EDUCATIONAL QUALITY AND DISPARITIES IN INCOME AND GROWTH ACROSS COUNTRIES

Hanol Lee[a] *and Jong-Wha Lee*[b]

December 2023

Abstract

We construct a comprehensive database of educational quality by cohort for 92 countries from 1970 to 2015 and analyze its impact on disparities in income and growth worldwide. To estimate educational quality, we utilize secondary students' scores on international mathematics and science tests. Additionally, we impute unobserved test scores for individual countries in non-participating survey years. Wage regressions using individual earnings data reveal considerable returns to educational quality. We estimate human capital stock by incorporating differences in educational quantity and quality by age group across countries and over time. Our newly-constructed human capital dataset enabled us to explore the role of educational quality and human capital in understanding cross-country income disparities. The findings from development and growth accounting exercises indicated a discernible contribution of educational quality to per capita income and its growth rate.

JEL classification*:* C8, I25, J24, O15
Keywords: educational quality, human capital, test score, development accounting, return to education

Acknowledgement: We sincerely thank Robert Barro, Wei Chi, So Kubota, Jong-Suk Han, Eric Hanushek, Do Won Kwak, Kwanho Shin, and Eunbi Song for their comments and discussions, with special gratitude to the editor and associate editor for their constructive comments and invaluable suggestions. This research was supported by a grant from Korea University (K2110021).

Declaration
Conflict of interest: The authors declare that they have no conflict of interest.

[a] Research Institute of Economics and Management, Southwestern University of Finance and Economics, 555, Liutai Avenue, Wenjiang District, Chengdu, Sichuan, China 611130. E-mail: hanollee@swufe.edu.cn. https://orcid.org/0000-0003-2292-0919

[b] *Correspondence*: Economics Department, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, Republic of Korea; Tel.: +82-2-32902216. E-mail: jongwha@korea.ac.kr. https://orcid.org/0000-0002-3152-4961

I. Introduction

The quality and quantity of education are important determinants of human capital that contribute to individual productivity, earnings, and national gross domestic product (GDP) per worker and wealth.[1] Many studies have attempted to measure the differences in educational quality across countries using various indicators of school outcomes. Generally, these studies compare educational quality across countries using internationally-comparable test scores for primary and secondary students (Hanushek & Kimko, 2000; Lee & Barro, 2001; Altinok et al., 2018; Angrist et al., 2021).

This study aims to construct a measure of educational quality and estimate human capital stock by incorporating differences in educational quantity and quality for age group cohorts across countries and over time. Using our new database, we explore the role of educational quality and human capital in explaining variations in income across countries.

We compile the available international mathematics and science test scores for secondary school students and create a comprehensive database of internationally-comparable test scores from 1970 to 2015 for 92 countries. To measure the educational quality of all cohorts in these countries, we employ imputation methods, such as interpolation/extrapolation and machine learning techniques, and fill in missing test scores for the years when individual countries did not participate.

We then utilize our new dataset on educational quality to create a new measure of human capital stock, which considers variations in both quantity and quality of education among different age groups across countries and over time. The construction of human capital stock necessitates an assessment of return to educational quality, in addition to that of educational quantity. One of the key contributions of this study is to directly estimate the return to educational quality using micro-level data on US immigrants from diverse countries. We leverage our newly constructed databases to explore the role of educational quality and human capital in explaining income disparities among countries using development and growth accounting methodologies.

[1] Human capital is determined by schooling, labor market experience, and innate abilities. It has many complex human attributes that are difficult to quantify. This study measures one component of human capital, educational attainment that can be compared across many countries. Educational quantity is often measured as the estimated average years of schooling for the adult population (Barro & Lee, 2013; Lee & Lee, 2016).

This study progresses in four steps—from constructing a new dataset to implementing an empirical strategy—and makes several contributions in each step. The first step is constructing a new educational quality database. Since Hanushek and Kimko's (2000) seminal work, many researchers have constructed test-score datasets by combining available information from international assessments. They compiled the international mathematics and science test scores for 39 countries from 1965 to 1991. Lee and Barro (2001) and Hanushek and Woessmann (2012) presented updated international test score datasets for more countries, including the Third International Mathematics and Science Study (TIMSS) and Program for International Student Assessment (PISA). Many recent studies have attempted to improve data coverage by combining international and regional achievement tests (Altinok & Murseli, 2007; Altinok et al., 2014; Kaarsen, 2014; Barro & Lee, 2015; Altinok et al., 2018; Angrist et al., 2021). However, considering that international and regional achievement tests are not directly comparable, combining them can cause measurement error.[2] Some studies, including Hanushek and Woessmann, Altinok et al., and Angrist et al., combined primary and secondary students' test scores (e.g., using a simple average) to construct an aggregate overall academic achievement measure. Some studies include reading test scores in their datasets.[3] However, although the test scores for primary and secondary students for reading, and other subjects are closely correlated, they have considerable discrepancies (Lee & Barro, 2001; Barro & Lee, 2015).

This study improves the cross-country comparability of educational quality relative to previous studies by focusing on nations that participated in international mathematics and science achievement tests for secondary school students after 1970, as these data are more comparable across countries and over time. Our original dataset covers 92 countries with international test

[2] These studies rescaled the test scores for regional assessments using data for a common sample of countries that participated in international and regional assessments. However, the information derived from this limited sample may not be appropriate for rescaling test scores for African and Latin American countries that participated only in regional assessments. For example, Botswana is the only African country with international and regional assessment scores at the secondary level. Botswana participated in the Southern and Eastern Africa Consortium for Monitoring Educational Quality and TIMSS in 2007.

[3] Reading test scores are available from PISA and the International Association for the Evaluation of Educational Achievement's (IEA) international tests starting in the 1980s. Angrist et al. (2021) use Early Grade Reading Assessment (EGRA) scores. However, EGRA, which measures the most basic foundational skills for literacy acquisition in early grades in about 15 minutes, may not be fully comparable with other international assessments, which are subject to more careful participation choices, testing regimes, language considerations, and score scaling.

score observations from 1970 to 2015 (i.e., for the age cohorts that were secondary school students and took tests each year during that period).

Although our original test score data are dispersed according to the availability of international assessments, we construct a time series of observations, mostly at four-year intervals, for all 92 countries from 1970 to 2015. We utilize two imputation methods to address unobserved test scores in nonparticipating years for each of these countries. First, we employ interpolations and extrapolations, a method commonly used in previous studies such as Hanushek and Woessmann (2012) and Altinok et al. (2018). This approach helps fill in missing observations. Additionally, we adopt a machine-learning technique to impute missing data. By utilizing a diverse set of predictors, the machine-learning technique provides estimates of unobserved test scores. It can fit complex functional forms with numerous features to data and search for functions that yield accurate out-of-sample predictions (Mullainathan & Spiess, 2017; Athey & Imbens, 2019).

The second step is to estimate the returns to educational quality and quantity in the labor market by combining our new cohort average educational quality data with individual earnings data. Our approach includes measuring how the wage of an immigrant in the US labor market responds to the quality of education that the cohort of the immigrant attained at an appropriate age in the country where the immigrant migrated, using a dataset of US immigrants from a broad number of countries across time. A sizable body of literature on the use of standard Mincer-type wage equations (Mincer, 1974) finds a positive association between educational quality and income after controlling for years of schooling in various countries.[4] For example, Mulligan (1999) finds that a one standard deviation increase in test scores leads to 11% higher annual earnings in the US labor market. Some studies have attempted to assess whether the educational quality attained by immigrants to the US in their original countries drives differences in their individual earnings (Bratsberg & Terrell, 2002; Hanushek & Kimko, 2000; Hendricks, 2002; Hanushek & Woessmann, 2012; Schoellman, 2012; Li & Sweetman, 2014). This study expands on this research by adding our educational quality measure to the wage equation using microlevel wage data from the American Community Survey on US Immigrants.[5] We use the Mincer-type equation to assess

[4] See the literature surveys in Speakman and Welch (2006) and Hanushek and Woessmann (2008).

[5] Several other datasets contain individual wage information. The International Income Distribution Database used by Islam et al. (2019) and Jedwab et al. (2022) covers more than 1,000 surveys for 145 countries from 1990 to 2016. However, this dataset is not publicly available. Lagakos et al. (2018b) use

whether the quality of education, measured by internationally-comparable test scores in our new dataset, significantly influences US immigrants' wages after controlling for years of schooling and other determinants. Our approach directly estimates the return to educational quality, whereas previous studies estimate country-specific returns to educational quantity and consider them as educational quality measures.

Our estimation of the wage equation, based on the sample of actual test scores without imputed values, demonstrates considerable returns to educational quality. A one-standard-deviation increase in an international test score is associated with an average hourly-wage increase of 9.5% for US immigrants. These returns exceed the estimated return to an additional year of schooling. US data are ideal for this analysis because they include a large sample of immigrants from many countries who work under similar labor market conditions and institutions. They provide other control variables such as year of immigration, region of the workplace, and English language proficiency (Schoellman, 2012; Lagakos et al., 2018a). However, bias in the estimates of returns to educational quality may arise from omitted variables, measurement errors, and sample selection owing to immigration selectivity. We demonstrate that our estimates are robust with different specifications and samples.[6] Nevertheless, in estimating the equation for individual wages, we have only the origin-country cohort-average test score, which we use as an imperfect measure of an individual's educational quality. This could lead to a downward bias in the estimated returns to educational quality. We investigate this issue and assess the extent of bias.

The third step is to construct a new measure of human capital stock that incorporates the quality of education using our educational quality database and estimated rates of return. We call this measure "quality-adjusted human capital" and the conventional human capital stock measure based on years of schooling "quantity-based human capital." Our new dataset allows us to estimate quality-adjusted and quantity-based human capital for people aged 15–64 in 83 countries at five-

---

Integrated Public Use Microdata Series (IPUMS) data to estimate the wage equation. Although these data are publicly available, they do not include recent surveys. We need recent data on workers' wages by age (cohort) that we can match to the estimated qualities of their secondary education.

