as Nobel Laureate Paul M. Romer’s “endogenous growth theory”[27], which holds that ideas and learning, unlike physical capital, can make cumulative or even exponential contributions to output. Our simple models in this report are far from conclusive evidence that increasing returns to scale are a feature of knowledge capital—but they are intriguing, and consonant with that proposition.

[26] This, by the way, is just what initial modeling revealed: further modeling might reveal the “multiplicative” impact of school quality and schooling quantity might be even greater.

[27] Paul Romer, “Endogenous Technological Change,” *The Journal of Political Economy*, Vol. 98, No. 5, Part 2, October, 1990, pp. S71-S102, https://web.stanford.edu/~klenow/Romer_1990.pdf

Table 3.

lnGDP/CAP = Math Primary scores * MYS BarroLee

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HLO Primary School Math Mean Score # ~	0.000412***	0.000197***	0.000195***	0.000180***	0.000174***	0.000118***	0.000125***
	(12.85)	(6.36)	(8.84)	(7.91)	(7.86)	(4.63)	(4.87)
Life expectancy at birth, total (years)		0.0657***	0.0269***	0.0242***	0.0259***	0.0286***	0.0309***
		(10.00)	(4.17)	(3.69)	(4.21)	(4.42)	(5.01)
Urban population (% of total population)			0.0236***	0.0237***	0.0218***	0.0269***	0.0248***
			(7.55)	(7.62)	(6.97)	(9.77)	(8.90)
FI Index				0.0909*		0.137**	
				(2.22)		(3.31)	
HF Score					0.0125***		0.0125***
					(4.07)		(4.08)
East Asia						-0.490***	-0.495***
						(-5.85)	(-6.48)
Latin America and Caribbean						-0.553***	-0.508***
						(-6.17)	(-5.55)
Constant	7.904***	4.121***	5.381***	4.976***	4.853***	4.525***	4.629***
	(43.13)	(10.13)	(14.96)	(14.62)	(16.14)	(13.91)	(15.67)
No of Observations	202	202	202	202	202	202	202
R-Squared	0.514	0.737	0.816	0.818	0.823	0.853	0.855
F Statistics	165.1	197.6	219.3	190.1	240.5	144.8	170.3

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Ideally, for examining the relationship between knowledge capital and national economic performance, we would want a reliably devised and administered and internationally standardized measure of tested knowledge and skills for the entire national population with near complete worldwide coverage. We lack this for now, and are likely to be lacking it for the foreseeable future. We suspect that such a dataset, if it existed, would show that tested knowledge and skills are superior to sheer years of schooling as predictors of national economic performance in national populations. But this surmise for the moment is untestable.

Meanwhile, with real existing data, we can see that information from student academic achievement, in tandem with information about the sheer amount of schooling national populations have obtained, can sharpen our understanding of the contribution the "education factor" is making to national productivity today—and the contribution it stands to make over the coming decade as well.

IQ, Academic Achievement, and National Economic Performance

It is not exactly a secret that questions about measured "intelligence"—identifiable differences in cognitive ability across national populations, and the possible relationship to national economic performance of any such measurable differences are political dynamite in intellectual and policy circles nowadays. This is unfortunate. Intellectual due diligence should require critical examination of potentially relevant evidence. Possibly IQ indicators and allied standardized testing could help in understanding the "human factors" underlying national economic performance.

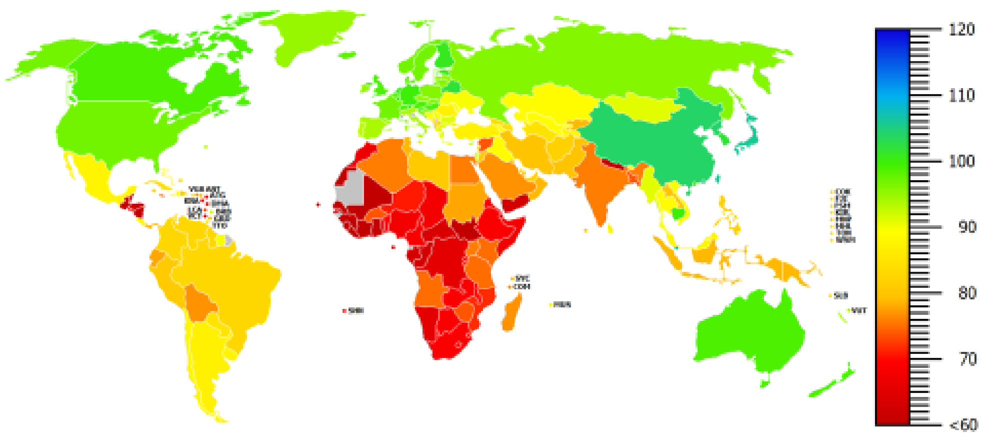
Though they naturally have limitations, carefully developed and standardized IQ tests convey meaningful information about examinees—just like the other historical branch of "psychometric" evaluations, standardized academic achievement tests. Few academic objections currently seem to be lodged against the "Flynn effect"[28], the apparent steady rise of mean IQs all across the world over the past century—yet this trend is derived from the very same data about cognitive performance that is decried in other contexts.

Several academic meta-studies have attempted to map out mean tested IQ across the world. Results of one such recent effort are displayed in Figure 9 below. Broadly speaking, IQ testing points to regional differences in measured cognitive performance: with countries in East Asia typically scoring in the top tier internationally, and countries in the sub-Sahara scoring in the lower ranking of the roster, and other regions in between. As it happens, the very same sorts of general regional differences are witnessed in all the international academic achievement datasets.

[28] Mohamed Nagdy, "The Flynn Effect: IQ Gains Over Time," in "Intelligence" at *Our World in Data*, https://ourworldindata.org/intelligence.

Figure 9.

IQ Scores Around the World - Lynn & Vanhanen 2012 Dataset



Source: Becker, D. (2019). The NIQ-dataset (V1.3.3). Chemnitz, Germany, https://www.researchgate.net/project/Worlds-IQ. (Accessed December 29, 2022

We thought it worthwhile to look at some of the international cognitive performance data and their possible bearing on national economic performance. We used the data from four recent peer-reviewed meta-studies[29], comparing them to the indicators we rely on in the rest of this report, and then examining statistically the relationship between IQ measures, academic achievement measures, and measures of national economic performance.

The overall relationship between these four IQ indicators and the other indicators in our report are presented in Figures 10a and 10b, which show "heat map" correlation triangles for the IQ indicators (mean national scores) with international achievement tests (mean national scores) and our other socioeconomic indicators, respectively. (See Figures10a and 10b)

[29] Becker, D. (2019). The NIQ-dataset (V1.3.3). Chemnitz, Germany, https://www.researchgate.net/project/Worlds-IQ.

Figure10a.

