
Supplementary information

Skill levels and gains in university STEM education in China, India, Russia and the United States

In the format provided by the
authors and unedited

Supplementary Online Materials for

Skill Levels and Gains in College STEM Education
in China, India, Russia, and the United States

Prashant Loyalka, Ou Lydia Liu, Guirong Li, Elena Kardanov, Igor Chirikov,
Shangfeng Hu, Ningning Yu, Liping Ma, Fei Guo, Tara Beteille, Namrata
Tognatta, Lin Gu, Guangming Ling, Denis Federiakin, Huan Wang, Saurabh
Khanna, Ashutosh Bhuradia, Zhaolei Shi, Yanyan Li

Correspondence to: loyalka@stanford.edu, chirikov@berkeley.edu

This PDF file includes:

Supplementary Materials and Methods
Supplementary References
Supplementary Tables 1-7

Other Supplementary Materials for this manuscript includes the following:

Materials and Methods

A. More Information about Sampling and Implementation in China, India, and Russia

We drew nationally representative samples of broadly, yet comparably, defined computer science (CS) and electrical engineering (EE) students in China, India, and Russia. The sampling procedure was comprised of several steps.

In the first step, we identified a broad set of cross-nationally comparable CS and EE majors or departments. These departments comprise approximately 34%, 24% and 24% of all undergraduate (bachelor's degree) STEM majors in China, India and Russia (authors calculations using administrative data with complete national coverage in each country). In China, the CS departments included *Computer Science and Technology*, *Computer Software*, *Software Engineering*, *Network Engineering*, *Internet of Things*, *Information Security*, *Information Engineering*, and *Informational and Computational Science*. The EE departments in China included *Electrical Engineering*, *Electrical Engineering and Automation*, *Measurement and Control Technology and Instrumentation*, *Electronic Science and Technology*, *Electronic and Information Engineering*, *Electronic and*

Information Science and Technology, Communication Engineering, Optoelectronic Information Science and Engineering, Microelectronics Science and Engineering), and Automation.

In India, the CS departments included *Computer Engineering, Computer Science Engineering, Information Science and Engineering, and Information Technology*. The EE departments in India included *Electrical Engineering, Electronics and Communication Engineering, Electronics and Electrical Engineering, Electronics and Instrumentation Engineering, and Electronics and Telecommunications Engineering*.

In Russia, the CS departments included *IT and Computer Facilities, Informational Systems and Technologies, Applied IT, Program Engineering, Mathematics and Computer Science, Fundamental IT and Information Technologies, Software and Administration of Information Systems, Information Security*. The EE departments in Russia included *Information and Communication Technology and Communication Systems, Design and Technology of Electronic Instrumentation, Radio Engineering, Electronics and Nanoelectronics, Electrical Power and Electric Engineering, Laser Equipment and Laser Technologies, Optics Engineering, Instrument Construction, Photonics and Optoinformatics*.

In the second step, we randomly selected higher education institutions (“universities” in China and Russia and “colleges” in India) from each country that had these comparable CS and EE programs. To do so, we used administrative data on (the population frame of) all higher education institutions with CS and EE programs in each country. From China, we selected institutions from six provinces that represent the country’s northern (Beijing and Shandong), central (Henan and Shaanxi), and southern (Guangdong and Sichuan) regions; the six provinces also represent economically more developed (Beijing and Guangdong ranked 2nd and 8th in GDP per capita out of 31 provincial-level administrative divisions), mid-developed (Shandong and Shaanxi ranked 10th and 14th) and less developed (Sichuan and Henan ranked 22nd and 24th) regions (National Bureau of Statistics of China 2015). From each of the six provinces, we took a simple random sample of 6 universities that offered bachelor’s degree programs in CS or EE.¹ In India, we randomly selected 50 colleges from elite and non-elite colleges that offered either CS or EE bachelor’s degree programs.² Specifically, we used simple random sampling to select elite colleges and probability proportional to size sampling to select non-elite colleges. In Russia, we took a simple random sample of 27 universities from all universities that offered bachelor’s programs in CS and EE and a simple random sample of 7 universities from all universities that

¹ The vast majority of universities in China have both types of programs (rather than only one type of program).

² Similar to China, the vast majority of colleges in India have both types of programs (rather than only one type of program).

offered bachelor's programs in CS but none in EE. Only two universities in Russia offered bachelor's programs in EE but not CS. For the sake of simplicity, the sampling frame in Russia did not include the (only) two universities that offered EE but not CS programs.³

The national samples of universities represented the range of elite and non-elite institutions in each country (Carnoy et al. 2013). In China, "elite" universities were defined on the basis of their designation as Project 985 or 211 universities. Project 985 universities are the top 39 universities in China whereas Project 211 universities are the top 112 universities in China, and both have received preferential government funding. In India, elite institutions were defined as the India Institutes of Technology (IITs), the Indian Institutes of Information Technology (IIITs), the National Institutes of Technology (NITs), and other institutions that ranked in the top 100 of the National Institutional Ranking Framework (NIRF) rankings developed by the Ministry of Human Resource Development, Government of India. In Russia, "elite" universities were defined as National Research Universities, "5-100" universities, and Federal universities. These universities receive additional funding from the Russian government.