[6] We also use cross-country data from the Programme for the International Assessment of Adult Competencies (PIAAC) survey as an alternative sample. This dataset provides internationally-comparable data but covers only Organisation for Economic Cooperation and Development (OECD) member-countries. Our estimation of the wage equation shows significant returns to educational quality, although lower than those from the US immigrant sample, for this PIAAC sample. This result is available from the authors upon request.

year intervals from 1985 to 2015. Many studies have constructed aggregate human capital measures by considering differences in the quality of education based on school inputs, such as teachers' human capital, educational spending, and pupil–teacher ratios (Bils & Klenow, 2000; Caselli, 2005; Erosa et al., 2010; Manuelli & Seshadri, 2014). Recent studies have constructed educational quality measures based on international test score data and combined them with the quantity of education to build aggregate human capital measures (Kaarsen, 2014; Filmer et al., 2020; Angrist et al., 2021). Our approach complements these studies by constructing aggregate human capital measures for many countries and years based on internationally-comparable educational quality measures. Contrary to previous studies, we directly estimate the return to educational quality and explicitly account for the time lag between schooling and the human capital stock attained by adult populations.

Finally, utilizing our newly-constructed databases, we explore the role of educational quality and human capital in understanding cross-country income disparities. By employing development and growth accounting methodologies, we appraise the contribution of quality-adjusted human capital stock to cross-country differences in the levels and growth rates of per worker output relative to that of quantity-based human capital stock. Previous studies, including Hendricks (2002), Caselli (2005), Schoellman (2012), Kaarsen (2014), Hendricks and Schoellman (2018), and Angrist et al. (2021), have used development accounting to measure the contribution of educational quality to cross-country differences in per-worker output. Our new balanced panel dataset on educational quality and human capital can contribute to this line of research by allowing us to ascertain the role of educational quality in understanding cross-country variations in both the levels and growth rates of per-worker output across countries.

The remainder of this paper is organized as follows: Section II describes the construction of our new educational quality database using international test scores. Section III examines the effect of educational quality on wages. Section IV constructs quality-adjusted and quantity-based human capital measures. Section V evaluates the role of educational quality and human capital in explaining cross-country differences in the levels and growth rates of per-worker output. Section VI offers conclusions.

II. Constructing a New Educational Quality Dataset

A. Data on international test scores

We collect international mathematics and science test scores for secondary school students across countries from 1970 to 2015. The primary data sources are TIMSS and PISA. The TIMSS aims to evaluate the mathematics and science achievements of students in Grades 4–8. Six surveys have been conducted since 1995. Waves occur every four years, with the latest in 2015. This wave includes 60 countries and benchmark regions. We use the test scores of students in Grade 8 to construct a measure of secondary education quality. PISA, launched in 1995, measures 15-year-olds' reading, mathematics, and science skills. We use data from six surveys (2000, 2003, 2006, 2009, 2012, and 2015). The 2015 survey included data from 65 countries and regions.

We also use data from the First and Second International Mathematics Study and the Second International Science Study conducted by IEA in the early 1970s and 1980s, and International Assessment of Educational Progress (IAEP) conducted in 1988 and 1990–1991.

Table 1. Number of Available International Mathematics and Science Assessments by Survey Year

No.	Year	Study	Organization	Subject	Countries
1	1970–72	First International Science Study	IEA	S	16
2	1980–82	Second International Mathematics Study	IEA	M	17
3	1983–84	Second International Science Study	IEA	S	17
4	1988, 90–91	International Assessment of Educational Progress	NCES	M, S	6, 19
5	1995, 1999, 2003, 2007, 2011, 2015	Trends in International Mathematics and Science Study (TIMSS)	IEA	M, S	39, 37, 44, 48, 41, 36
6	2000, 2003, 2006, 2009, 2012, 2015	Programme for International Student Assessment (PISA)	OECD	M, S	42, 41, 56, 69, 63, 68

Notes: The number of countries denotes the number of observations included in the study. M and S indicate mathematics and science, respectively.

Table 1 summarizes international mathematics and science assessments of secondary students since 1970. Overall, 102 countries and regions participated in these assessments at least once. We exclude observations from countries whose survey results are not representative of all students.[7] We also exclude countries if their national indicators are not available in the World Bank database of economic and educational statistics because these indicators are important for filling in missing observations and constructing our panel dataset. Our dataset consists of 92 countries, with 647 and 665 observations for mathematics and science, respectively.

B. Consolidating international tests

We initially utilize TIMSS as our primary data source and establish a linkage with PISA. TIMSS allows us to construct consistent international test scores starting from 1995, while the inclusion of PISA test scores add 160 additional survey year observations across countries.[8] The first TIMSS test score in 1995 was reported on a scale ranging from 0 to 1,000, with an average of 500 and a standard deviation of 100. To ensure comparability across all waves of assessment, subsequent TIMSS scores are transformed (Provasnik et al., 2012). PISA test scores are also reported on a scale ranging from 0 to 1,000 and exhibit a high correlation with TIMSS scores. Specifically, for the 92 countries that participated in both tests within our sample, the scores demonstrate correlations of 0.88 in mathematics and 0.92 in science.

These international assessments carefully designed sampling frameworks to ensure that the selected schools in each country are representative in terms of region, gender, selection policy, and performance. TIMSS and PISA are not directly comparable because they have different content and population targets. We adjust PISA test scores to the TIMSS scale using an equipercentile linking method, developed by Braun and Holland (1982), which defines a nonlinear relationship between score scales by setting equal cumulative distribution functions for two different assessments.[9] Online Appendix A comprehensively explains the methodology for converting PISA test scores to the TIMSS scale.

[7] We exclude China and India because the surveys were taken in a few select provinces.

[8] The following 19 countries participated only in PISA: Albania, Argentina, Azerbaijan, Brazil, Costa Rica, Croatia, Kyrgyz Republic, Liechtenstein, Luxembourg, Macao Special Administrative Region, Mauritius, Mexico, Montenegro, Panama, Peru, Poland, Trinidad and Tobago, Uruguay, and Vietnam.

[9] Patel and Sandefur (2020) develop a new methodology by combining two international assessments, the TIMSS and the Progress in International Reading Literacy Study (PIRLS), and two regional

As the IEA and IAEP assessments prior to 1995 use different samples and testing techniques, they are not comparable to TIMSS or PISA. Following Hanushek and Kimko's (2000) approach, we adjust them to be comparable to TIMSS over time using the corresponding National Assessment of Educational Progress (NAEP) scores for US students as an anchor.[10] Considering that the US regularly participates in all international assessments, its NAEP score patterns can be used to compare test score levels over time. In addition, following Hanushek and Woessmann's (2012) approach, we equalize the variances of test scores for 13 core Organisation for Economic Cooperation and Development (OECD) countries across these earlier international tests.

C. Constructing the panel dataset

We aim to construct a complete panel dataset of test scores for 1970, 1980, 1984, 1990, 1995, 1999, 2003, 2007, 2011, and 2015. We use all available survey data for the sample period. When a year has multiple observations, we primarily use TIMSS test scores. During the sample period, 442 mathematics and 459 science test scores are obtained.

Figure 1 plots mathematics and science test scores for countries that participated in assessments of both subjects. The scores are highly correlated. Therefore, we construct our educational quality measure using a simple average of mathematics and science test scores.[11] We obtain 468 actual observations of test scores; and a country has 5.1 observations on average during the period. We expect to fill 452 cells (about 49% of 920 total cells) in our panel data for the 92 countries.

---

assessments, Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) and Programme d'analyse des systèmes éducatifs de la CONFEMEN (PASEC). They estimate the conversion functions among different tests using the results of these exams for a single sample of primary school students in Bihar, India. This methodology is not applicable to our data construction, which does not combine international and regional assessments or math and science (TIMSS) and reading (PIRLS) scores. We also focus on test scores of secondary rather than primary school students.

[10] We rescale the scores for earlier assessments using mean NAEP scores for 13-year-old US students. Using NAEP scores for 17-year-old US students does not affect the main results.

[11] The results in Sections III and IV are robust to using individual mathematics or science test scores to measure educational quality.