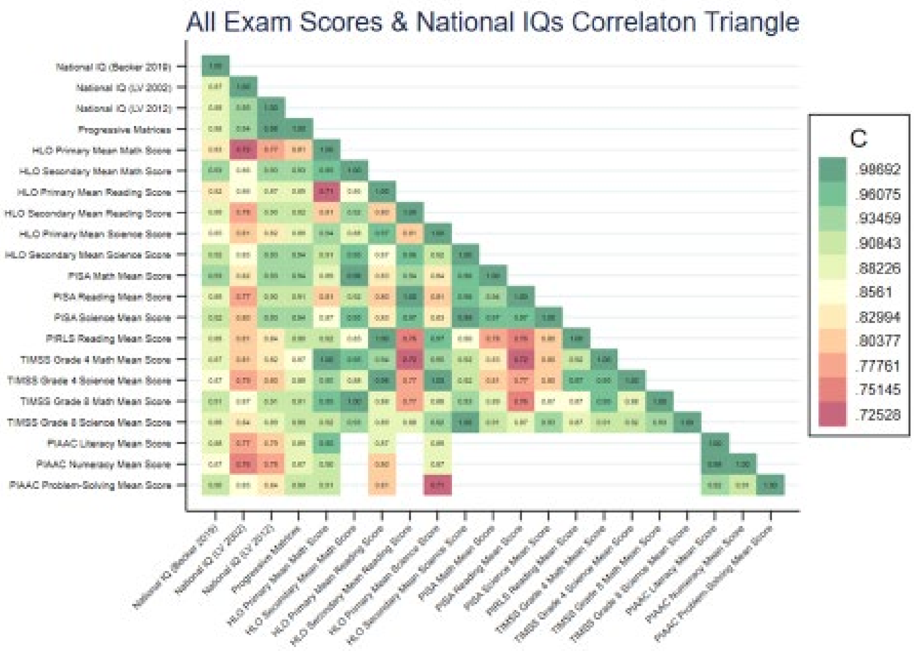
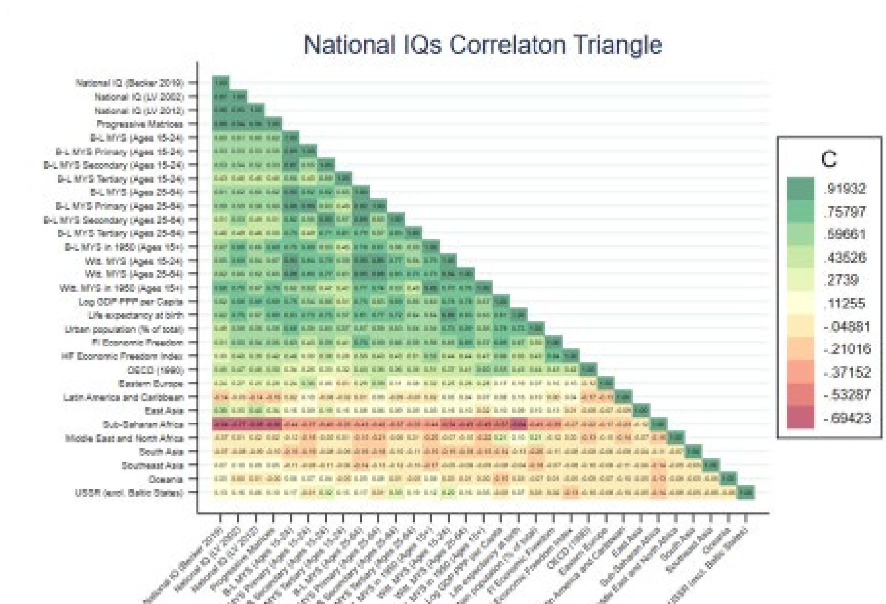


Figure 10b.



National mean scores in IQ and academic achievement correlate remarkably closely. This should not surprise. If these is any surprise in Figure 10a, it is how "weak"—relatively speaking—the association appears between some indicators of measured IQ and tested academic achievement. Mean scores form the Lynn-Vanhanen 2002 world IQ dataset and the HLO country data on mean primary math scores, for example, have an R-squared of "only" 50 percent. But framing the comparison this way only serves to emphasize how very powerful the correlation between IQ and international academic achievement appears to be. Of the 24 separate pairwise correlations between international IQ scores and mean HLO test scores in Figure 10a, fully 21 have R-squareds of over 65 percent, and almost half have R-squared correlations approaching or exceeding the remarkable level of 80 percent.

For its part, Figure 10b shows that IQ measures correlate strongly with per capita GDP—not nearly as strongly as with academic achievement, but with roughly the same degree of association as academic achievement indicators. IQ also seems to correlate roughly as strongly as mean test scores with some "human factor" indicators (life expectancy, "economic freedom"), and more strongly with some (urbanization), but less well with others (mean years of adult schooling).

An inquiry into the relationship between IQ and national economic performance naturally begs the question of *how* IQ might influence per capita GDP—i.e., the *mechanisms* through which any influence might be effected. Typically researchers have hypothesized that influence occurs at the level of the individual: with higher mean IQ scores raising productivity the way higher life expectancy, or schooling, or other "human potential factors" might. An alternative hypothesis is the "hive mind"[30] conjecture, which envisions a more macro, "institutional" influence: by this theory IQ in effect influences "business climate", since propensity to cooperate, defer gratification (savings and investment), etc., i.e. economically propitious dispositions, seem to track with IQ within observed populations.

Plausible as the IQ/national productivity connection may seem in the abstract, in our own analyses measured cognitive performance did not provide much if any additional predictive information in our existing models using "human factors" to predict per capita GDP. Table 4 below is illustrative: in these models, which use the same indicators as Tables 1 through 3, adding the IQ variable to models already

[30] Garett Jones, *Hive Mind*, (Redwood City: Stanford University Press, 2015), https://www.sup.org/books/title/?id=23082

including student test scores, adult years of schooling, life expectancy, urbanization and "business climate" does nothing to increase the models' predictive power (or R-squared). Moreover, in each of these models the relationship between added IQ and per capita GDP appears to be slightly *negative*—and unlike all the other "human factor" indicators, the calculated influence of the added IQ indicators were not statistically meaningful (their "coefficients" lacks "statistical significance")

Table 4.

Primary Math Score, MYS BL
(Psychometric Intelligence and School)

	(1)	(2)	(3)	(4)	(5)
HLO Primary School Math Mean Score	0.00202**	0.00229*	0.00250**	0.00209	0.00218
	(3.14)	(2.54)	(2.81)	(1.66)	(1.75)
Psychometric Intelligence and s-e	-0.00444	-0.00491	-0.00400	-0.00235	-0.000466
	(-0.66)	(-0.74)	(-0.59)	(-0.38)	(-0.08)
B-L MYS (Ages 25-64)	0.0786**	0.0818**	0.0745**	0.0840**	0.0786*
	(3.11)	(3.30)	(3.01)	(2.65)	(2.49)
Life expectancy at birth, total (years)	0.0277***	0.0266***	0.0252***	0.0162*	0.0150*
	(4.27)	(4.39)	(3.88)	(2.48)	(2.18)
Urban population (% of total population)	0.0242***	0.0223***	0.0242***	0.0239***	0.0255***
	(7.62)	(6.96)	(7.63)	(7.09)	(7.96)
Economic Freedom Index		0.0521***		0.0182**	
		(3.84)		(2.62)	
Economic Freedom of the World Index (pr)			0.0881		0.135*
			(1.93)		(2.14)
Eastern Europe				-0.0441	-0.0480
				(-0.41)	(-0.45)
Latin America and Caribbean				-0.358**	-0.352**
				(-3.11)	(-3.05)
East Asia				-0.584***	-0.522***
				(-4.58)	(-4.68)
Sub-Saharan Africa				0.0252	0.0673
				(0.12)	(0.32)
Middle East and North Africa				0.165	0.213
				(1.06)	(1.33)
South Asia				0	0
				(.)	(.)
Southeast Asia				-0.0859	-0.0450
				(-0.39)	(-0.19)
Oceania				0	0
				(.)	(.)
USSR (excl. Baltic States)				-0.363	-0.369
				(-3.05)	(-3.04)
Constant	4.516***	4.297***	4.227***	4.173***	3.988***
	(12.40)	(13.22)	(12.38)	(6.88)	(6.44)
Observations	202	202	202	202	202
R-Squared	0.821	0.828	0.823	0.865	0.862

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Our statistical analysis does not imply cognitive performance "doesn't matter" in national economic performance. Rather, it suggests that the information we might obtain from IQ scores in attempting to predict national economic performance is in effect already available from the other "human factor" indicators—so that IQ information appears to be superfluous for our purposes here.

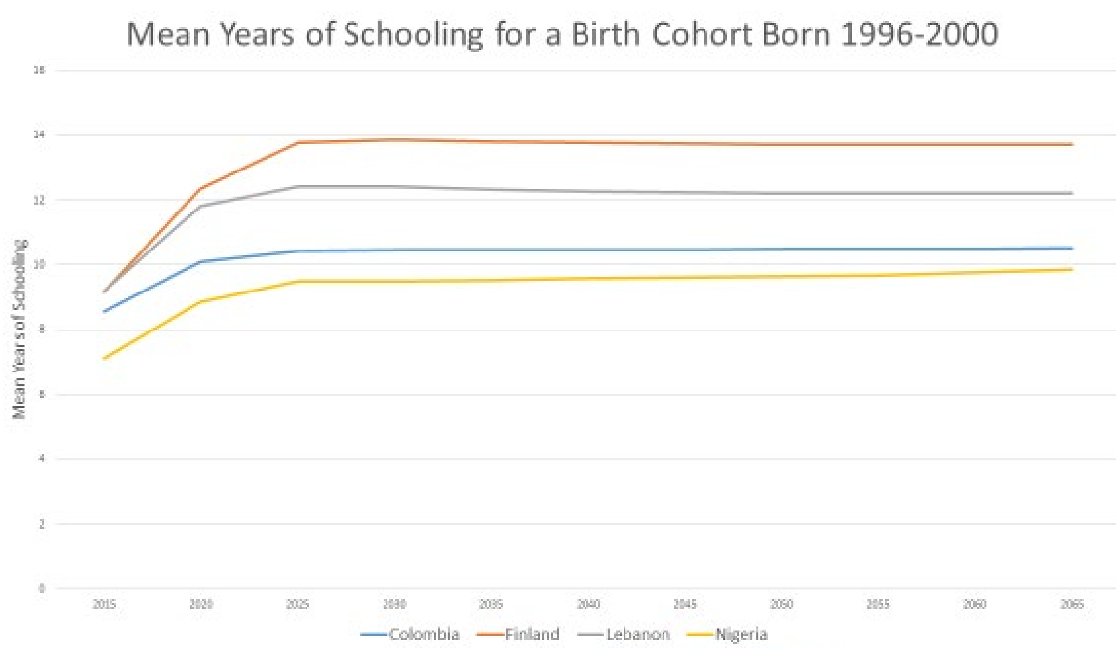
<u>Augmentation of Skills and Decline of Skills over the Life Course</u>