In the third step, we randomly sampled students within CS and EE programs in the selected universities. In China, we randomly sampled up to 2 CS departments and 2 EE departments from each sample university. If there was only one CS or EE department in the university, we sampled only that department. We then randomly sampled 1 administrative group or "class" of first year (freshman year) students within each department. In India, we randomly sampled 1 CS department and 1 EE department from each college. Finally, in Russia, we randomly sampled up to 3 CS departments and 3 EE departments from each university. If there were less than three CS or EE departments in the university, we sampled all departments. We then randomly sampled 1 administrative group or class of first year students within each department.

Fourth, in each randomly sampled class (or randomly sampled department in the case of India), we sampled all first year students. Furthermore, within each year, half of the students were randomly assigned to take year-specific math and physics exams, one quarter of the students were randomly assigned to take a critical thinking exam, and one quarter of the students were randomly assigned to take a quantitative literacy exam. In this paper, we do not report the results of the quantitative literacy exam.

³ After we randomly sampled institutions, one institution in each country decided not to participate in the study. Using the same sampling procedures, we replaced these single institutions in each country with another randomly sampled institution. Subsequently, all selected institutions as well as their subsidiary administrative units participated in the study. As such, the total effect on representativeness from exclusions/replacements is negligible.

We accounted for the sampling design in all of our analyses. For all students in China, India and Russia, we created and used sample weights that reflected the inverse probability of being sampled at the university, department, (in the case of China and Russia) class, and student levels.⁴ We adjusted our standard estimates to account for the clustering of students within universities.⁵

B. Exam Creation

The math and physics exams were specifically created to assess and compare math and physics skill levels and gains among start of first and end of second year students in CS and EE programs across countries. Details of the test development and validation process is explained in more detail in Kardanova et al. (2016). The content of the start of first year math and physics exams for were aligned with common and core content that students cover in high school curricula and on high-stakes college entrance exams; the content of end of second year math and physics exams were aligned with the common and core content that students cover in the first two years of their undergraduate programs. The content validity, appropriateness, and translation of large pools of exam items were confirmed, item-by-item, with dozens of experts at elite and non-elite universities from the different countries. The larger pools of exam items were also piloted with approximately 4,000 start of first and start of third year CS and EE students in China, India, and Russia. Afterwards, the psychometric properties (item quality, reliability, unidimensionality, validity, scalability and cross-national comparability) of the exams were validated. The final math and physics exams, for freshmen and juniors separately, each contained 35 items and lasted for 40 minutes.

The *HEIghten*[®] Critical Thinking assessment was developed by ETS and launched in 2015. The test was rigorously piloted with over 3,000 students at 35 two- and four-year institutions in the United States. Validity evidence supports the use of the test (Liu et al. 2016). The assessment framework captures both analytic and synthetic dimensions of critical thinking and is aligned with most current research on critical thinking in higher education (Liu, Frankel and Roohr 2014). Multiple item types are used to assess students' critical thinking skills, including extensive critical thinking sets. Critical thinking sets provide a list of facts, sometimes supplemented by a table or a graph and two or more statements of

⁴ The sample weights were created by multiplying the inverse probabilities of being sampled at each stage (university, department, class, and student). The sample weights were created in consultation with trained statisticians. After the sample weights were created, we validated that the sample represented essential characteristics of the population of CS and EE students (in particular, total size and percentage of elite students) within an appropriate degree of error.

⁵ The standard errors of the assessment outcomes are one to five times larger under the complex sampling design compared to the simple random sampling.

opinions, which are related to one another and also to the listed facts. The assessment also includes shorter passages that address similar skills to those assessed by the critical thinking sets but in a smaller scope, and item sets that prescribe conditions applicable to a fictional situation and ask students to extrapolate conclusions based on those conditions (National Academy of Science 2005). All the items are embedded in authentic, real-life contexts that are relevant to higher education. The test has 26 items and takes about 45 minutes to complete.

The critical thinking test assesses generic, transferable skills which are not aligned with any particular curriculum and therefore there is no evidence suggesting that the critical thinking test is more sensitive to students in a particular country than to student in other countries. In addition, although critical thinking is considered an important skill, it is generally not included in any admissions test in China, India or Russia and therefore students would not be motivated to improve the skill through test prep.