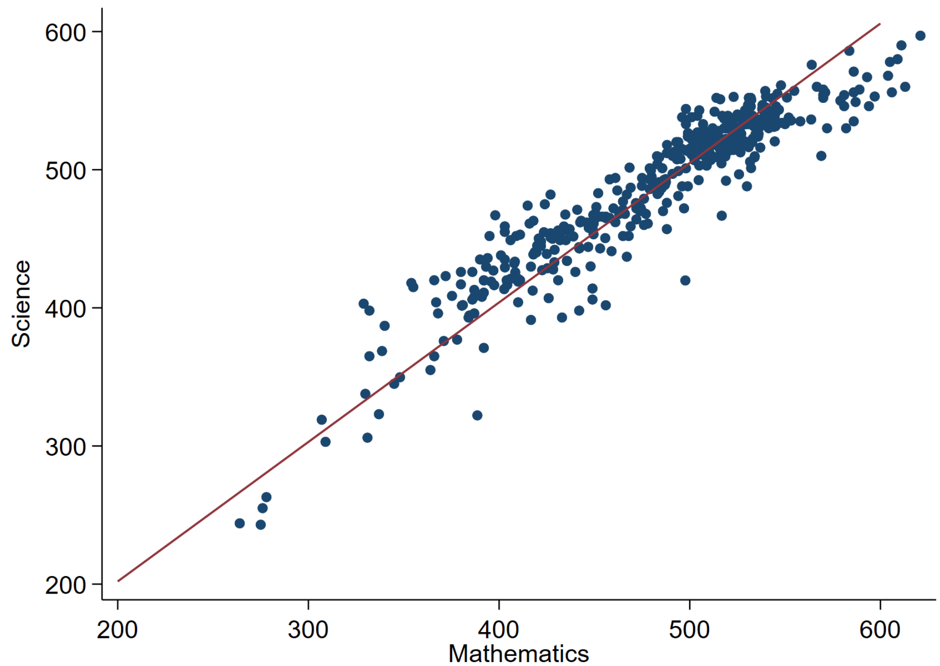


Figure 1. Comparison of Mathematics and Science Test Scores

We employ two imputation methods—interpolation/extrapolation and machine-learning techniques—to estimate unobserved test scores for years when individual countries did not participate in the surveys. This enables us to create a balanced panel dataset. The interpolation/extrapolation technique utilizes linear interpolations with a time trend to fill in missing observations. Additionally, for earlier years, we use extrapolations assuming that the missing observations are the same as the nearest available test score. Meanwhile, the machine-learning technique utilizes a wide range of inputs, including available test scores and a set of economic and educational variables, to estimate the unobserved test scores. We employ least absolute shrinkage and selection operator (LASSO) regression, which performs both feature selection and regularization to improve the predictive accuracy of the resulting statistical model. Further details of the machine-learning technique can be found in Online Appendix B. Notably, the estimates obtained from both interpolation and machine-learning techniques exhibit high similarity, with a correlation coefficient of 0.993.

Combining the estimates, we construct a complete dataset of educational achievements of secondary school students in 92 countries for 10 selected years, between 1970 and 2015. Online

Appendix Table C.1 presents the list of these countries. Figure 2 displays the trends in average test scores from 1970 to 2015 for the world, advanced economies (OECD), and developing economies (non-OECD) based on our new educational quality database. In this database, we employed interpolated estimates to address missing observations. Moreover, the trends observed using alternative estimates obtained through machine-learning techniques exhibit a high degree of similarity.[12] The test scores increase consistently over time for all groups. On average, secondary school students in advanced economies perform better on international mathematics and science assessments than students in developing economies throughout the period.

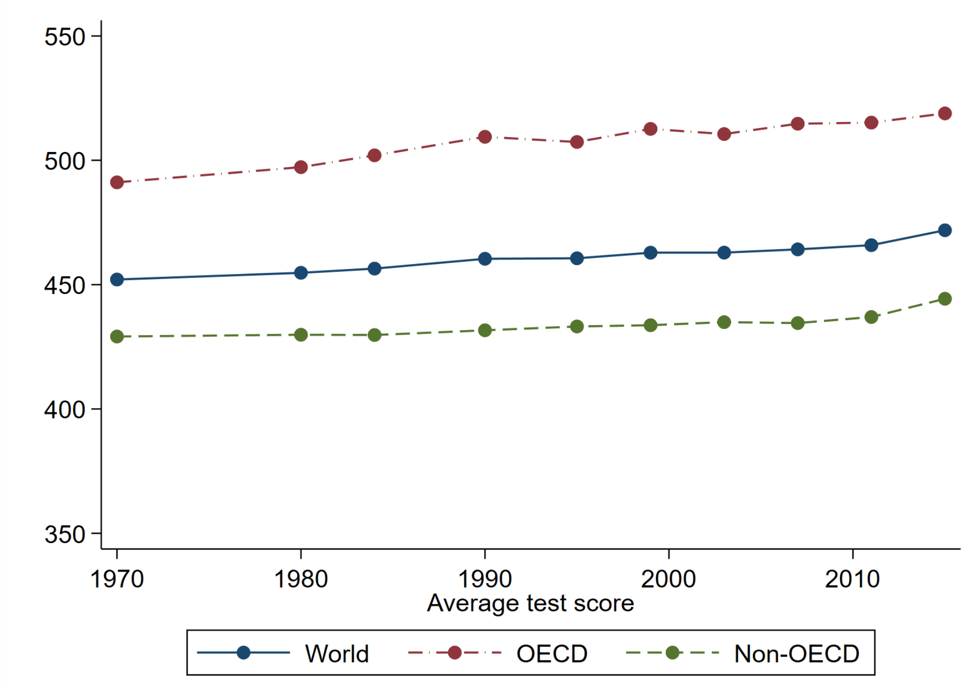


Figure 2. Trends in Average Test Scores for Secondary School Students, 1970–2015

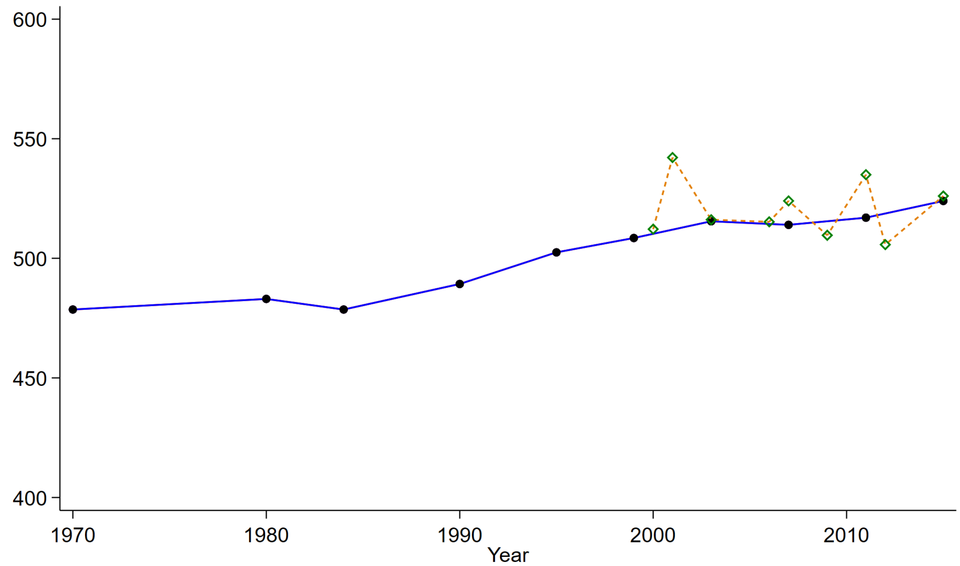
Note: The figures are unweighted averages of test scores for the world, advanced economies (OECD), and developing economies (non-OECD).

[12] The estimation results of the wage equation and estimates of human capital stock measures using alternative test scores, which were imputed through a machine-learning technique, are nearly identical to those reported in the following sections.

Online Appendix Figure C presents trends in average test scores and compares our two estimates (based on interpolation and machine-learning techniques) with those from the latest database constructed by Angrist et al. (2021) for individual countries. They are closely related, albeit with some differences. Angrist et al.'s scores, which combine the actual test scores of primary and secondary school students in three different subjects, often fluctuate over time. By contrast, our estimates, which rely on more comparable data on secondary students' math and science test scores, appear more stable over time. In addition, we generate much longer time-series data using more actual test scores available from earlier surveys.

Figure 3 compares our estimates with those from Angrist et al.'s database for two countries (the US and the Philippines). Our scores steadily increase over time for the US, while those of Angrist et al. (2021) fluctuate. For the Philippines, we obtained more actual data on secondary school students' math and science test scores from earlier surveys. Although we do not use actual reading test scores for primary school students in 2013 as in Angrist et al. (2021), our estimates steadily increase from 2003 to 2015.

A. United States



B. Philippines

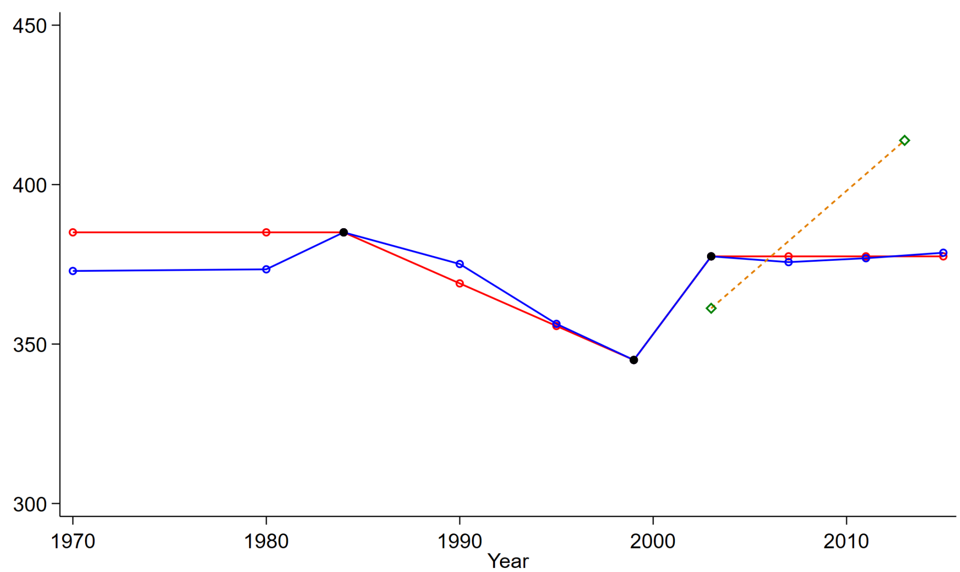


Figure 3. Comparison of Educational Quality Estimates, Selected Countries

Notes: Angrist et al.'s (2021) estimates, denoted by a hollow diamond, are actual test scores in mathematics, science, and reading for both primary and secondary students. If multiple observations are available for the same year, a simple average is used. In our estimates, the actual test scores for secondary students are represented by black dots, whereas the imputed scores are depicted as hollow circles. The interpolated scores are shown in red and machine learning estimates are displayed in blue.