Our statistical analysis in this report uses what might be called "snapshot" data—student achievement scores and the other "human factor" indicators here measure conditions at a single point in time. But knowledge and skills change over a person's life course. Since PISA tests students on their reading, math and science skills at age 15, and PIRLS and TIMSS test students at even younger ages, these "point in time" scores may not correspond well with a person's knowledge and skills when they are of economically active ages—the period of adult life which may begin as early as one’s teens and continue into one’s 60s or beyond.

In a world where educational attainment has been rising almost everywhere, and seems set to continue increasing in the decades ahead, assessing one’s knowledge and skills at age 15 or earlier seems very likely to underestimate workforce potential for many.

Figure 11 below makes the point. It shows projected levels of educational attainment from ages 15 through 65 for babies born in 2000 in four select countries with varied socioeconomic profiles: Finland; Colombia; Lebanon; and Nigeria. PISA tests students at age 15—but between ages 15 and 30, the mean years of schooling for the cohorts that were PISA-tested in 2015 would rise by another 1.8 to 4.6 years in these four countries, or by 20% to 50%. In these and many other settings, current pupil test scores stand to under-predict subsequent economic potential of tested students.

Figure 11.



Source: World Bank, “GDP per Capita, PPP (Constant 2017 International $),” accessed July 16, 2021, https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD.

On the other hand, a pupil’s skills can decay in later life for reasons including lack of regular use. Decay of school skills over the life course is not a mere hypothetical. It appears to be taking place today on a large scale in lower income regions.

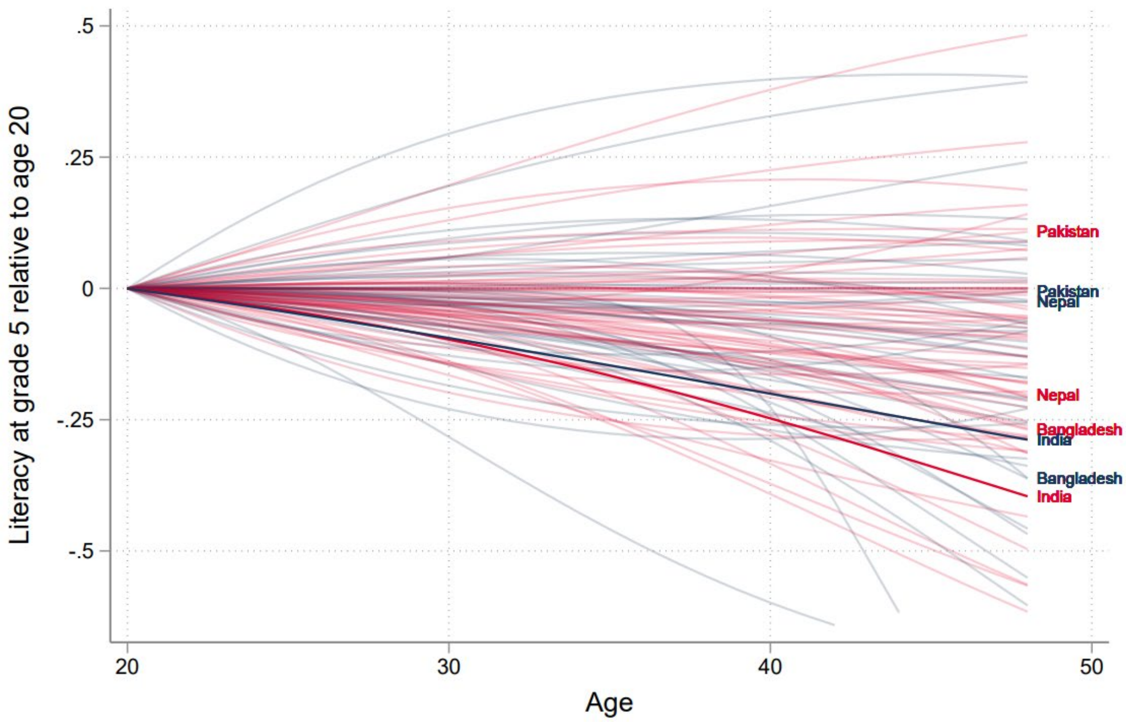
An original new study relying upon demographic and health surveys (DHS) utilizes literacy data to make the point.[31] Family planning agencies have been conducting DHS surveys for decades and they are available for dozens upon dozens of less developed countries; these do not test knowledge and skills but they do customarily inquire whether respondents can read and write.

DHS data from South Asia seem to suggest that literacy has tended to decline between ages 20 and 49 for women from Nepal, Bangladesh and India—and for Indian and Bangladeshi men as well. (See Figure 12) DHS studies also detail the same pattern in a number of African countries, especially among rural women.

Until we have life-course data on tested knowledge and skills for large numbers of countries, measurement of global knowledge capital profiles will be full of gaps. Taken together, as we have already seen, educational data and school quality data may afford a serviceable proxy for what we really need. Certainly adding some information about a population's time spent in school enhances the accuracy of predictions for national economic performance above and beyond what we can glean simply from a reading of student achievement in middle school by itself. But these will be inherently better suited to taking into account prospective improvements in skills than skills decay—and the decay of skills will also weigh on national economic potential, and by extension the global balance of power, in decades to come. Educators and economists have yet to focus on the phenomenon of skills decay and study it systematically; the phenomenon may prove to be an important missing piece in our understanding of knowledge capital.

[31] Alexis Le Nestour, Laura Moscoviz and Justin Sandefur, "The Long-term Decline of School Quality in the Developing World," Center for Global Development, September 2021, https://www.cgdev.org/sites/default/files/consultation-draft-le-nestour-et-al-school-quality-decline.pdf.

Figure 12.



<u>Determinants of Academic Achievement Test Performance Across Countries: Socioeconomic Factors and "Regional Effects"</u>

As we have seen, national academic achievement scores—or more precisely, the skills and knowledge for which they proxy—bear a clear and strong association with national economic performance. We have also seen that overall distributions of student achievement scores vary tremendously between national populations—so much, in fact, that top performers in certain countries would be counted as poor performers in certain other countries; while poor performers on some nations would rank as top performers in other nations.

We may therefore ask: what accounts for international differences in these economically portentous achievement test profiles? Are there systematic patterns to the differences between countries? What factors tend to make for high-scoring national populations—or for poor levels of national test score achievement?

In this section we examine correlates and determinants of academic achievement across the world, probing the correspondence between socioeconomic and other factors and national test score performance. We will here concentrate on the determinants of mean HLO scores, and summarize our findings from our statistical investigation here. (More detailed results can be found in Appendix E for those with an interest.)

The "correlation triangles" presented earlier in this report already demonstrated the strong correlation at the national level between mean test score outcomes and indicators for health, education, income and other developmental variables. In practice this means we can model predictors of national test performance just as we modeled predictors of national economic performance, re-arranging variables in our modeling equations so that they can show the contribution of those socio-economic factors to differences in international student test scores.

For predicting student achievement scores we use the four factors of 1) mean adult years of schooling, 2) per capita GDP, 3) measures of "economic freedom"; and 4) urbanization ratios. We also add so called "dummy variables" to see if there are any strong "regional effects" evident after controlling for theses socioeconomic factors. In this section we deploy those predictors on HLO "merged data"—mean test scores from both primary and secondary students for over 160 countries over the 2000-2017 period.[32] We model both the immediate, current correspondence between HLO scores and their presumptive predictors (i.e. "zero-lag" models), and the correspondence between the predictors and HLO scores ten years later ("ten-year" lag models).