Following a similar procedure that is employed by the Programme for International Student Assessment (PISA) to link scores of students who took the same assessment in different languages and countries (OECD 2014), scores on the HEIghten Critical Thinking assessments for students from China, India, Russia, and the USA were linked using a statistical procedure called concurrent calibration and using a 2-parameter item response (2-PL IRT) model.⁶ The procedure assumes that all items function the same across countries and uses a threshold (RMSEA=0.15) to evaluate whether this assumption holds for each individual item. The items that meet this criterion are referred to as anchor items, as their item parameters are considered the same across countries. The items that do not meet this criterion are considered country-specific items and their parameters function differently across countries. As a result of this iterative process, all anchor items and country-specific items were identified. The ability of students estimated based on the anchor items and country-specific items was made comparable through concurrent calibration. These ability estimates were

⁶ An almost identical procedure was used to link math (and physics) scores of students (and thus place the scores on a common scale across years of study and countries). After doing so, we also made a slight adjustment to the scores of baseline students in China that had to take the test on paper. Adjustments were based on the results of a large-scale trial that randomly assigned computer versus paper versions of the critical thinking, math, and physics test among CS and EE students in China. Cross-country comparisons of skill levels are substantively the same with and without the adjustment. Skill gain estimates in China are, in fact, slightly higher (that is closer to zero or less negative) as a result of the adjustment. As an additional sanity check, we also examined differences in scaled (critical thinking, math, and physics) scores across the first and third year student cohorts from the baseline wave in each country (including China). Estimating skill gains by comparing differences in skill levels across cohorts in different years is common practice in the higher education assessment literature (Saavedra and Saavedra 2011; Huber and Kuncel 2016). The cross-cohort differences generally align with the skill gain estimates from the longitudinal data and do not change the substantive conclusions of the paper.

then put on a common scale (integers between 30 and 50 points) following a widely used true score to scale score linking procedure (Kolen and Brennan 2004). The software program MLTDM was used to conduct the concurrent calibration (von Davier 2005).

While it might be true that freshmen in China benefited from extensive math and physics test prep in high school, high school students in India and Russia also experienced intense math and physics test prep. In contrast to students in China, students in India and Russia did not show losses in math skills between the start of the first year and end of the second year, however. This suggests that China's loss in math skills is related to college experiences rather than the impact of test prep in high school.

C. Exam Translation and Translation Review

The math and physics exams underwent a rigorous translation, adaptation, and language review process. For China and Russia, the math and physics exams were translated, back translated, and reviewed by multiple experts on our research team that were native or fluent in English and various country-specific languages and which also had ample exposure to the subject matter areas. Since English is the language of instruction in STEM majors in India, only the language of test items was reviewed and modified by our team of experts. The translated language of each exam item was further checked by 12 local experts (professors that teach math and physics in CS and EE departments – 6 from elite institutions and 6 from non-elite institutions) in each of the three countries. Outside expert suggestions were carefully reviewed and incorporated by our internal team.

The translation and adaption of the *HEIghten*[®] CT assessment into Chinese and Russian was conducted by Capstan, a linguistic quality control company based in Belgium that has undertaken translation work for large-scale programs such as the Programme for International Student Assessment (PISA) and Programme for International Assessment of Adult Competencies (PIAAC). We adopted the most rigorous, three-step translation and adaptation model that Capstan offers with double translation, reconciliation, and verification. Our internal team of experts also worked with Capstan to ensure that translations were of the highest quality.

D. Motivation to Participate in the Assessment

We address test motivation in two ways. We first ran a pilot randomized experiment with three different motivational prompts: (a) offering students a results report; (b) telling students the test was competitive with other countries/institutions; (c) telling students the purpose/content of the test) and a control group. Since only the first had a positive impact on scores (~ 0.10 SD), all students in the main study were offered the option of receiving a score report before they took the exams. Second, according to standard practice, we omitted a

small minority of students in China, India, and Russia that answered less than 75% of items. The results are substantively the same whether or not we include these students.

E. Gains Estimates – Adjustments for Missing Data

In addition to reporting skill gains estimates for China, India, and Russia that are unadjusted for missing test score values (and rely on the “missing completely at random” assumption), we use multiple imputation to report estimates that are adjusted for missing data (and rely on the more conservative “missing at random” assumption).⁷ In particular, we impute missing data in two ways. First, we impute missing follow-up test score values for students that participated in the baseline but not in the follow-up; using the imputed data, we estimate skill gains for all students that were present in the baseline (see Supplementary Table 3.A and Supplementary Table 3.B, “A: Adjusted Gains”). Second, we additionally impute missing baseline test score values for students who participated in the follow-up but not in the baseline; we then estimate skill gains for all students (that were present in either the baseline or the endline) (see Supplementary Table 3.A and Supplementary Table 3.B, “B: Adjusted Gains”). In either case, estimates that rely on imputed data are very close to (and substantively the same as) the non-adjusted estimates.