III. Effect of Educational Quality on Earnings

A. Empirical framework

Using our new educational quality database, we assess the role of educational quality in determining individual wages across countries. Human capital, which is a worker's stock of skills and knowledge, is a major determinant of lifetime earnings.

Human capital is determined by schooling, labor market experience, family factors, and innate individual abilities.

(1) $$H = \mathrm{F}\,(\,Schooling, Experience, Family, Ability, \ldots\,) + \mu$$

We focus on the role of schooling in earnings through human capital accumulation. Educational attainment is a composite of quantity (*S*) and quality (*Q*).

(2) $$Schooling\; = \Phi(\mathrm{Q})\mathrm{S},\;\; \Phi^{'} > 0$$

Educational quantity is typically measured by the number of years of schooling. We measured education quality using internationally-comparable test scores.

We can set up an augmented version of the Mincer-type wage equation with educational quantity and quality as independent variables as follows:

(3) $$ln\big(W_{i,j,t}\big) = \alpha + \beta_1\, S_{i,j,t} + \beta_2 Q_{i,j,t} + \beta_3 Exp_{i,j,t} + \beta_4 Exp_{i,j,t}^2 + \beta_5 X_{i,j,t} + u_j + \mu_t + e_{i,j,t}.$$

$ln\;(W_{i,j,t})$ is the logarithm of hourly earnings of individual i in country of origin *j* at time *t* (the year of labor survey); $S_{i,j,t}$ is the number of years of schooling attained by individual *i*'s cohort in country of origin *j*, $Q_{i,j,t}$ is the secondary-level test score of individual *i*'s cohort in country of origin *j*; $Exp_{i,j,t}$ is individual i's years of potential work experience (i.e., age minus years of schooling minus six); and $X_{i,j,t}$ denotes other factors, including individual and labor market characteristics. This specification is estimated using US immigrant data derived from a combination of two survey datasets encompassing individuals from a wide range of countries of origin. We control for country-of-origin and time (survey year) fixed effects. Although we use micro-level individual data to estimate Equation (3), we only have cohort-level data on educational quality. We assume that individuals within a birth cohort in a country have the same quality of education as measured by the average test scores at the secondary level.

In this formulation, the coefficient of years of schooling (i.e., β_1) measures the impact of educational quantity, that is, an additional year of schooling, on the logarithm of the wage, controlling for educational quality. β_2 measures the impact of workers' educational quality on their wages, controlling for years of schooling. Considering that these coefficients reveal the marginal returns to educational quantity and quality, they can be used to construct a composite measure of human capital stock.

Equation (3) is estimated using wage and schooling variables in logarithmic terms. Therefore, the educational quantity and quality terms are entered separately, in line with previous studies, including Hanushek and Woessmann (2008). This property suggests that differences in educational quality are associated with (logarithmic) wage differentials, holding years of schooling constant, as discussed by Speakman and Welch (2006). This equation can easily interpret the impact of educational quality on wages separately from that of educational quantity. This equation also allows educational quality to affect wages through other channels, such as human capital externalities and technological progress, independent of educational quantity.

In an alternative equation, increasing educational quality increases wages by improving returns to additional schooling (i.e., β_1) (Card & Krueger, 1992).

$$ln(W_{i,j,t}) = \alpha + \beta_1 S_{i,j,t} + \beta_2 Q_{i,j,t} S_{i,j,t} + \beta_3 Exp_{i,j,t} + \beta_4 Exp^2_{i,j,t} + \beta_5 X_{i,j,t} + u_j + \mu_t + e_{i,j,t} \quad (4)$$

This equation indicates that the wage differential owing to years of schooling increases faster as educational quality rises. β_2 measures the impact of educational quality on wages for the mean quantity of education, that is, $\frac{\partial ln(W_{i,j,t})}{\partial Q_{i,j,t}} = \beta_2 \cdot mean(S_{i,j,t})$.

We adopt the augmented Mincer equations, Equations (3) and (4), to assess whether educational quality significantly affects wages after controlling for years of schooling and other major determinants. Some studies, including Schoellman (2012) and Li and Sweetman (2014), adopt a two-step approach. In the first step, they estimate Equation (4) using individual-level data without educational quality, allowing the coefficient of years of schooling to vary by country. Therefore, the estimated coefficient of the years of schooling measures the returns to schooling. In the second step, they investigate the relationship between educational quality and the estimated returns to schooling using country-level data. Our approach estimates Equations (3) and (4) directly using educational quality data that differ across cohorts within a country.

Our estimation technique faces some empirical issues. Exogenous variations in test scores could not be explicitly identified. Bias can occur owing to omitted variables that influence an individual's test scores and earnings. For example, a worker's unobserved ability and family background may be correlated with both educational quality, measured by test scores, and future wages. Considering the potential omitted variable bias, we are unsure of the extent to which β_2 reflects a direct causal effect. Our empirical estimations of Equations (3) and (4) include a set of control variables reflecting individual characteristics, and country-of-origin and survey-year fixed effects, which can help reduce omitted variable bias. Our empirical framework, using a cohort-level education quality measure for each individual, also helps reduce endogeneity. Although we have attempted to address the endogeneity issue, we still lack convincing evidence. In principle, this endogeneity bias can be addressed with instruments; however, it is challenging to identify good instrumental variables to assess the independent effects of test scores in our empirical framework using US immigrant data. We are also unable to conduct an experiment that randomly introduces varying levels of educational quality among a group of individuals and observes subsequent labor market outcomes. Similarly, employing a quasi-experimental technique such as regression discontinuity is not feasible.[13]

Bias can also occur because of sample selection concerning the likelihood of labor market participation. Individuals can decide whether or not to participate in the labor market. Those with higher earnings potential are more likely to participate in the labor market than those with lower earnings potential. Educational quantity and quality affect not only wages, but also employment status. Adults with better skills, obtained through higher-quality education, may be able to obtain wages above their reservation wages, leading to more labor-market participation. Conversely, low-skilled workers are less likely to find jobs. To prevent this sample selection bias, we adopt Heckman's (1979) two-stage sample selection model.

Another potential source of bias arises from measurement errors. The imputed data used for international test scores, whether through interpolation or machine-learning techniques, may be susceptible to potential mismeasurement. Therefore, we estimate the wage regressions using

[13] For comprehensive discussions on econometric issues, particularly those pertaining to endogeneity in estimating returns to educational quantity and quality in wage regressions, refer to the surveys conducted by Card (1999), Gunderson and Oreopoulos (2020), and Hoekstra (2020). These surveys also provide insights into existing studies that have addressed these issues.

the sample with only original data, excluding imputed data. We find that the estimation results are not significantly different when the sample includes imputed data.

Another measurement error may arise from using the origin-country cohort average test score as an imprecise measure of the individual's educational quality. Let us examine the main estimation equation: $lnW_i = \alpha + \beta Q_i + u_i$, where Q_i is the quality of schooling that individuals received and is represented by an individual-level test score. Q_i is measured with error. We only have cohort-level data on educational quality. For each individual in a cohort, we can express $Q_i = \bar{Q} + \mu_i$, where $\bar{Q}$ represents the aggregate country-level average test score for all individuals in the same cohort, and μ_i represents the individual-level measurement error. Accordingly, the estimation equation transforms into $lnW_i = \alpha + \beta\bar{Q} + \beta\mu_i + u_i$. In the context of standard measurement error framework, the ordinary least squares (OLS) estimator $\hat{\beta}$ is smaller than the true β. If we define a measure of bias as $\beta/\widehat{\beta}$, it is calculated by $(var\bar{Q} + 2var\mu_i/(var\bar{Q} + var\mu_i)$.[14] Therefore, the magnitude of bias would be larger if the variance of the individual-level test score within countries ($\mu_i = Q_i - \bar{Q}$) relative to the country-level average test score ($\bar{Q}$) is larger.

In order to assess the extent of bias, we calculate the variances of average test scores and individual test scores within countries using the TIMSS individual-level dataset. The standard deviation of the average test scores across countries is approximately 72, whereas the standard deviation of the individual test scores within countries is around 80. Consequently, the bias measure is estimated to be 1.55, suggesting that the true estimate is approximately 55% higher than the OLS estimate $\hat{\beta}$. To address this, we employ a "bias-corrected estimate," represented as ($\hat{\beta}\times$ 1.55), as an alternative measure of the returns to educational quality.

Note that this bias measure is derived under the assumption that the individual-level measurement error results solely from variations in the quality of schooling that individuals

[14] The ordinary least squares estimator $\hat{\beta}$ is given by $cov(\bar{Q}, \alpha + \beta\bar{Q} + \beta\mu_i + u_i)/cov(\bar{Q}, \bar{Q}) = \beta + cov(\bar{Q}, \beta\mu_i)/cov(\bar{Q}, \bar{Q}) = \beta(1 + cov(\bar{Q}, \mu_i)/cov(\bar{Q}, \bar{Q}))$. By replacing $\bar{Q}$ with $Q_i - \mu_i$, we have $\hat{\beta} = \beta(1 + cov(Q_i - \mu_i, \mu_i)/cov(Q_i - \mu_i, Q_i - \mu_i))$. Under the assumption of no correlation between Q_i and μ_i, we obtain $\hat{\beta} = \beta(varQ_i/(varQ_i + var\mu_i))$. Therefore, $\hat{\beta}$ is smaller than the true β. Note that $varQ_i/(varQ_i + var\mu_i) = (var\bar{Q} + var(Q_i - \bar{Q}))/(var\bar{Q} + 2var(Q_i - \bar{Q}))$. Accordingly, the bias measure $\beta/\widehat{\beta}$ becomes larger as the variance of the individual-level test score relative to the country-level average test score increases.

receive. However, other factors, such as unobserved students' abilities, can contribute to differences in individual-level test scores within and between countries.[15] It is plausible that some portion of the calculated 55% bias is attributable to variations in these factors rather than educational quality. Although determining the exact magnitude of the bias resulting from all measurement errors is challenging, we can regard our calculation of 1.55 as an upper bound for the bias resulting from the measurement error caused by employing the cohort average test score as a proxy for an individual's educational quality.