In brief: our four factors predict (statisticians would say correspond with, or track) up to 70 percent of the observed international differences in mean HLO achievement test scores between countries and over time for reading, math and science—and for all national test scores pooled together. Although the specific modeled results differed somewhat depending on which indicators we selected[33], all our modeling told a story quite similar to the one shown below in Table 5, which uses our indicators to predict mean scores for all tested student populations in the HLO dataset. (See Figure 5)

[32] We also examined the relationship between the predictors and primary/secondary achievement separately—as noted already, the correlation between socioeconomic factors and student achievement was consistently stronger for primary school results. Just why this should be the case is a question for future investigation.

[33] With two indicators for years of schooling, two for "economic freedom", and 10 for HLO mean years of schooling (3 primary for reading/math science, 3 secondary and 3 merged for the same, and one total pooled indicator), there were a lot of permutations for any of our modeling "families"—up to 40, to be precise.

Table 5.

All Mean Score = MYS BarroLee + HF score + …

	(1)	(2)	(3)	(4)	(5)	(6)
BL MYS (Ages 25-64)	18.95***	13.63***	12.86***	12.31***	6.389***	6.654***
	(24.33)	(12.07)	(11.44)	(11.31)	(5.28)	(6.31)
Economic Freedom Index	1.275***	0.422	0.298	0.486	0.212	
	(5.88)	(3.71)	(1.58)	(1.95)	(1.43)	
Natural log of GDP PPP per capita-2017)		24.01***	15.14***	23.06***	16.54***	20.14***
		(6.08)	(3.08)	(4.83)	(3.39)	(6.28)
Life expectancy at birth, total (years)			1.443**	1.951***	1.157	
			(3.18)	(5.81)	(1.93)	
Urban population (% of total population)				-0.605***	-0.223	
				(-4.34)	(-1.37)	
East Asia					47.20***	48.01***
					(7.75)	(8.27)
Latin America and Caribbean					-52.64***	-54.42***
					(-9.55)	(-11.33)
Middle East and North Africa					-61.77***	-65.83***
					(-10.72)	(-11.64)
South Asia					-100.1***	-105.0***
					(-7.33)	(-7.39)
Sub-Saharan Africa					-66.73***	-77.28***
					(-4.99)	(-7.69)
Eastern Europe					1.562	
					(0.29)	
Southeast Asia					-23.89*	-22.86**
					(-2.56)	(-2.73)
Oceania					-133.0***	-139.6***
					(-13.86)	(-19.58)
USSR (excl. Baltic States)					-1.463	
					(-0.14)	
Constant	191.5***	65.12*	37.33	-23.18	162.7***	221.5***
	(26.78)	(2.63)	(1.47)	(-0.83)	(3.58)	(7.34)
No of Observations	475	441	441	441	441	441
R-Squared	0.646	0.679	0.688	0.697	0.794	0.798

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

In this illustrative example, our four indicators for adult years of schooling, per capita income, life expectancy and "economic freedom" tracked with nearly 70 percent of the global observed differences in mean achievement scores over 2000-2017 (the period covered by the HLO dataset). We also see that mean years of adult schooling and per capita appear to be the strongest correlates with HLO achievement scores, providing the most predictive information on national outcomes. This makes intuitive sense, since all else being equal we would expect countries with higher income levels to be able to purchase a better quality of education for students, and we would likewise expect countries with strong established schooling systems, wherein high school and at least some higher education is the norm, likewise to be better suited at imparting high quality education than places where school systems are more rudimentary and education is much more limited on the part of the adult population.

While our simple four factor models manage to predict a creditable ~70 percent of all international variation in measured student achievement, this leaves about 30 percent of those differences still to be explained—a rather considerable residual. What accounts for the remainder of these international performance differences?

An important part of the answer has something to do with *geographical location*. In our statistical analysis we found that knowing what region of the world a country happened to be part of provided meaningful and consistent information about students' knowledge and skills. The impact of geographic location on academic achievement is revealed by so-called "dummy variables", which show the "educational effect" associated with residence in a given region, after controlling for key elements of national development. Our study found these regional effects to be remarkably stable across our many "models" examining student achievement—and some of these effects were also remarkably large. Evidently, some parts of the world "punch above their weight" in scholastic achievement, while others turn out to be underperformers.

For example: in Table 5, countries in the South Asian region tend to score about 100 points less on mean achievement scores than would be predicted on their socioeconomic profiles alone. This is a big shortfall on a test whose mean global level is notionally scored at 500 points; it would imply students from this region end up with the equivalent of up to two and a half years less learning from their time in school than their notional "typical" global counterpart.

By the same token: sub-Saharan countries *per se* score about 70 points less in our models here than developmental factors by themselves would predict; MENA countries, at least 60 points less; countries from the Latin America/Caribbean regions, 50-plus points less. Given the rather regular "bell curve" distributions of test scores from one country to another, this also means a much smaller share of "high academic achievers" for these populations, and a correspondingly larger share of students with poor grasp of the basic skills in reading, math and science.

Conversely, countries in the East Asian region routinely and regularly over-perform academically in relation to their socioeconomic or developmental profiles. In the Table above, for example, coming from East Asia adds nearly 50 points to a country's predicted tested level of knowledge and skills—the equivalent of over a year's schooling. Since our "regional effects" are scaled implicitly against OECD countries from Europe and North America (the main spots on the map we do not use "dummy variables" to track) this implies that East Asia countries get more for less than their affluent Western counterparts.

In other words, all other things being equal, our models suggest that students from East Asian societies today will match or exceed the national mean scores of their Western counterparts even if their societies have decidedly lower income levels, or lower levels of adult educational attainment, or both. (Note the potential relevance of these findings for the case of the People's Republic of China.)

When we add in the "regional effect" to our models, their predictive power rises from about 70% to around 80% percent or even slightly higher. As with our other models, these ones are almost as good at predicting outcomes ten years into the future on the basis of current data as they are in predicting current outcomes.

But 80 percent accuracy in predictions would still leave almost a fifth of the international variation in measured student knowledge capital unexplained. And the fact of such major regional differences in academic performance raises the question of why they appear to be such fundamental feature of the global demographic landscape.

Remember—these regional affects speak to discrepancies in knowledge capital with very considerable economic implications. To go by our illustrative models in this paper; two countries—one from East Asia, the other from South Asia—with the same developmental characteristics in health, education, income and "economic climate" could expect to see a difference in mean national achievement scores on the order of 150 points—a discrepancy that in itself would be predicted in our models to results in a per capita GDP gap ten years hence on the order of 40%.

The matter of regional effects in academic performance (and by extension, knowledge capital) is recognized by psychometricians, students of education policy, and educational economists—and the phenomenon is usually chalked up to some combination of "cultural" and cognitive" differences. But few have been interested in parsing the matter further.

Yet it is possible in principle to do so—simply by adding cognitive or IQ indicators to our simple models. And the argument for adding a cognitive measure is compelling. For the geography of "regional effects" from Table 5 closely maps the global terrain of measured mean national IQ differences in Figure 9. Just as East Asia has a positive "regional premium" in its predicted achievement scores, so it records relatively high IQ scores on international testing. Conversely, South Asian and African "regional effects" are negative—they subtract from the scores that would be predicted by socioeconomic factors alone—and these regions also test more poorly at the moment than the global norm in standardized cognitive examinations. Bringing IQ indicators into consideration may therefor help us parse out how much the regional residual is in fact a proxy for international differences in cognitive factors.

We show an illustrative result from such modeling in Table 6. (See Table 6)

Table 6.