F. Skills/Gains Estimates – Controlling for Family Background and Out-of-College Activities

The observed differences in skill levels and gains across countries, and between elite and non-elite institutions can be attributed to higher education systems and institutions and not to differences in family background and out-of-colleges activities (e.g., outside tutoring, internships, paid job, volunteering) among countries and institutional types. Specifically, in estimating differences in skill levels and gains between countries and between elite and non-elite institutions, we control for family background and out-of-college activities. We control for family background since our baseline surveys in China, India, and Russia asked start of year 1 students about whether their fathers and mothers went to college as well as whether they had certain assets in the home (fridge, microwave, laptop/tablet, AC, car, washing machine). We use the asset data, along with polychoric principal components analysis, to construct a family wealth

⁷ We use the built-in `–mi-` command suite in Stata 15.1 (20 imputations). The following variables were used in the multiple imputation procedure: subject-specific test score (baseline or endline), baseline department mean of subject-specific test score, age, gender, father went to college (yes/no), mother went to college (yes/no).

index—see Kolenikov and Angeles, 2009). In our follow up surveys of end of year 2 students, we were also able to collect data on out-of-college activities. For end of year 2 students that took the critical thinking test in China, India, and Russia, we collected data on and are able to control for a wide array of out-of-college activities (hours of tutoring, hours of paid work, entrepreneurial activities (yes/no), community service or volunteer work (yes/no), and internship (yes/no)). For end of year 2 students that took the math and physics tests in China and Russia (but not India), we collected data on and are able to control for two main out-of-college activities (hours of tutoring, hours of paid work).

According to Supplementary Table 4, cross-country differences in critical thinking skill levels are largely the same when controlling for family background and out-of-college activity.

- China continues to have much higher levels of critical thinking skills at the start of year 1 and the end of year 2 than India (without controls, the differences were approximately 1.4 SDs; with controls, the differences are nearly the same at 1.2 to 1.3 SDs).
- China also continues to have higher levels of critical thinking skills at the start of year 1 and the end of year 2 than Russia (without controls, the differences were approximately 0.35 SDs; with controls the differences are moderately higher from 0.55 SDs to 0.73 SDs).
- Russia continues to have higher levels of critical thinking skills than India (without controls, the differences were approximately 1 to 1.05 SDs; with controls the differences are moderately lower from 0.56 SDs to 0.73 SDs).

These changing score gaps between Russia and the other two countries are predictable, since students in Russia have higher levels of family wealth and are much more likely to have parents that are college educated (and these types of socioeconomic factors are almost always positively correlated with critical thinking test scores).

We also find that cross-country differences in critical thinking skill gains are all negligible and statistically not significant whether or not we control for family background and out-of-college activities (Supplementary Table 4).

We find almost exactly the same pattern for math and physics skill levels and gains (see Supplementary Table 5 and Supplementary Table 6). The magnitudes of the large differences in skill levels between China and India are very similar whether or not we control for family background and out-of-college activities. When adjusting for family background, Russia falls behind India in math and physics skill levels (Supplementary Table 5). The large gap between Russia and China widens even further when adjusting for family background and out-of-college activities (Supplementary Table 5 and Supplementary Table 6). Similar to critical thinking, controlling for the family background and out-of-college activity

does not substantively change the cross-country differences in math and physics skill gains.

Furthermore, controlling for the family background and out-of-college activity does not substantively change within country differences between elite and non-elite universities (Supplementary Table 7). The magnitudes of differences in critical thinking, math and physics skills levels and gains between elite and non-elite institutions in China and India are very similar whether or not we control for family background, out-of-college activities or both. In Russia, the small gap between elite and non-elite institutions in critical thinking skills at the start of year 1 closes when we control for both family background and out-of-college activities (from 0.3 SDs with no controls to 0.1 SDs with controls) but the magnitudes of differences in other skills levels and gains are very similar.

Finally, we do not control for rural-urban differences in our analyses for two reasons. First, the vast majority of colleges in each country are in urban areas. Not one of the colleges in our sample is in a rural area. There are thus no rural-urban differences in regards to institutional location or setting. Second, the definition of rural versus urban at the individual level is not consistent within or across countries. In China, individual designations of rural versus urban are typically defined by one's official residential permit status; in recent years, different regions of China have instituted policies which, to varying degrees, allow individuals with rural permits to switch to urban permits. In India and Russia, an individual's rural versus urban status is not by official designation but rather self-reported and subject to individual interpretation. In fact, given inconsistencies in reporting in Russia, we did not ask study participants to self-report their rural versus urban status. That all being said, although we did not control for individual designations of rural versus urban in our analyses, the wealth index we control for, which is consistently measured within and across countries, is a reasonable proxy for socioeconomic differences across rural and urban areas.