B. Data

To estimate the augmented Mincer equation, we use the dataset of immigrants in the US labor market, distinguishing workers *i* by their country of origin *j*, where they were born and educated before immigrating to the US.

US immigrant data is taken from the American Community Survey five-year sample, a 5-in-100 national random sample of the population provided by IPUMS, US (Ruggles et al., 2020). We combine the annual datasets of the American Community Survey for five years and generate a complete dataset with one observation over a period of five years to reduce the effect of missing data, particularly for sparsely-populated regions and sub-populations. We utilize two five-year samples, specifically from 2009 to 2013 and from 2014 to 2018, resulting in a dataset with two observations separated by approximately five-year intervals. We investigate the impact of schooling quality in immigrants' home countries on their wages in the US labor market considering immigrants from various countries of origin.

Immigrants are identified based on their country of origin. We restrict our sample to immigrants aged 25–53 who arrived in the US after their expected date of graduation from the highest level of education that they attained. We select individuals who worked at least 30 hours

[15] Previous research has demonstrated that differences in abilities play an important role in accounting for the variation in students' academic achievements, although the magnitude of the bias is considered minor (Card, 1999; Gunderson & Oreopolous, 2020). Additionally, studies show that some cross-country variations in student achievement are also attributed to unobserved factors including intellectual abilities (Hanushek & Woessmann, 2011). Lynn (1982) contends that the intellectual abilities of East Asian students, rather than their efforts or other school factors, are connected to their individual success in mathematics. Various researchers, including Flynn (1987), have observed secular gains in IQ, even among preschool children, in various countries. These gains may be indicative of environmental factors such as improvements in nutrition and medical care, as well as reductions in fertility.

per week for 48 weeks in the previous year, because the survey data report an interval measure of weeks worked: 1–13, 14–26, 27–39, 40–47, 48–49, and 50–52. We exclude self-employed workers and those who only attended school. The Census includes a measure of schooling attainment, which we recode as years of schooling in the usual manner.

To calculate average hourly wages, we used data on the previous year's annual wage income (in 2010 US dollars), weeks worked, and hours worked per week. To avoid the influence of outliers, we exclude immigrants in the bottom or top 1% of earners in each country of origin. We include several control variables such as gender, workplace region, disability status, workplace in a metropolitan area, self-reported English language proficiency, and years of immigration.

Test scores are based on the estimates in this study. We assume that the respondents were 15 years old at the time of testing. We assign international test scores of the students in their country of origin using the respondents' birth years. We standardize these test scores in the subsequent regression analyses to obtain a mean of zero and a standard deviation of one. Online Appendix Table C.2 provides summary statistics by birth country for US immigrants from 76 countries. Immigrants from developed countries tend to have more years of schooling and higher-quality schooling.

US immigrant data cover numerous countries. However, they are subject to sample selection bias owing to immigration selectivity (Borjas, 1999). Immigrants are self-selected. The choice to migrate to the US depends on the benefits and costs. If migration brings sufficiently high earnings to cover the migration costs, these individuals will migrate. Immigrants may believe that they are more likely to find better opportunities in the US than in their home country. Controlling for the potential immigrant selection bias is challenging. Accordingly, we include country-of-origin and survey-year fixed effects and individual characteristics such as English language proficiency in our estimation. Our approach uses a direct measure of educational quality for individual countries, and estimates a common rate of return for all countries. Therefore, it can minimize bias from estimating country-specific returns to educational quantity, which is used to measure educational quality for individual countries, as in previous studies.

C. Estimation results

Table 2 reports the regression results for Equation (3) using data on US immigrants. We use the restricted sample of the actual test scores. Online Appendix Table C.3. shows the estimation results using the extended sample, which includes the imputed test scores by interpolations or extrapolations. The estimates are similar to those in the baseline case of Column (2) in Table 2.

Table 2. Regression for Wages of US Immigrants

	(1)	(2)	(3)
Years of schooling	0.073***	0.075***	0.075***
	(0.005)	(0.005)	(0.005)
Experience	0.047***	0.048***	0.046***
	(0.007)	(0.007)	(0.007)
Experience square	-0.075***	-0.073***	-0.071***
	(0.018)	(0.018)	(0.018)
Average test score		0.095**	
		(0.048)	
Average test score× Years of schooling			0.004
			(0.002)
Other controls	Yes	Yes	Yes
Country-of-origin fixed effects	Yes	Yes	Yes
Survey-year fixed effects	Yes	Yes	Yes
Number of observations (N)	20,503	20,503	20,503
R^2	0.417	0.418	0.417

Notes: US immigrant data are from the American Community Survey 2009–2013 and 2014–2018 five-year samples The dependent variable is the logarithmic value of gross hourly wages (in 2010 US dollars). The sample consists of full-time employees aged 25–53. Test scores are standardized to have a mean of zero and a standard deviation of one. The regression is based on a sample of actual test scores without imputed values. Least-squares regressions weighted by the sampling weights are adopted. All regressions control for region of workplace, gender, disability status, workplace in a metropolitan area, self-reported English language proficiency, and year of immigration. The value of the experience square term is divided by 100. Robust standard errors are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

First, we estimate this equation without an educational quality measure. The results are consistent with our predictions. In Column (1) of Table 2, the coefficient of years of schooling is 0.073, implying that an additional year of education increases wages by 7.3%.

Column (2) of Table 2 shows the estimates, including the average test score. The coefficient of years of schooling remains the same. As our test score measure is standardized, its estimated coefficient can be interpreted as the percentage increase in wages associated with a one standard deviation increase in the average test score, that is, educational quality. This coefficient is 0.095 and statistically significant, suggesting that a one-standard-deviation increase in the average test score is associated with an average increase in hourly wages of 9.5%.[16] This effect is substantial, exceeding the impact of an additional year of schooling. The results confirm that, on average, immigrants who received secondary education in countries with higher test scores before immigration earn higher wages in the US.

Table 2 presents the estimation results of Equation (4). Column (3) includes the interaction between test scores and average years of schooling instead of the educational quality variable. The coefficient of years of schooling does not change. The coefficient of the interaction between average years of schooling and average test score is statistically insignificant. For the mean years of schooling (14.9 years), the estimated coefficient implies that a one standard deviation increase in the average test score increases hourly wages by about 6.0%.[17]

Therefore, Specification (4) performs better than Specification (5), suggesting that educational quality affects wages separately from educational quantity.

Figure 4 provides a graphical description of the partial relationship between wage and educational quality as measured by the test score. The horizontal axis represents the normalized test scores. The vertical axis shows the values of the partial regression residuals, which refer to the wage rate filtered for the estimated effects of explanatory variables other than the test score. The residuals are computed from Column (2) of Table 2. This figure clearly demonstrates a positive relationship between the two variables, and this estimated relationship is not driven by outlier observations. This implies that the educational quality for each cohort in an origin country has a significantly positive impact on the wages of immigrants of the cohort from that country in the US labor market. An improvement in educational quality from the old to the younger cohort for each country results in a higher relative wage of immigrants from the younger cohort.

[16] The standard deviation of country average scores is 66.8 in the sample of wage regression.

[17] We also perform a two-step estimation, in line with Schoellman (2012) and Li and Sweetman (2014), that estimates country-specific returns to educational quantity and then estimates the impact of educational quality (i.e., test scores) on these returns. The quantitative impact of educational quality on wages in this two-step estimation is much smaller than that from our reduced-form estimation. The results are available on request.

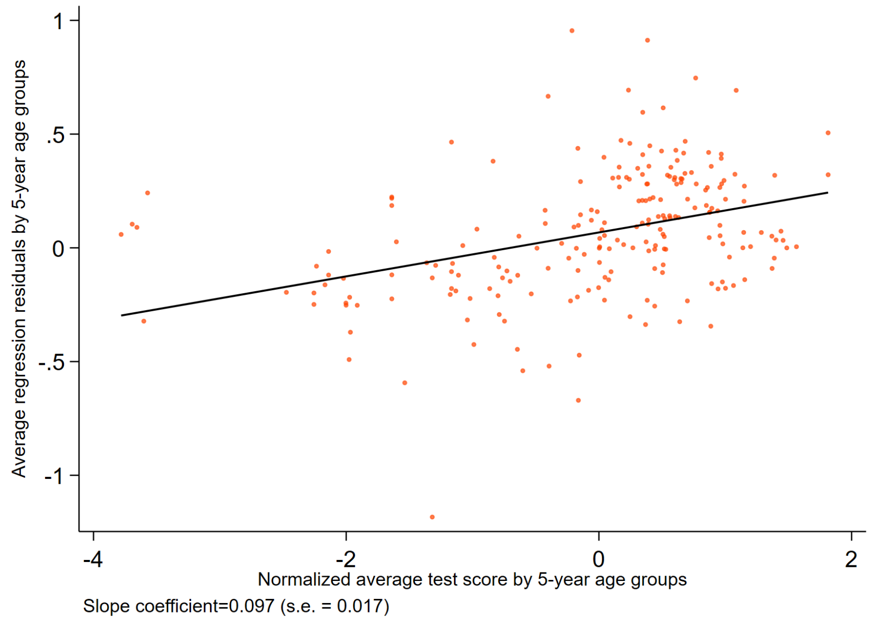


Figure 4. Scatter Plots between Test Scores and Partial Regression Residuals

Notes: Partial regression residuals are computed by the residuals of regressing the wage variable against all explanatory variables except the test score in the specification of Columns 2 and 4 of Table 2. The residuals and test scores are averaged for each 5-year cohort.