Secondary Math Score, MYS Witt
(Raven's Progressive Matrices estimates)

	(1)	(2)	(3)	(4)	(5)	(6)
Raven's Progressive Matrices estimates'	7.221***	7.220***	7.287***	7.249***	6.653***	6.651***
	(21.60)	(20.72)	(20.57)	(20.78)	(17.99)	(18.34)
Witt. MYS (Ages 25-64)	0.426	-0.0399	-0.613	-0.702	-2.935*	-2.657*
	(0.39)	(-0.04)	(-0.54)	(-0.71)	(-2.40)	(-2.30)
Life expectancy at birth, total (years)	-0.0923	-0.156	-0.197	-0.250	1.492**	1.421**
	(-0.27)	(-0.41)	(-0.51)	(-0.66)	(2.88)	(2.75)
Urban population (% of total population)	-0.112	-0.167	-0.200*	-0.188*	-0.0479	-0.0497
	(-1.30)	(-1.89)	(-2.31)	(-2.08)	(-0.48)	(-0.50)
Natural log of GDP PPP per capita-2017)		2.280	1.229	1.526	1.266	1.922
		(0.71)	(0.36)	(0.46)	(0.35)	(0.55)
Economic Freedom Index			0.298		0.284	
			(1.66)		(1.52)	
Economic Freedom of the World Index (g-)				4.252		2.518
				(1.92)		(1.23)
Eastern Europe					6.678	5.874
					(1.60)	(1.40)
Latin America and Caribbean					-14.65**	-12.97**
					(-3.02)	(-2.65)
East Asia					25.30***	25.81***
					(4.60)	(4.75)
Sub-Saharan Africa					33.54*	34.42*
					(2.49)	(2.57)
Middle East and North Africa					-12.31**	-10.83*
					(-2.71)	(-2.32)
South Asia					23.95**	25.41***
					(3.29)	(3.43)
Southeast Asia					-2.041	-0.688
					(-0.31)	(-0.10)
Oceania					0	0
					(.)	(.)
USSR (excl. Baltic States)					21.32**	21.13**
					(2.69)	(2.63)
Constant	-184.9***	-194.5***	-197.5***	-204.0***	-252.0***	-258.2***
	(-7.50)	(-7.48)	(-7.67)	(-8.12)	(-5.41)	(-5.46)
Observations	447	443	442	443	442	443
R-Squared	0.868	0.868	0.869	0.869	0.890	0.890

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Several points here are worth mentioning. First: in this example, bringing the cognitive factor into the mix raises the explanatory power of our simple models to nearly 90 percent—a very high level of predictive power in the social sciences for any complicated phenomenon. Second: taken together, regional effects and cognitive effects explain for as much as two third of the "unexplained residual" in international test scores that remained after socioeconomic factors were taken into account. Third, the social and economic indicators lose most of their specific "statistical significance" in these larger models: unsurprisingly, given the high degree of correlation among the increasing number the indicators in the models that include both IQ and regional effects. But fourth, what is pertinent is the high remaining statistical significance of both IQ and regional effects—the "nature" and the "nurture" factors in national deviations from socioeconomically predicted outcomes.

Finally, controlling for IQ or cognitive factors dramatically reduced the estimated "regional effects" in our models (though these regional effects remain highly meaningful in a statistical sense even after adding and IQ indicator). For instance: the South Asia effect is slashed by over two thirds; and the MENA and Latin America/Caribbean effects, by even more. The big positive East Asia effect, quite notably, is cut roughly in half.

One interpretation of such results would be that "cultural" and "cognitive" factors *both matter* in explaining why some places punch above, or below, their socioeconomic weight when it comes to academic achievement. Cultural factors could include (among others) values, traditions and familial incentives that prize academic excellence. Cognitive factors would include traits, including inherited traits, which facilitate such things as problem solving, reasoning ability, and recall capabilities.[34] Deviations in national knowledge capital from levels predicted by levels of socioeconomic development alone appear to be largely explained by some mix of cultural and cognitive factors, as our study illustrates: though the exact balance may change between countries and over time, and investigating that balance is a question beyond the scope of this report.

<u>The Mystery of China's Performance In Academic Achievement Tests</u>

Our study of knowledge capital and its statistical proxies—their determinants, and their relationship to national economic performance—now brings us full circle: back to our starting point in this study, the absolutely remarkable scores over the past decade for randomly sampled students from the People's Republic of China in international standardized examinations of academic achievement.

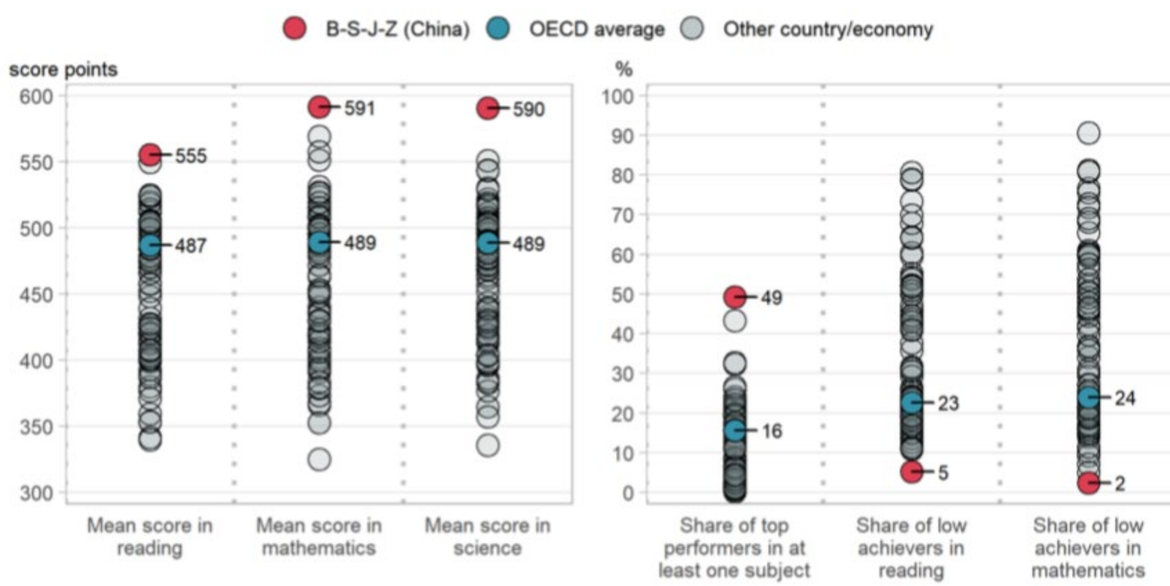
These soundings come from four successive waves of testing (2009, 2015, 2015, and 2018) by PISA, arguably the world's leading authority for developing, administering, and evaluating such tests. PISA has reviewed the results from China and formally validated them.

[34] It is possible that inherited traits might be influenced by historical traditions and practices, possibly including the prizing of academic excellence—or by improving living standards—or by both; in any event the "Flynn effect" points to rising cognitive capabilities the world over, temporal regional differences notwithstanding. Cf. Nagdy, "The Flynn Effect" in "Intelligence" at *Our World in Data*, https://ourworldindata.org/intelligence.

The results from China are simply stunning. In 2018 students from China were not only the top performers in all tested fields—reading, math and science—but far ahead of the runner up in two of those areas. (See Figure 13) As we see by comparison with OECD countries and other top performers, the students from China earned far and away the highest cumulative mean scores from these three areas of testing—placing the Chinese student test takers not only distinctly ahead of perennial star performer Singapore, but far above such poster children of international educational excellence as Estonia and Finland, and still further ahead of the USA. (See Figure 14) The gap between USA performance and that of the Chinese student is on the order of two years of learning in each subject.

Figure 13.