References

1. National Bureau of Statistics of China. *China statistical yearbook* (2015; <http://www.stats.gov.cn/tjsj/ndsj/2015/indexeh.htm>).
2. Huber, C.R. & Kuncel, N. R. Does college teach critical thinking? A meta-analysis. *Review of Educational Research* **86**(2), 431-468 (2016).

3. Liu, O. L., Mao, L., Frankel, L. & Xu, J. Assessing critical thinking in higher education: the HEIghten™ approach and preliminary validity evidence. *Assessment and Evaluation in Higher Education* **41**(5), 677-694 (2016).
4. Organization for Economic Cooperation and Development. *The PISA 2012 technical report* (OECD, 2014; <https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf>).
5. Kolen, M. J. & Brennan, R. L. *Test Equating, Scaling, and Linking: Methods and Practice* (Springer, New York, 2004).
6. Davier, M. *Multidimensional latent trait modelling (MDLTM)* [Software program] (NJ: Educational Testing Service, Princeton, 2006).
7. Kolenikov, S., & Angeles, G. Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer?. *Review of Income and Wealth*, **55**(1), 128-165 (2009).

Supplementary Table 1. Mean Characteristics of First-Year Undergraduate Computer Science and Electrical Engineering Students across Countries and Types of Institutions

	China			India			Russia		
	All	Elite	Non-elite	All	Elite	Non-elite	All	Elite	Non-elite
Age (years)	19.092 [0.056]	18.740 [0.044]	19.202 [0.052]	18.171 [0.065]	18.504 [0.093]	18.156 [0.067]	18.559 [0.055]	18.358 [0.063]	18.623 [0.060]
Female (%)	0.289 [0.022]	0.297 [0.050]	0.286 [0.025]	0.388 [0.033]	0.142 [0.025]	0.399 [0.034]	0.206 [0.014]	0.168 [0.014]	0.218 [0.017]
Dad college degree (%)	0.246 [0.040]	0.505 [0.069]	0.163 [0.018]	0.423 [0.033]	0.635 [0.027]	0.413 [0.034]	0.832 [0.012]	0.885 [0.015]	0.814 [0.013]
Mom college degree (%)	0.204 [0.042]	0.471 [0.081]	0.117 [0.016]	0.342 [0.032]	0.564 [0.049]	0.332 [0.033]	0.892 [0.010]	0.925 [0.006]	0.881 [0.012]
Took entrance exam (%)	0.998 [0.001]	0.995 [0.004]	0.999 [0.000]	0.954 [0.009]	0.989 [0.004]	0.952 [0.000]	0.846 [0.033]	0.951 [0.015]	0.813 [0.036]

Notes: (1) Means in first row with standard errors in second row in brackets; (2) means adjusted for survey design (weights, clustering, strata); (3) Definitions of elite: China - all 985 and 211 institutions; India – IITs, NITs, and other top 100 MHRD ranked colleges; Russia - all national research and federal universities; (4) N = 16,054.

Supplementary Table 2. Time Spent Studying (hours per day)

	China			India			Russia		
	All	Elite	Non- elite	All	Elite	Non- elite	All	Elite	Non- elite
Time on college studies	7.89	8.45	7.71	9.83	7.64	9.94	8.77	9.01	8.68
Time on outside tutoring or mentoring	0.42	0.23	0.48	0.73	0.54	0.74	0.09	0.05	0.10
Total study time	8.31	8.68	8.19	10.57	8.18	10.68	8.86	9.06	8.78
% of total study time on college studies	95.0%	97.4%	94.1%	93.1%	93.4%	93.1%	99.0%	99.5%	98.9%

Notes: (1) Time on college studies refers to average number of hours per day in class, plus average number of hours doing schoolwork directly related to class. Total study time refers to time on college studies plus time on outside tutoring. (2) China, India, and Russia data are national random samples of first-year students from four-year undergraduate CS and EE related majors, using the sample of students that took the math test (and the associated questionnaire with time use questions): $N_{\text{China}} = 1,895$, $N_{\text{China, elite}} = 347$; $N_{\text{India}} = 3,821$, $N_{\text{India, elite}} = 610$; $N_{\text{Russia}} = 925$, $N_{\text{Russia, elite}} = 204$. Analytical estimates from China, India, and Russia are calculated using sampling weights so that they are representative of well-defined national populations. (3) Definitions of elite: China - all 985 and 211 institutions; India – IITs, NITs, and other top 100 MHRD ranked colleges; Russia - all national research and federal universities.