D. Robustness

In general, the estimated returns to immigrants' educational quality remain robust when using different samples and additional controls. As shown in Table 3, the estimates change little when we include US natives or excluding immigrants from Mexico, who are overrepresented in our sample. When we include self-employed immigrants, the estimated return to educational quality increases marginally to 10.5%. The estimate increases to 12.2% when we exclude naturalized immigrants. When we exclude immigrants who entered the US less than three years after their expected graduation dates, the return increases significantly to 15.1%. A sizable fraction of immigrants' earnings is imputed in the American Community Survey (Bertoli & Stillman, 2019). When we exclude these imputed observations, the return decreases to 8.8%. Our baseline model uses a quadratic polynomial for potential work experience. Workers may accumulate various levels of human capital when working in their home countries prior to immigration (Lagakos et al., 2018a). When we include potential work experience before immigration as another control variable, the estimated return to educational quality does not change.[18] We also include cohort fixed effects in the regressions to control for cohort-specific shocks, and determine that the estimated return to educational quality decreases to 9.3%.

To address the potential selection bias from non-employment, we adopt Heckman's (1979) two-step model. We utilize information on an individual's number of children to identify the participation equation. The application of the Heckman two-step model reveals a lower estimated returns to test scores (8.3%) compared to the baseline model. Consequently, disregarding selection bias tends to lead to an overestimation of the overall returns on educational quality.

In sum, the estimated returns to educational quality range from 8.3% to 12.2% among US immigrants, depending on the specific sample and additional controls. In the next section, we construct a measure of human capital stock using the results in Column (2) of Table 2 for US immigrants as a baseline. The rate of return to one standard deviation of educational quality, measured by average test scores, is assumed to be 9.5%, and the rate of return to an additional year of education is assumed to be 7.5%. We also construct an alternative measure of quality-adjusted human capital per worker by assuming the bias-corrected estimate of return to educational quality as 14.7 (9.5×1.55).

[18] We also test whether returns to educational quality vary by the duration of residence, that is, the number of years that the worker spent in the US since the year of immigration. We find that these returns decline with the duration of residence in the US.

Table 3. Regressions with Alternative Samples and Additional Control Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Baseline (Column 2, Table 2)	Alternative sample						Additional control		Control for non-labor market participation
		Including US natives	Without Mexico	Including self-employed	Non-naturalized only	3-year buffer	Excluding imputed wages	A quadratic polynomial in home-country potential work experience	5-year cohort fixed effects	Heckman selection model
Years of schooling	0.075***	0.075***	0.076***	0.077***	0.074***	0.073***	0.077***	0.074***	0.080***	0.083***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)	(0.005)
Experience	0.048***	0.048***	0.048***	0.043***	0.054***	0.039***	0.053***	0.064***	0.053***	0.052***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.009)	(0.007)	(0.012)	(0.008)	(0.006)
Experience square	-0.073***	-0.074***	-0.074***	-0.060***	-0.091***	-0.041**	-0.085***	-0.103***	-0.074***	-0.081***
	(0.018)	(0.017)	(0.018)	(0.018)	(0.020)	(0.021)	(0.018)	(0.019)	(0.021)	(0.017)
Average test score	0.095**	0.095**	0.096**	0.105**	0.122**	0.151**	0.088	0.095**	0.093*	0.083*
	(0.048)	(0.048)	(0.048)	(0.049)	(0.055)	(0.059)	(0.053)	(0.048)	(0.048)	(0.049)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country-of-origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Survey-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	20,503	984,214	14,032	21,462	15,755	14,643	16,029	20,503	20,503	36,485
R^2	0.418	0.415	0.415	0.382	0.419	0.437	0.455	0.420	0.420	

Notes: The three-year buffer excludes immigrants who entered the US less than three years after their expected graduation date. The exclusion restriction in the selection equation in the Heckman model is a variable indicating the number of children of an individual. See also the notes in Table 2.

IV. Construction of Quality-Adjusted Human Capital Stock

We assume that both years and quality of schooling are important for human capital accumulation. We construct a measure of aggregate quality-adjusted human capital as follows: [19]

$$\text{(5)} \qquad \mathrm{HC_q} = \sum_{\mathrm{a}} \sum_{k} \mathrm{e}^{(\beta_s^k S_k^a + \beta_q^k q_k^a)} l_k^a,$$

where a is the age cohort, ranging from 15–19 to 60–64. k is one of the seven education levels: no schooling, incomplete primary, complete primary, lower secondary, upper secondary, incomplete tertiary, and complete tertiary. S_k^a is the number of years of schooling at level k for cohort group a, q_k^a measures the quality of educational level k attained by cohort group $a,$ and l_k^a is the fraction of cohort group a that has attained educational level k. In this equation, β_s^k measures the efficiency of a unit of labor with educational level k relative to a unit with no formal education and is the marginal return to an additional year of schooling at level k. β_q^k measures the marginal return to the quality of education at level k.

If the marginal returns to educational quantity and quality are constant for all educational levels, then Equation (5) can be rewritten as:

$$\text{(6)} \qquad \mathrm{HC_q} = \sum_{\mathrm{a}} \sum_{\mathrm{k}} \mathrm{e}^{(\beta_s S_k^a + \beta_q q_k^a)} l_k^a.$$

This equation is consistent with the augmented Mincer-type wage regression given by Equation (3).[20] It suggests that for a given educational quality, the logarithmic human capital differential (or logarithmic wage differential) increases proportionally with average years of education by β_s. For a given quantity of education, the logarithmic human capital differential also increases proportionally with educational quality by β_q. In this specification, the educational quality differential is presumed to influence the logarithmic human capital differential regardless of educational quantity. We use Equation (6) to construct a measure of quality-adjusted human capital stock.[21]

[19] This equation implies that human capital per worker across all educational levels is the weighted sum of the shares of workers multiplied by their marginal products (or wage rates). Wages are determined by educational level and quality. This equation implies that the wage rate of a person with no schooling is normalized to equal one.

[20] This specification does not include returns to experience. Some recent studies, such as Lagakos et al. (2018a) and Jedwab et al. (2022), attempt to estimate wage-experience profiles and returns to experience across countries. They estimate different returns to experience in developed and developing economies. We leave the measurement of human capital stock incorporating work experience for future study.

[21] This measure of aggregate human capital stock assumes perfect substitution between workers with different educational attainments. Assuming a lower elasticity of substitution between high-educated and

By contrast, quantity-based human capital stock, which is the conventional human capital stock measure that does not account for educational quality differences, is simplified to

(7) $$\mathrm{HC} = \sum_{\mathrm{a}} \sum_{\mathrm{k}} \mathrm{e}^{\beta_s s_k^a} l_k^a.$$

To measure aggregate human capital stock, we assume that β_q remains consistent across cohorts and countries. Consequently, we hypothesize that the logarithmic human capital differential increases proportionally with test scores, guided by the same β_q both across countries and over time. An issue arises when we use the estimated coefficient in a regression using individual wages to calculate country-level human capital stock. The coefficient, derived from variations in cohorts within countries, may not accurately reflect the causal relationship between school quality and human capital at the country level. Estimating a parameter that can identify the causal effects of test scores on human capital using country-level data is a formidable challenge. A reverse causality problem is more pronounced if high-income countries have parents and teachers with higher average human capital, resulting in higher average test scores among their children. We acknowledge that there are several constraints in Specification (6) for measuring country-level human capital stock while incorporating education quality. Nevertheless, given that previous studies have estimated human capital stock based on years of education using (7), we adopt a similar method to ensure comparability.

We also make several assumptions regarding educational quality and individuals' cognitive skills. First, we assume that the quality of primary and secondary education is the same for individuals in each cohort within a country and is measured by average secondary-level test scores.

Second, we assume that the educational quality attained by tertiary students at age 20 is the same as that achieved by secondary students at age 15, who are concurrently enrolled in school in that country. No internationally comparable data quantitatively measure educational quality at the primary or tertiary levels relative to that at the secondary level across countries and years.

Third, we assume that the test scores of secondary school students before 1970 are similar to the estimates of those in 1970. This assumption is necessary because we lack data on test scores for people who attended school before 1970. We estimate educational quality for people aged 15–

low-educated workers tends to increase differences in the human capital stock over time and across economies (Lee & Lee 2016). This issue seems less important for incorporating educational quality into our aggregate human capital stock because it is assumed to be equal across all education levels in a given year and country.

64. Therefore, we assume that the test scores of workers aged 50 and 64 in 2000 are similar to those of workers aged 45–49 who attended secondary school in 1970. We may have overestimated the educational quality for older cohorts, particularly in earlier years, for countries whose educational quality improved before 1970.

To construct quality-adjusted human capital per worker in Equation (6), using the results in Column (2) of Table 2 for US immigrants, we assume that the marginal rate of returns to an additional year of education is constant at 7.5% for the average educational quality and that a one-standard-deviation increase in educational quality, measured by average test scores, is associated with a 9.5% increase in earnings for the average years of schooling.[22] Data on the proportion of the population that attained each educational level by age group are available in five-year intervals from 1950 to 2015 from Barro and Lee (2013).[23] We also create an alternative measure of quality-adjusted human capital per worker by substituting the estimate of the return to educational quality with a bias-corrected figure of 14.7%.