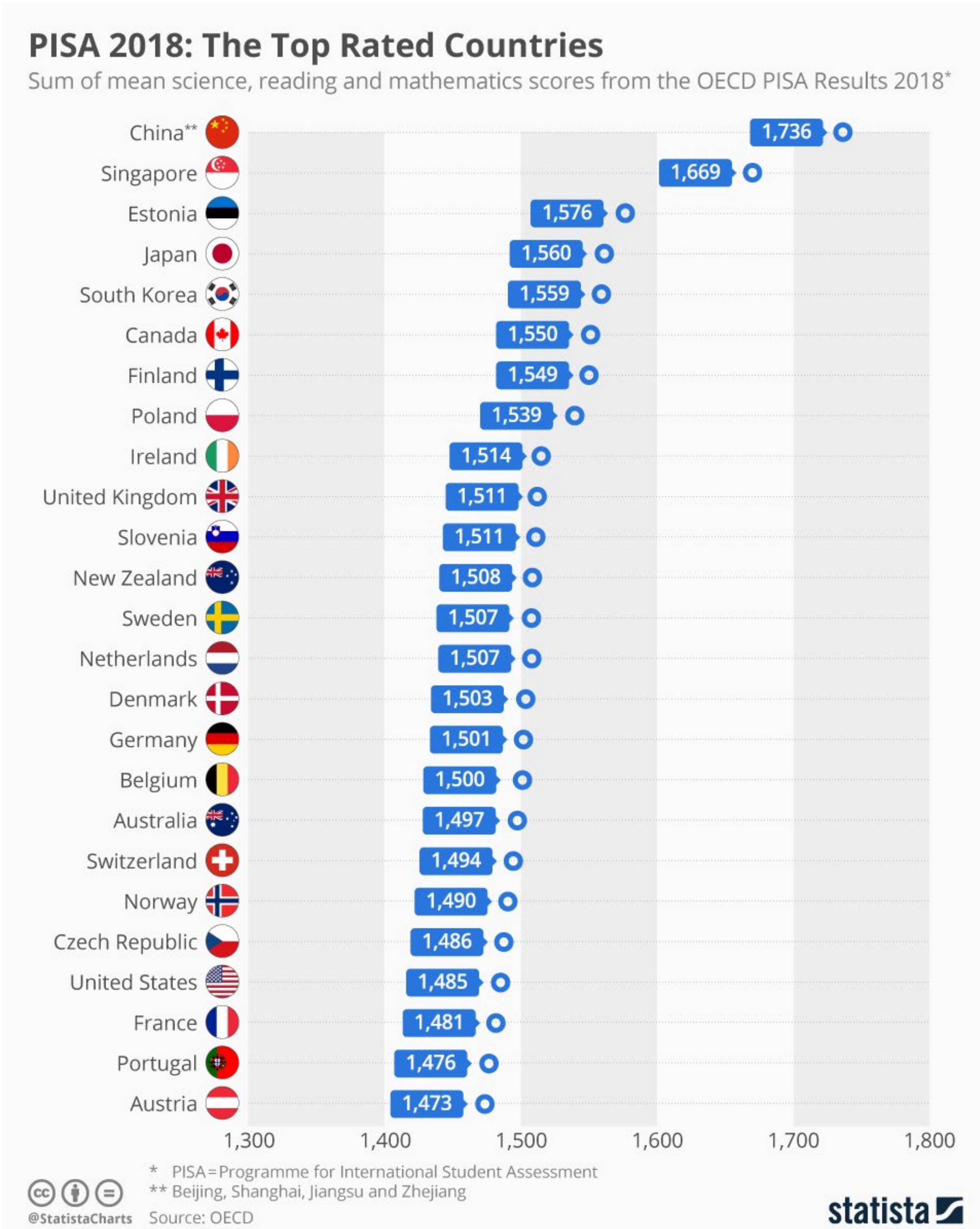
Figure 1. Snapshot of performance in reading, mathematics and science



Note: Only countries and economies with available data are shown.
Source: OECD, PISA 2018 Database, Tables I.1 and I.10.1.

Source: https://edtechchina.medium.com/china-1-on-2018-pisa-is-the-country-really-an-education-powerhouse-as-the-rankings-suggest-8b626cc1ae92

Figure 14.



Source: https://www.statista.com/chart/7104/pisa-top-rated-countries-regions-2016/

Though PISA has vouched for the Chinese scores, they deserve further scrutiny. We offer some in this section, posing and attempting answer several pressing questions about these results.

The first question concerns the representativeness of the data for “China”. PISA testing has taken place in just a small part of this enormous nation: a small and constantly shifting pastiche of the most developed portions of the Mainland, starting with just Shanghai in 2009 and 2012, then expanding to Beijing-Shanghai-Jiangsu-Guangdong in2015, and most recently switching to Beijing-Shanghai-Jiangsu-Zhejiang in 2018. But there are other unanswerable questions about testing representativeness even within these highly unrepresentative regions. The method by which given schools, particular classes, and even specific students were selected by the PRC Ministry of Education for PISA evaluation remains opaque.[35] In the country where a recent premier has complained about “man-made” statistics, authorities do not enjoy benefit of the doubt when methodologies are murky. There is also the hardly trivial issue of test prep for the students selected. This matter is also murky: according to some educational experts PISA rules allow for “crash courses” before testing in selected schools, although these PISA protocols do not seem readily accessible online. Just how Chinese authorities would have utilized such permitted leeway is impossible for outsiders to know.

Second: there is the question of anomalies and curiosities from the China data themselves. The short version is that reported academic achievement for China’s selected regions jumped between 2015 and 2018 in a sharp and unusual manner—one we contrast with the regularity of PISA trends for Massachusetts and Moscow, academically the top-performing regions of the USA and the Russian Federation, respectively. (We highlight some of these in Appendix F for those with an interest.)

Third is the question of what China’s true PISA tested knowledge capital scores might look like for truly a genuinely representative all-China sample of 15-year-old students. We try to approximate those levels through the statistical models we have developed for predicting worldwide mean national test score levels, as recorded in the World Bank HLO dataset, on the basis of socioeconomic factors, ‘regional effects”, and cognitive indicators.

---

[35] Among other questions: were Shanghai and Beijing migrant children from disadvantaged rural regions included in local testing? By Chinese policy these migrants are “non-*hukou*” and therefore not *de jure* residents of those cities.

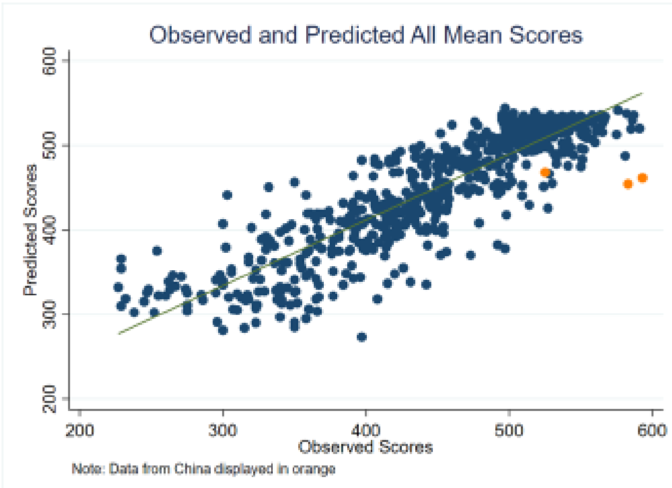
The HLO dataset only extends up to 2017, so it does not include the PISA 2018 findings for China. But it does include China scores for 2009, 2012 and 2015. We start by comparing those reported values to the predicted values for all-China for those same years from to our global models for estimating international knowledge capital as proxies by international standardized student achievement tests. (We have many models by which to make such estimates, and they included in Appendix F for readers who may have an interest.)

Our many models for estimating global determinants of student achievement calculate slightly different "predictions" for all-China test results for 2009, 2012 and 2015, but they share two common characteristics. First: they estimate China's "true" level of proxied knowledge and skills to be below those reported to PISA for those years—and typically, far below. Second: they estimate all-China's mean test scores to be rising rapidly, in consonance with the country's extremely rapid pace of socio-economic modernization.

The "main" model below, which predicts the average for mean scores in reading, math and science combined on the basis of socioeconomic factors and 'regional effects", seems most apposite for an initial illustration of the discrepancies between China achievement test score reports and estimated all-China results. (See Figure 15)

Figure 15.

Modeling All-China Mean Test Scores for Academic Achievement In Relation To Reported PISA Scores for Tested China Students: 2009, 2012, 2015



China Scores		
	Actual	Predicted
2009	583	454
2012	593	461
2015	525	468

All HLO Scores = 7.07 (Witt MYS) + 14.55 (log GDP) + 1.64 (life expectancy) + regional dummies (less East Asia) + 146.17
(7.00) (4.96) (4.07)
R-squared: 0.7770 Number of Observations: 701 Number of countries: 148

 5

As we can see, the gap between reported scores for China students and the estimated all-China scores that our model calculates is dramatic—ranging from roughly 130 points in 2009 and 2012 (when only Shanghai was being tested) to "only" 58 points in 2015 (when Beijing, and the provinces of Jiangsu and Guangdong, were added to "China").

If these estimates of China's "true" level of proxied student knowledge capital proved roughly accurate, where would that place China internationally? Table 7 provides an impression. China's scores would fall far below Vietnam's—a country currently with a decidedly lower level of economic development, but also a noted and increasingly studied Asian "over-performer" in student achievement. On the other hand, China would score above almost all the countries of South Asia, Latin America/Caribbean and MENA region, excepting virtually only Chile and Cuba. Interestingly enough, China would quite possibly look to be in the same league as the sub-Sahara's top academic achievers, Gabon and Kenya. And to go by the HLO figures, Turkey and Malaysia might also be reasonable comparators for our estimated all-China population when comes to "knowledge capital" as proxied through academic achievement exams. (See Table 7)

This selection of would-be comparators may help to place China's reported achievement scores in greater international perspective. There is a wide range of countries from a wide range of developmental levels that have actually scored in the range where we estimated China could have fallen, given our predictive models for HLO scores: those countries include what the World Bank classifies as "lower middle income" and "upper middle income" economies[36]. (The Bank categorizes China as an "upper middle income" economy.) The point is that our statistical models suggest China may have performing much as one would expect a country in its broad income classification to be performing—at least in the 2009-2015 period.