Supplementary Table 3.A. Adjusted and Unadjusted Skill Gains: Across China, India, and Russia (Standard Deviation Units)

	China	India	Russia
PANEL A: CRITICAL THINKING			
Unadjusted Y1 to Y2 Gains	0.104 (0.088)	0.038** (0.019)	0.102* (0.056)
A: Adjusted Y1 to Y2 Gains	0.081 (0.075)	0.064*** (0.021)	0.077 (0.070)
B: Adjusted Y1 to Y2 Gains	0.094 (0.074)	0.017 (0.020)	0.080 (0.068)
Unadjusted Y3 to Y4 Gains	-0.476*** (0.092)	-0.112*** (0.028)	-0.107** (0.049)
A: Adjusted Y3 to Y4 Gains	-0.428*** (0.101)	-0.105*** (0.027)	-0.119* (0.059)
B: Adjusted Y3 to Y4 Gains	-0.427*** (0.091)	-0.121*** (0.026)	-0.125** (0.057)
PANEL B: MATH			
Unadjusted Y1 to Y2 Gains	-0.312*** (0.043)	0.388*** (0.015)	0.079* (0.042)
A: Adjusted Y1 to Y2 Gains	-0.279*** (0.053)	0.403*** (0.017)	0.114** (0.049)
B: Adjusted Y1 to Y2 Gains	-0.246*** (0.063)	0.343*** (0.018)	0.098** (0.048)
Panel C: PHYSICS			
Unadjusted Y1 to Y2 Gains	-0.061 (0.041)	0.244*** (0.037)	0.168*** (0.070)
A: Adjusted Y1 to Y2 Gains	-0.042 (0.056)	0.257*** (0.032)	0.261*** (0.049)
B: Adjusted Y1 to Y2 Gains	-0.003 (0.062)	0.240*** (0.039)	0.245*** (0.050)

Notes: (1) For critical thinking, students took exams in the first semester of their freshman/junior year and then again at the end of the second semester of their sophomore/senior year. For math and physics, students took exams in the first semester of their freshman year and then again at the end of the second semester of their sophomore year. Unadjusted gains are calculated based on students present in both baseline and follow up phases. Adjusted gains estimates “A” use all students in the baseline (imputing scores if they were not present at follow up). Adjusted gains estimates “B” use all students in the baseline and follow up (imputing scores if they were not in the baseline or follow up). (2) Levels and gains estimates are reported as effect sizes (in standard deviation units). Scaled exam scores were divided by the subject-specific baseline mean and standard deviation of the China, India, Russia cross-national sample of exam takers. (3) China, India, and Russia data are national random samples of students from four-year undergraduate computer science and electrical engineering related majors. Analytical estimates from China, India, and Russia are calculated using sampling weights so that they are representative of well-defined national populations. (4) Definitions of elite: China - all 985 and 211 institutions; India – IITs, NITs, and other top 100 MHRD ranked colleges; Russia - all national research and federal universities. (5) Standard errors reported in parentheses are adjusted for clustering at the institution level. (6) *p < .10, **p < .05, ***p < .01.

Supplementary Table 3.B. Adjusted and Unadjusted Skill Gains: Elite versus Non-elite Institutions (Standard Deviation Units)

	China			India			Russia		
	Elite	Non-Elite	Difference	Elite	Non-Elite	Difference	Elite	Non-Elite	Difference
PANEL A: CRITICAL THINKING									
Unadjusted Y1 to Y2 Gains	0.417 (0.267)	0.008 (0.062)	0.409 (0.259)	-0.014 (0.075)	0.041** (0.019)	-0.057 (0.074)	0.054 (0.201)	0.119*** (0.035)	-0.065 (0.189)
A: Adjusted Y1 to Y2 Gains	0.337 (0.203)	0.006 (0.066)		0.019 (0.073)	0.065*** (0.021)		-0.074 (0.185)	0.135** (0.050)	
B: Adjusted Y1 to Y2 Gains	0.301 (0.181)	0.015 (0.069)		0.033 (0.061)	0.023 (0.019)		-0.024 (0.181)	0.117** (0.050)	
Unadjusted Y3 to Y4 Gains	-0.397 (0.249)	-0.505*** (0.085)	0.108 (0.246)	-0.343** (0.114)	-0.102*** (0.028)	-0.241** (0.111)	-0.024 (0.090)	-0.126** (0.056)	0.102 (0.100)
A: Adjusted Y3 to Y4 Gains	-0.239 (0.222)	-0.452*** (0.068)		-0.350** (0.114)	-0.098*** (0.027)		0.056 (0.132)	-0.159** (0.059)	
B: Adjusted Y3 to Y4 Gains	-0.264 (0.199)	-0.431*** (0.073)		-0.364** (0.114)	-0.112*** (0.027)		0.036 (0.138)	-0.160*** (0.056)	
PANEL B: MATH									
Unadjusted Y1 to Y2 Gains	-0.209* (0.091)	-0.333*** (0.050)	0.125 (0.099)	0.306*** (0.040)	0.387*** (0.016)	-0.081* (0.041)	0.257*** (0.031)	0.043 (0.046)	0.213*** (0.053)
A: Adjusted Y1 to Y2 Gains	-0.140 (0.087)	-0.300*** (0.052)		0.316*** (0.043)	0.400*** (0.017)		0.247*** (0.043)	0.094 (0.055)	
B: Adjusted Y1 to Y2 Gains	-0.083 (0.118)	-0.300*** (0.051)		0.315*** (0.053)	0.345*** (0.017)		0.216*** (0.035)	0.074 (0.054)	
Panel C: PHYSICS									
Unadjusted Y1 to Y2 Gains	0.044 (0.116)	-0.076* (0.042)	0.120 (0.117)	-0.403*** (0.104)	0.272*** (0.034)	-0.675*** (0.105)	-0.168 (0.123)	0.244*** (0.056)	-0.413*** (0.127)
A: Adjusted Y1 to Y2 Gains	0.077 (0.093)	-0.047 (0.051)		-0.391*** (0.093)	0.282*** (0.030)		0.064 (0.044)	0.325*** (0.045)	
B: Adjusted Y1 to Y2 Gains	0.152 (0.094)	-0.049 (0.053)		-0.411*** (0.094)	0.272*** (0.035)		0.065 (0.049)	0.305*** (0.046)	