The final dataset contains estimates of the average quality-adjusted human capital stock for people aged 15–64 for 83 countries at five-year intervals from 1985 to 2015. Figure 5 presents our baseline human capital measure, which accounts for the quality of education for individual countries in 2015, compared to the conventional quantity-based human capital measure without the quality of education, as estimated by Equation (7). The scatterplot clearly shows large cross-country differences in both the quality-adjusted and quantity-based human capital. The two human capital measures are highly correlated across countries, but quality-adjusted human capital stock levels are lower than quantity-based human capital stock levels in many countries. Cross-country differences in quality-adjusted human capital are also greater than those in quantity-based human capital. For example, consider Korea and Morocco, which had the highest and lowest levels of quality-adjusted human capital in 2015, respectively. In Korea, this level is 5.9, which is more than double that of Morocco (2.6). In contrast, the difference in the two countries' quantity-based

[22] Hanushek et al. (2017) construct an estimate of quality-adjusted human capital stock for 47 US states considering returns to education quantity and quality of around 8% and 17%, respectively, based on previous studies using US labor data.

[23] We use the latest version of the Barro-Lee dataset that estimates educational attainment for people between 15 and 64 years old, disaggregated by 10-year age group, in 146 countries at five-year intervals from 1950 to 2015. The dataset is available from www.barrolee.com.

human capital levels is much smaller, with Korea at 2.7 and Morocco at 1.7, as this measure does not account for educational quality differences between the two countries.

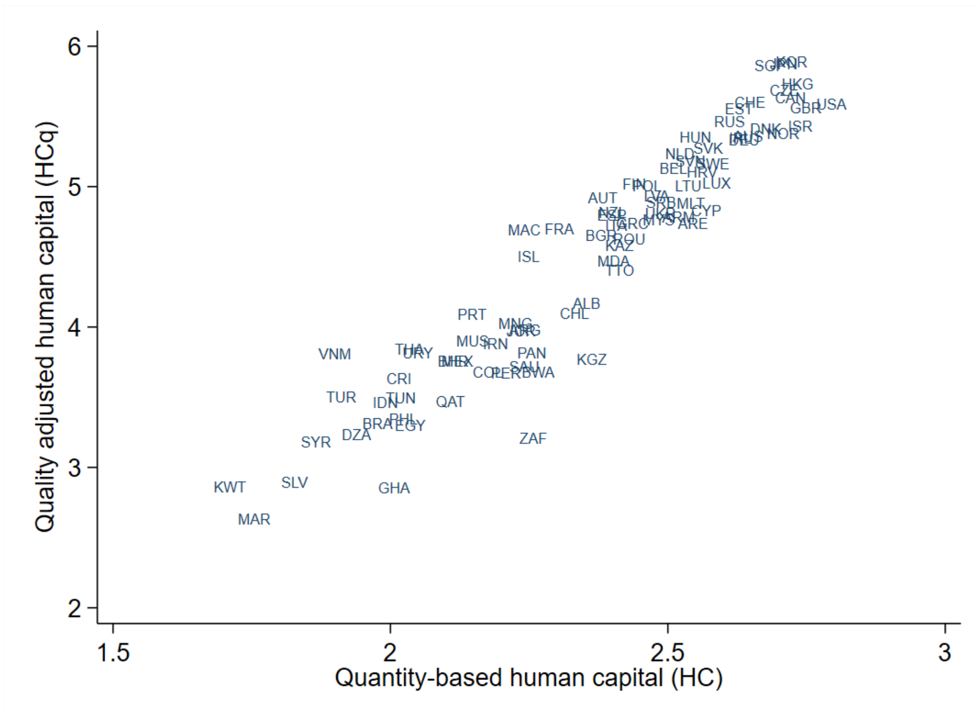


Figure 5. Comparison of Quantity-Based (HC) and Quality-Adjusted (HCq) Human Capital in 2015

Notes: Quantity-based human capital stock refers to a conventional human capital measure based on years of schooling. Quality-adjusted human capital refers to human capital that incorporates quality of education.

Figure 6 shows the evolution of quality-adjusted and quantity-based human capital for the world, OECD, and non-OECD countries over the sample period. Since 1985, quality-adjusted human capital stock estimates have grown faster than quantity-based human capital stock estimates worldwide in both advanced and developing countries. The difference in quality-adjusted human capital stock between advanced and developing countries grew more than that in quantity-based human capital stock over this period.

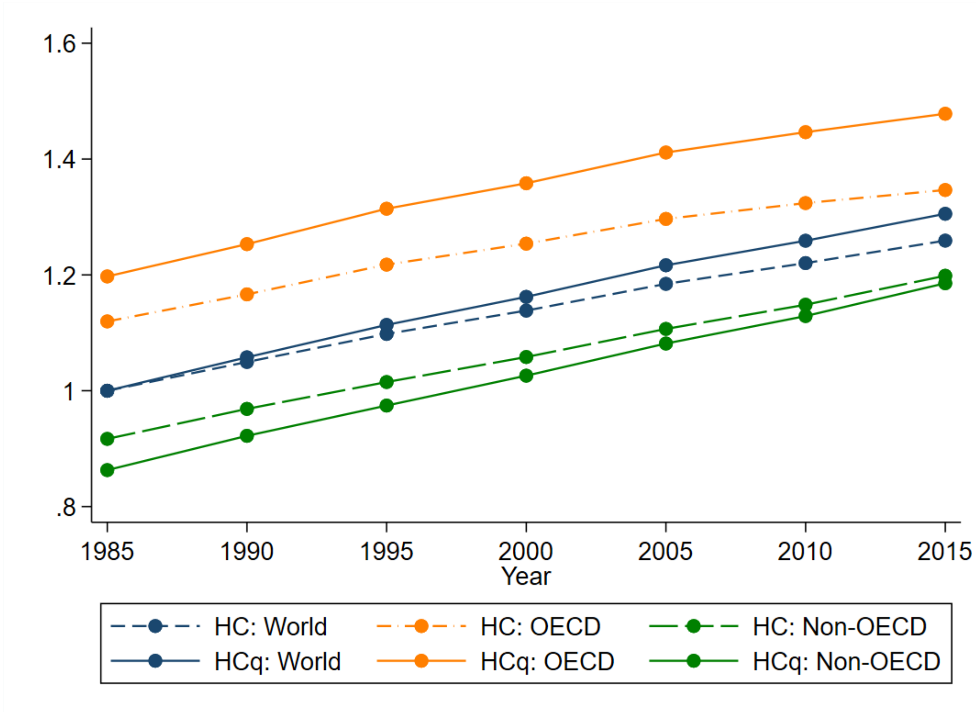


Figure 6. Trends in Quantity-Based (HC) and Quality-Adjusted (HCq) Human Capital, 1985–2015

Notes: The figures show the unweighted averages of human capital stock measures for the world, OECD, and non-OECD countries. Their values for the world in 1985 are normalized to one.

V. Educational Quality, Human Capital, and Cross-country Income Differences

We use our newly-constructed databases to investigate the role of educational quality and human capital in explaining income disparities among countries. As discussed in Section IV, identifying precise causal relationships between school quality and human capital across countries poses a challenge. Similarly, discerning the causal relationship between a country's aggregate human capital and its overall income is challenging. In this section, we rely on development and growth accounting methodologies. Although these methodologies cannot establish direct causal relationships, they serve as useful tools for determining the extent to which differences in human capital explain cross-country variations in the level and growth of income or output.

We consider a simple Cobb–Douglas production function, such as

(8) $$Y = K^{(1-\alpha)}(AhL)^{\alpha} = K^{(1-\alpha)}(hL)^{\alpha}A^{\alpha},$$

where Y is real GDP, K is physical capital, L is the number of workers, h is human capital per worker, A is total factor productivity (TFP), α denotes the labor share of output, and $1-\alpha$ denotes the capital share of output. Technological progress is assumed to be labor-augmenting (i.e., Harrod neutral).

Following Klenow and Rodriguez-Clare (1997), and Hall and Jones (1999), the production function can be rewritten as follows:

(9) $$y = A(k/y)^{(1-\alpha)/\alpha}h.$$

Using this equation, we can categorize the differences in per-worker output (y) across economies by the differences in per-worker human capital stock (h) and other inputs, such as the capital-output ratio (K/Y) and an unobservable residual (TFP).

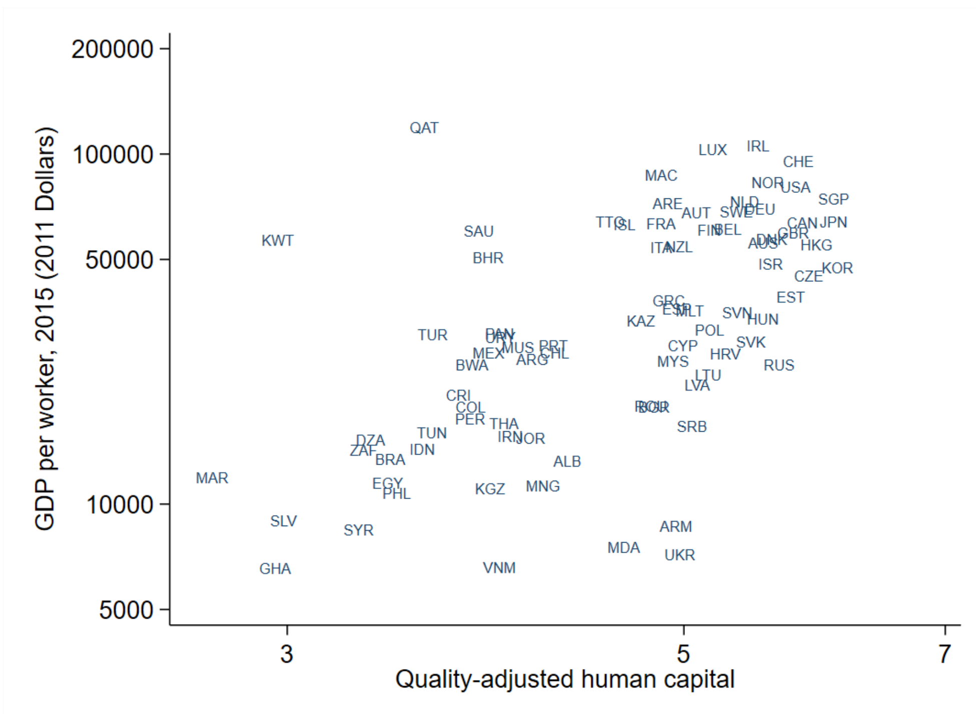


Figure 7. Quality-Adjusted Human Capital Stock and GDP Per Worker across Economies in 2015

Note: GDP per worker data are sourced from PWT version 9.1 (Feenstra et al., 2015), and both the x- and y-axes use a logarithmic scale.