Table 7.

Selected Illustrative Comparator Countries: Actual HLO Scores vs. Modeled All-China Scores

Country	HLO Mean National Score	2020 Per Capita GDP (PPP 2017 Intl Dollars)
Vietnam	519	8,200
Mauritius	472	20,522
All-China Modeled 2015	468	12,612 (2015)
Malaysia	468	26,472
All-China Modeled 2012	461	10,370 (2012)
Turkey	459	28,393
Gabon	456	14,321
Kenya	455	4,340
All -China Modeled 2009	454	8,069 (2009)
Cambodia	452	4,192
Mexico	429	17,852

Sources: https://blogs.worldbank.org/opendata/harmonized-learning-outcomes-transforming-learning-assessment-data-national-education; https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD; Figure 15 in this report.

[36] "The World by Income and Region," *The World Bank*, https://datatopics.worldbank.org/world-development-indicators/the-world-by-income-and-region.html.

But what about China 2018? We can use the model in Figure 15 to predict corresponding all-China men student scores, based on the changes in China's level of socioeconomic development over those years in the datasets we use for this study. In all, improvements in adult years of schooling, per capita income, and life expectancy between 2015 and 2018 would be predicted to increase the all-China student mean score—but not radically. By our calculations, those social trends would track with an increment of about 5 points—meaning the all-China modeled mean score average for all three tests would be a little over 470 points (472). Part of what this calculation indicates is how very slowly student achievement scores can be expected to improve even in places that are undergoing hyper-rapid development: a point that underscores our previously expressed skepticism regarding the great leap officially reported for China student achievement scores in the scant three years between 2015 and 2018.

Our modeled 2018 all-China averaged achievement scores estimate of c. 472 can be contrasted with the corresponding PISA reported China student aggregate average of 579. That discrepancy of 107 points would point to a radically different 2018 "knowledge capital" profile in all-China from the one reported to PISA for select locations within China.

One last stylized comparison may be worthwhile for the China data. This would be a notional contrast between academic achievement profiles for Singapore 2012 and Shanghai 2012, relying upon reported PISA data, socioeconomic data on Singapore, our models for predicting academic achievement, and some back-of the-envelope estimates about the developmental differences between Singapore and Shanghai c. 2012.

Since both Shanghai and Singapore are counted as "East Asia region" in our models, there would be no difference in regional effects. And if we treat both places as more or less completely urbanized, we can ignore any differences with respect to that variable.[37] The differences pertinent to the model in Figure 15 would be: per capita income; adult education; and life expectancy.

[37] According to official statistics Shanghai is not quite 100 percent urban—the entire municipality is larger than the state of Delaware—but we treat it as entirely urban nonetheless for purposes of simplicity. We also exclude differences in "business climate not least because we do not have any measurements for Shanghai in our "business climate datasets". It is a fair bet that "business climate" would be judged more auspicious in Singapore than Shanghai—and such a difference would generally suggest more favorable modeled scores for Singapore in our method. Here again we disregard a factor that would likely weigh in Singapore's advantage in our modeling.

Say for the sake of argument that Shanghai's PPP-adjusted productivity level was two third of South Korea's in 2012 (arguably a generous assumption): that would mean its per capita output level was roughly one third of Singapore 2012. With regard to the difference in mean years of adult schooling—let posit, again generously, the Shanghai 2012 was only one year behind Singapore. And with respect to the Shanghai-Singapore life expectancy at birth—let us assume there was none at all in 2012 (this last assumption not as generous to Shanghai as the previous two, however, since mortality levels in the two populations appear to be quite similar).

By these notional parameters, our model above would tell us to expect Shanghai's mean aggregate score average for reading, math and science to be about 20 points *lower* than Singapore 2012. Since Singapore aggregated HLO mean score reading/math/science average for 2012 was 564 points, that would mean a Shanghai 2012 score around 545 would track would our model. On a global basis, that would count as a very high score. That same year, for example, perennially high-performing Finland's corresponding figure was just under 540. But the actual figure Shanghai 2012 reported to PISA was 593—almost 30 points *higher* than Singapore 2012; over 50 points higher than Finland 2012; and fully 70 points higher than Massachusetts 2012, America's top-performing region year in and year out.

Now it is true that some around the world *do* happen to over-perform, and massively—as we have seen, the global map is dotted with conspicuous outliers in tested scholastic performance. This is part of the fascination in attempting to explain academic achievement.

Our illustrative back of the envelope calculations are not determinative: they merely indicate the scale of over-performance that would be required for Shanghai's reported 2012 numbers to be credible. What our thought experiment here suggests is that judged on their own merits, China's reported numbers for Shanghai appear "highly ambitious". And since Shanghai has long been the most economically developed province in China, surmises about knowledge capital in the rest of China might advisedly be adjusted accordingly.

Concluding Observations And Next Steps

Since the beginning of the 21st Century, the world has witnessed an explosive surge in detailed data about the human potential of the individuals who make up the planet's national populations. Much of this information bears directly on national economic potential, and thus on the global balance of power. Our project aims to exploit the potentialities of this data revolution for taking strategic demography "beyond the headcount approach".

Even a decade ago this report would have been impossible because we would have lacked too much of the required information on student achievement. Indeed we still lack the data we would *ideally* wish for a dive into knowledge capital: namely, robust and comprehensive information on the economically relevant knowledge and skills of both youth and adult populations for most countries around the globe. Instead we use just data gleaned from the aperture of tested students. And as we all know, replication of testing on identical populations cannot be guaranteed to elicit identical results.

Given those inherent limitations, our statistical investigation into knowledge capital was perforce more "untidy" than previous phases of this project. We suspect that with better data, and better statistical instruments, we might have been able to demonstrate an even deeper and more powerful relationship between a country's knowledge and skills and its economic performance than we present in this report. But untidiness notwithstanding, those very data nonetheless proved capable of illuminating the link between knowledge capital and national economic performance—and showed that relationship to be powerful, ripe with strategic import.

To summarize and recap briefly our main findings:

- In depth examination of statistical information on knowledge capital– knowledge and skills as proxied by national results from international standardized academic achievement testing—confirms that a country's student test scores provide highly meaningful information about both current differences in national economic potential (per capita productivity) and also national economic potential ten years in the future.
- Our best models suggest (after controlling for other developmental factors, including a population's total years of schooling) that a difference of 100 points on a country's mean scores in student achievement tracks with about

a 25 percentage point difference in national per capita productivity levels ten years hence.

● Our analysis further suggests that taken together, quality of education (as measured by student achievement scores) and the quantity of education (as measured by years of schooling for the working age 25-64 adult population) account for somewhere between a quarter and a third of economic growth in our postwar era—that is to say, that taken together, measured academic achievement and measured educational attainment can explain close to thirty percent of the overall productivity disparity between the world's most productive and its least productive economies.

● There are hints and clues in the data we examined that the impact on national economic potential of more schooling and higher academic achievement may be *multiplicative*—that" knowledge capital" may enjoy *increasing returns.* One manor school of modern economic thinking posits just such a relationship: the New (aka Endogenous) Growth Theory, associated with Nobel Economics Laureate Paul Romer, holds that unlike capital and land, returns from ideas and knowledge are not subject to diminishing returns. Our findings, while only exploratory, appear to be consonant with that proposition.

● Two major trends are currently transforming the global terrain of knowledge capital—and both will shape the global economic balance. The first of course is *the continuing expansion of education*. With a pronounced and seemingly insatiable demand for more education almost everywhere, national levels of educational attainment seem on course to keep increasing all around the world. This will mean the skills and knowledge of working age populations that power national economies on the whole are set to increase in the decades immediately ahead.

But the second trend—less familiar but no less real—is *skills decay in adult populations*. New research has identified this second phenomenon in low income countries, and is beginning to scope its dimensions, which include loss of literacy amongst those who could once read and write. We should expect that a parallel syndrome of skills decay exists in richer countries, too.

Our models from this project can crudely, but serviceably, approximate the implications of increasing intergenerational educational attainment for national economic performance: by estimating the impact of conjoint changes in schooling levels and student achievement. But understanding the implications of skills decay in developing and affluent countries—its prevalence and severity—is an undertaking yet even to be organized, and a problem whose strategic consequences remain all but completely unexamined.

● In this study we also modeled the determinants of academic achievement at the national level. Socioeconomic factors—levels of national schooling, income, urbanization, and "business climate"—could account for up to 70 percent of inter-country differences in mean achievement scores. But that still left a large unexplained residual, with many countries scoring considerably above, or below, the level predicted by socioeconomic factors alone.