Notes: (1) For critical thinking, students took exams in the first semester of their freshman/junior year and then again at the end of the second semester of their sophomore/senior year. For math and physics, students took exams in the first semester of their freshman year and then again at the end of the second semester of their sophomore year. Unadjusted gains are calculated based on students present in both baseline and follow up phases. Adjusted gains estimates “A” use all students in the baseline (imputing scores if they were not present at follow up). Adjusted gains estimates “B” use all students in the baseline and follow up (imputing scores if they were not in the baseline or follow-up). (2) Levels and gains estimates are reported as effect sizes (in standard deviation units). Scaled exam scores were divided by the subject-specific baseline mean and standard deviation of the China, India, Russia cross-national sample of exam takers. (3) China, India, and Russia data are national random samples of students from four-year undergraduate computer science and electrical engineering related majors. Analytical estimates from China, India, and Russia are calculated using sampling weights so that they are representative of well-defined national populations. (4) Definitions of elite: China - all 985 and 211 institutions; India – IITs, NITs, and other top 100 MHRD ranked colleges; Russia - all national research and federal universities. (5) Standard errors reported in parentheses are adjusted for clustering at the institution level. (6) *p < .10, **p < .05, ***p < .01.

Supplementary Table 4. By Country Differences in Critical Thinking Skill Levels and Gains, Adjusted for Family Background and Out-of-College Activity

	(1)	(2)	(3)	(4)	(5)	(6)
	Levels				Gains	
	Start of Year 1	End of Year 2	Start of Year 1	End of Year 2		
India - China	-1.286*** (0.093)	-1.327*** (0.153)	-1.196*** (0.078)	-1.286*** (0.131)	-0.036 (0.077)	-0.087 (0.088)
Russia – China	-0.603*** (0.142)	-0.597*** (0.200)	-0.642*** (0.104)	-0.725*** (0.191)	0.009 (0.102)	-0.007 (0.106)
Russia – India	0.683*** (0.103)	0.730*** (0.110)	0.555*** (0.087)	0.561*** (0.121)	0.045 (0.073)	0.081 (0.064)
Family Background Controls	YES	YES	YES	YES	YES	YES
Out-of-college Activity Controls	NO	NO	YES	YES	NO	YES

Notes: (1) Family background controls include mother went to college (yes/no), father went to college (yes/no), and wealth index created using polychoricpca in Stata and yes/no possession of six household assets (fridge, microwave, laptop/tablet, AC, car, washing machine)—see Kolenikov and Angeles (2009). (2) Out-of-college activities include hours of tutoring, hours of paid work, entrepreneurial activities (yes/no), community service or volunteer work (yes/no), and internship (yes/no).

Supplementary Table 5. By Country Differences in Math and Physics Skill Levels and Gains, Adjusted for Family Background

	(1)	(2)	(3)	(4)	(5)	(6)
	Levels				Gains	
	Math - Start of Year 1	Math - End of Year 2	Physics - Start of Year 1	Physics - End of Year 2	Math Gains (End of Year 2 - Start of Year 1)	Physics Gains (End of Year 2 - Start of Year 1)
India - China	-1.206*** (0.142)	-0.557*** (0.159)	-1.014*** (0.187)	-0.771*** (0.188)	0.674*** (0.053)	0.289*** (0.058)
Russia – China	-1.147*** (0.178)	-0.776*** (0.194)	-1.451*** (0.249)	-1.020*** (0.243)	0.375*** (0.073)	0.334*** (0.060)
Russia – India	0.059 (0.081)	-0.220** (0.094)	-0.437*** (0.116)	-0.248*** (0.088)	-0.299*** (0.055)	0.044 (0.043)
Family Background Controls	YES	YES	YES	YES	YES	YES

Notes: (1) Family background controls include mother went to college (yes/no), father went to college (yes/no), and wealth index created using polychoricpca in Stata and yes/no possession of six household assets (fridge, microwave, laptop/tablet, AC, car, washing machine)—see Kolenikov and Angeles (2009).