But by adding two additional factors, our models turned out to predict up to 90 percent of the academic achievement differences between countries.

The first of these was *geography*. Even after controlling for other factors, strong regional patterns to academic achievement remain. Countries from the East Asian region over-perform academically; conversely, South Asian, sub-Saharan, Middle Eastern and Latin American countries generally under-perform.

The second was *cognitive ability*. International differences in measured IQ levels provide meaningful independent predictive information about differences in national academic achievement, even after socioeconomic factors and geography are taken into account.

The distinct, statistically significant contributions of these two separate factors to explaining international variations in tested student skills and knowledge would be consistent with the argument that both "culture" and innate population traits contribute in national academic achievement—that "nurture" and "nature" both matter, so to speak.

● Finally, we examined the spectacular student test scores the PRC has reported to international testing authorities, scores those same international testing authorities have validated. In the PISA 2018 wave, 15 year olds from four provinces in China registered the world's very top scores in reading, math and science—scoring far above the average for OECD countries that same year. We used our models for predicting international academic achievement to see how an East Asian country with China's socioeconomic profile would be expected to fare in such testing. Estimates from those models assigned China a level of overall student achievement comparable with that recorded in places like Turkey, Malaysia, and Mauritius—that is to say, in other middle income economies. Our most relevant models, in other words, suggested that China's nationwide achievement levels in student knowledge and skills tests would be expected to look more or less like those of a country at its current developmental level.

● We modeled these all-China scores because we lack actual all-China data on academic achievement. However, officially reported China achievement data did permit us to use our models "reality check of sorts. In the 2012 PISA wave, the PRC participated—but only with students from Shanghai. Shanghai is unrepresentative of the rest of China: it has long been the country's wealthiest and most highly educated region. But Shanghai is arguably comparable to other places PISA tested—in particular, the city state of Singapore, which is always near the very top in global knowledge and skills testing.

The similarities between Shanghai and Singapore are striking: both are rapidly developing East Asian urban areas with high levels of health and education. But income and education are lower in Shanghai: and in our models such differences would presage distinctly lowers score for Shanghai than for Singapore, not higher scores, as the PRC reported to PISA in 2012.

To be sure: our ballpark estimates for Shanghai would have implied highly impressive levels of achievement—our notional scores for Shanghai would have been in the same league the as the actual scores for Finland, long a cynosure for the educational policy world. Nevertheless, our own suggestive estimates for Shanghai's educational achievement were roughly 50 points below the levels reported for Shanghai to PISA—a gap in the same league as the one that currently separates the actual scores of France and Turkey.

Outside the confines of the PRC government, there is precious little in the way of data on nationwide patterns of academic achievement for China, reliable or otherwise. Our statistical analysis in this report can only tell us what global patterns would suggest about academic performance in China, or more specifically for a country's modeled profile—not what that profile actually is. But those modeled results offer at least a starting point for addressing the mystery of academic achievement in contemporary China.

Based on our modeling of socioeconomic factors and “regional effects”, we arrive at estimates of tested knowledge and skills drastically lower for China nationwide than those reported for China students to PISA. Instead of being “top of the class”, China would by these numbers fall closer to “middle of the pack” internationally; the nation as a whole would test about as far below Singapore as countries like Nicaragua and Burkina Faso test below Turkey (or by our models, China itself).[38]

But because we lack actual test data for the vastness of China’s interior, we cannot categorically know this is the case. And there is one family of models we developed that would assign OECD-level academic achievement scores to China as a whole. These are the ones that predict academic achievement on the basis of socioeconomic factors, “regional effects”, and reported national IQ levels.[39]

Such models fare very well in accounting for global differences in academic performance. But we happen to lack confidence in their particular estimates regarding China because we have some doubts about the “cognitive” measures for China in the IQ datasets our project utilized.

Virtually all IQ datasets assign China an above-global-average level of national cognitive ability—indeed in these listings China typically ranks near the very top in the global distribution of mean IQ levels. Such soundings may perhaps track with the true results for the children of urban China who are most likely to be examined in such evaluations. But urban children comprise a minority of China’s youth. Most youngsters in China still grow up in rural areas. And China’s lofty IQ ratings in relevant international datasets simply do not square with the gathering empirical evidence about cognitive development in places for from Shanghai and Beijing.

In 2017, *Science* magazine reported on research by Stanford-based scholars indicating that half of rural China’s children fell more than a standard deviation

[38] Interestingly enough, the World Bank’s own HLO dataset derives an independent estimate of test score performance for China very close to the modeled numbers we report above (456 points for World Bank HLO, versus our own 454-468 points in Table 7). While the calculated results are quite similar, the World Bank does not provide information on the methodology it deployed to arrive at its adjustments of officially reported China scores.

[39] These may be found in Appendix F.

below the norm on IQ tests (i.e. 90 points or lower)[40]. Given rural China's share of China's total child and youth population, that finding would apply to a third or more of China's rising generation nationwide.

Note that such numbers would imply that lower cognitive development children are actually *over-represented* in the Chinese population.[41] And since that *Science* story, an increasing corpus of peer reviewed studies have appeared in scholarly journals suggesting mean IQs in various parts of rural China tested far below the notional global norm of 100 points.[42]

The true contours of knowledge and skills in the Chinese population will have a direct and powerful influence on the balance of power over the coming generation. But the cloud of empirical uncertainty surrounding China's true terrain of knowledge capital has not yet been dispelled—indeed, as yet it has barely been pierced.

Our research in this report only suggests what that terrain *might* look like. Learning what it *does* look like will require an in-depth assessment of both quantitative and qualitative information to paint a picture of Chinese realities. Such an effort could allow us to place the actual and prospective dimensions of China's knowledge capital in international—and thus strategic—perspective.

To be sure: the PRC is an increasingly closed system nowadays, and Chinese authorities have no reason to abet independent assessment of a national capability that they regard, at the end of the day, as strategic to China's future. But for all its likely difficulties, assessing the knowledge capital situation in contemporary China looks to be a different sort of research challenge from the one that confronted, say, Murray Feshbach in the Cold War era when he undertook his great work on demographic and health conditions in the USSR.

---

[40] Dennis Normile, "One in Three Chinese Children Faces an Education Apocalypse. An Ambitious Experiment Hopes to Save Them," *Science*, September 21, 2017, https://www.science.org/content/article/one-three-chinese-children-faces-education-apocalypse-ambitious-experiment-hopes-save

[41] I.e., roughly a third of the total population scoring below 90 points, versus the roughly 16 percent that the normed "bell curve" anticipates for a distribution one standard deviation away from the notional IQ mean of 100 points.

[42] To cite just one of many: Xinyue He, Huan Wang, Fang Chang, Sarah-Eve Dill, Han Liu, Bin Tang and Yaojiang Shi, "IQ, Grit, and Academic Achievement: Evidence from Rural China," *International Journal of Educational Development*, Volume 80, 2021, which found the mean IQ of tested rural China of be 89 points, https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/iq_grit_and_academic_achievement-_evidence_from_rural_china.pdf.

For one thing, current PRC decision-making relies much more than did its Soviet counterpart on processes (or at the very least the appurtenances) of policy research. The PRC Ministry of Education has a constellation of research institutes that contribute to policy evaluation and development: the Ministry's English language website lists over 30 of these[43]. For another, and also unlike the Soviet system, which extolled its own research over that taking place in the outside world, and maintained a parallel academic universe for publishing and promoting its researchers, the PRC actively seeks draw upon scholarly talent in the West, and to hold its own research to Western standard by publishing in peer-reviewed Western journals—a disposition that perforce opens a social science aperture into China which did not exist for the USSR. Finally, unlike the Soviet system—whose ultimate failure can be attributed at least in part to its manifest incapability to cope with a worldwide "information revolution", the PRC has embraced that same revolution with enthusiasm. The very state that has brought AI and big data to bear in its "social credit rating system" is hungry for all manner of other data as well, and collects them assiduously—including such things as annual scores from the national college examination tests (*gaokao*) and panel surveys on educational performance by students around the country. Just how (or how well) the Chinese government uses such data is of course another question—the point for the moment is that such data *actually exist*.

We should entertain no illusions about the obstacles that would likely be faced by any effort to map out China's actual contours of knowledge capital. But we should likewise understand the potential benefits of such an effort.

[43] "Affiliated Institutions," Ministry of Education, The People's Republic of China, http://en.moe.gov.cn/about_MOE/affiliated_institutions/.