Supplementary Table 6. By Country Differences in Math and Physics Skill Levels and Gains, Adjusted for Family Background and Out-of-college Activity (China and Russia only)

	(1)	(2)	(3)	(4)	(5)	(6)
	Levels				Gains	
	Math - Start of Year 1	Math - End of Year 2	Physics - Start of Year 1	Physics - End of Year 2	Math Gains (End of Year 2 - Start of Year 1)	Physics Gains (End of Year 2 - Start of Year 1)
Russia - China	-1.404*** (0.280)	-1.119*** (0.289)	-1.714*** (0.430)	-1.518*** (0.422)	0.376*** (0.062)	0.194** (0.081)
Family Background Controls	YES	YES	YES	YES	YES	YES
Out-of-college Activity Controls	YES	YES	YES	YES	YES	YES

Notes: (1) Family background controls include mother went to college (yes/no), father went to college (yes/no), and wealth index created using polychoricpca in Stata and yes/no possession of six household assets (fridge, microwave, laptop/tablet, AC, car, washing machine)—see Kolenikov and Angeles (2009). (2) Out-of-college activities include hours of tutoring, hours of paid work.

Supplementary Table 7. Within Country Differences between Elite and Non-Elite Universities, Adjusted for Family Background and Out-of-college Activity

	China (elite - non-elite)			India (elite - non-elite)			Russia (elite - non-elite)		
	With no controls	With family background controls	With family background & out-of-college activity controls	With no controls	With family background controls	With family background & out-of-college activity controls	With no controls	With family background controls	With family background & out-of-college activity controls
PANEL A: CRITICAL THINKING									
Start of Year 1	0.871*** (0.171)	0.777*** (0.126)	0.658*** (0.093)	0.971*** (0.095)	0.831*** (0.085)	0.826*** (0.091)	0.289* (0.154)	0.333** (0.158)	0.144 (0.105)
End of Year 2	1.170*** (0.319)	1.094*** (0.307)	1.021*** (0.294)	0.980*** (0.067)	0.800*** (0.050)	0.754*** (0.050)	0.202 (0.256)	0.078 (0.257)	0.085 (0.261)
Y1 to Y2 Gains	0.409 (0.259)	0.330 (0.224)	0.391 (0.265)	-0.057 (0.074)	-0.032 (0.071)	-0.058 (0.074)	-0.065 (0.189)	-0.252 (0.172)	-0.054 (0.206)
PANEL B: MATH									
Start of Year 1	1.209*** (0.262)	1.128*** (0.238)	1.180*** (0.208)	1.217*** (0.077)	1.104*** (0.076)	n/a	0.135 (0.128)	0.115 (0.138)	-0.017 (0.114)
End of Year 2	1.425*** (0.335)	1.268*** (0.276)	1.252*** (0.261)	1.187*** (0.086)	1.073*** (0.083)	n/a	0.276** (0.132)	0.275** (0.139)	0.253* (0.130)
Y1 to Y2 Gains	0.125 (0.099)	0.135 (0.091)	0.090 (0.092)	-0.081* (0.041)	-0.069 (0.043)	n/a	0.213*** (0.053)	0.154* (0.082)	0.273*** (0.077)
PANEL C: PHYSICS									
Start of Year 1	1.433*** (0.431)	1.313*** (0.397)	1.255*** (0.338)	1.720*** (0.121)	1.552*** (0.116)	n/a	0.259* (0.131)	0.100 (0.191)	0.328*** (0.119)
End of Year 2	1.634*** (0.462)	1.477*** (0.402)	1.380*** (0.341)	1.037*** (0.068)	0.971*** (0.059)	n/a	0.019 (0.127)	-0.062 (0.149)	-0.068 (0.162)
Y1 to Y2 Gains	0.120 (0.117)	0.102 (0.099)	0.079 (0.111)	-0.675*** (0.105)	-0.649*** (0.091)	n/a	-0.413*** (0.127)	-0.172*** (0.056)	-0.409*** (0.138)

Notes: (1) Family background controls include mother went to college (yes/no), father went to college (yes/no), and wealth index created using polychoricpca in Stata and yes/no possession of six household assets (fridge, microwave, laptop/tablet, AC, car, washing machine)—see Kolenikov and Angeles (2009). (2) Out-of-college activities include hours of tutoring, hours of paid work.