Figure 7 plots GDP per worker against the quality-adjusted human capital stock for economies in 2015. We observe a clear positive bilateral relationship, especially when oil-producing countries, such as Bahrain, Kuwait, Saudi Arabia, and Qatar, are excluded. The countries with the highest levels of human capital stock, such as Japan, Singapore, and

Switzerland, tend to have higher per-worker output levels. In contrast, Ghana, El Salvador, and the Syrian Arab Republic have low levels of output per worker and human capital stock.

We can determine the extent to which the variation in per-worker output (y) across economies can be explained by the variation in human capital stock per worker (h) and other inputs. Accordingly, we take the logarithm of Equation (9). The variance in the logarithm of GDP per worker is the sum of the covariances of the logarithm of GDP per worker with the three additively-separable inputs. We calculate the variance in the logarithm of GDP per worker across countries, explained by the covariance with the logarithm of human capital per worker:

$$(10) \qquad \frac{Cov(lnh, lny)}{Var(ln\, y)}$$

Table 4 shows the estimates of the measure in Equation (10) for 2015 using a sample of 83 countries with complete observations. Data on GDP and workers are from the Penn World Table (PWT) version 9.1 (Feenstra, Inklaar, and Timmer 2015).

Table 4. Development Accounting Results for Cross-Country Differences in GDP Per Worker in 2015

Return specification	Variance of GDP per worker explained by: Quality-adjusted human capital per worker	Variance of GDP per worker explained by: Quantity-based human capital per worker
1. Baseline case *Return to quantity=7.5%; quality=9.5%*	0.157	0.086
2. Alternative case *Return to quantity=7.5%; quality=14.7%*	0.195	0.086

Notes: The sample includes 83 economies. The figures indicate the estimated share of the variance in GDP per worker as explained by each human capital measure. The data used for development accounting (Equation (10)) are sourced from the PWT version 9.1, Barro-Lee human capital data (Barro & Lee 2013), and our educational quality data. The baseline case uses a quality-adjusted human capital measure constructed by assuming the estimated returns to educational quality and quantity in Column (2) of Table 2. The alternative case uses a quality-adjusted human capital measure constructed by assuming an upper-bound rate of return to educational quality of 14.7.

In Table 4, the estimated share of the variance in GDP per worker explained by quality-adjusted human capital per worker is 0.157 in 2015 using the baseline human capital estimates. However, when we measure human capital per worker based on educational quantity, this share decreases to 0.086. therefore, the development accounting exercise shows that quality-adjusted human capital explains a significant fraction of cross-country differences in per-worker output. This result suggests that differences in per-worker human capital stock account for approximately 16% of the cross-country variation in per-worker GDP worldwide, and educational quality accounts for about seven percentage points of this variation. Table 4 also shows the estimates when the alternative measures of educational quality and human capital stock are used. When we use the rates of return to an additional year of schooling (7.5%) and educational quality (14.7%), the contribution of per worker human capital stock to the variance in per worker GDP increases to 19.5%, and educational quality contributes about 11 percentage points.

Our results suggest that 16–20% of the cross-country variation in per-worker GDP in 2015 can be attributed to human capital, with educational quality contributing 7–11 percentage points. The results indicate that the contributions of human capital stock and educational quality critically depend on the estimates of the return to educational quality. This result can be explained intuitively.[24] According to our human capital measure, an increase of one standard deviation in test scores across countries results in a β increase in logarithmic human capital. Consequently, if human capital follows a normal distribution, the gap in human capital between the country at the 84th percentile (i.e., Belgium) and the one at the 16th percentile (i.e., Jordan) is approximately β times 2 standard deviations. This value is either 0.095 or 0.147 depending on our β estimate. In our dataset, the actual gap in the logarithmic per worker GDP for this pair of countries is 1.37. Therefore, school quality explains approximately 7% or 11% of the gap in the logarithmic per worker GDP.

Our estimates of the contributions of educational quality to cross-country variation in per-worker GDP are generally smaller than those reported in existing studies. Other studies such as Schoellman (2012), Jones (2014), Manuelli and Seshadri (2014), and Angrist et al. (2021) find a greater contribution of human capital to per worker output in development accounting when differences in educational quality are incorporated. These differences mainly result from different assumptions, samples, and measurements that generate different returns to educational quality (and

[24] We appreciate the anonymous reviewer for suggesting this example.

quantity) across countries. We assume a consistent rate of return to educational quality across countries to construct the aggregate human capital stock. If high-income countries exhibit higher returns to educational quality, the estimated contributions increase significantly. Some studies also assume that a higher quality of education leads to a proportional increase in educational quantity (Bils & Klenow, 2000; Schoellman, 2012). Jones (2014) accounts for imperfect substitutability between different types of skills in production.

We also assess the contribution of human capital to cross-country differences in the growth rate of per-worker output. Our new balanced panel data on educational quality and human capital enables us to conduct this assessment using growth accounting exercise.

The growth accounting method categorizes the growth in output per worker (y) into three productive inputs: physical capital per worker (k), human capital per worker (h), and total factor productivity (TFP).

$$\frac{dy}{y} = (1-\alpha)\frac{dk}{k} + \alpha\frac{dh}{h} + \alpha\frac{dA}{A}, \tag{11}$$

where dx/x represents the percentage rate of change of x. We assume that the marginal products can be measured by factor prices and set α equal to the average labor share for the whole period.

Table 5. Growth Accounting for GDP Per Worker Aged 15–64, 1995–2015

		Contribution from	
Return specification	Global growth rate of per-worker GDP	Quality-adjusted human capital per worker	Quantity-based human capital per worker
1. Baseline case *Return to quantity=7.5%; quality=9.5%*	0.0171	0.0040 (23.3%)	0.0032 (19.0%)
2. Alternative case *Return to quantity=7.5%; quality=14.7%*	0.0171	0.0043 (25.3%)	0.0032 (19.0%)

Notes: The data are sourced from PWT version 9.1, Barro-Lee human capital data, and our educational quality data. The global GDP growth rate is the average annual GDP growth rate for 76 countries from 1995 to 2015, with weights assigned based on each country's share of real global GDP at the current purchasing power parity. The contribution of human capital per worker is the average growth rate multiplied by the average share of the GDP for that period.

Table 5 reports the results of the growth accounting decomposition for 1995 to 2015 for the 76 countries with complete observations during that period.[25] We compute weighted global averages using each country's share of real global GDP at current purchasing power parities. The global average annual per-worker GDP growth rate for 1995 to 2015 was 1.71%.

The results of growth accounting reveal a consistent contribution of educational quality and human capital to per worker growth, whether using baseline or alternative measures. Approximately 0.4 percentage points, or 25%, of the world's average per-worker GDP growth rate of 1.71% can be attributed to growth in quality-adjusted human capital per worker.[26] In contrast, human capital's contribution to per worker GDP growth is about 0.3 percentage points when measured based on educational quantity. Therefore, educational quality has contributed about 0.1 percentage points to per-worker GDP growth.[27] The estimates suggest a modest contribution of educational quality to per-worker GDP growth. This is attributable to the small, on average, increase in quality-adjusted human capital within countries during the sample period. These increases are associated with fewer changes in the average test scores of the cohorts of secondary school students in the earlier periods preceding the sample period.

VI. Conclusion

This study constructed a new database of educational quality by cohort across countries based on internationally-comparable secondary-level mathematics and science test scores, and examined the effects of educational quality on wages using micro-level data on US immigrants. We found significant returns to the quality and quantity of education. On average, a standard deviation increase of 1 in international test scores is associated with a wage increase of between 9.5% and 14.7%, which exceeds the returns to an additional year of schooling.

[25] We also apply the growth accounting decomposition from 1985 to 2015 to the 62 countries with complete data in that period. The estimated contributions of educational quality and human capital to per-worker GDP growth are similar.

[26] When we use the alternative measure of human capital stock that is based on the estimated returns to educational quantity and quality using the sample with imputed test scores, the estimated fractions of the average per-worker GDP growth rate explained by quality-adjusted human capital per worker are similar to the baseline estimates.

[27] The augmented Mincer-type specification in Equation (6) shows that the growth rate of quality-adjusted human capital is the sum of two components—the growth rate of educational quantity (i.e., quantity-based human capital) and the growth rate of educational quality.

Using our new dataset of cohort-based test scores and estimated returns to educational quantity and quality, we estimated the quality-adjusted human capital stock between 1985 and 2015. We found vast differences in educational quality and human capital stock between cohorts and across countries.

Our newly-constructed human capital dataset enabled us to explore the role of educational quality and human capital in understanding cross-country income disparities. The findings from the development and growth accounting exercises indicated that there is a discernible contribution of educational quality and human capital to per capita income and its growth rate.

One caveat of this study is the assumption that educational quality generates consistent returns across countries over time. Differences in economic structures and labor markets can alter returns to educational quality across economies. Factors such as low economic growth and labor-market frictions, for example, may diminish these returns. Further investigation is needed to determine whether the test scores of secondary students genuinely reflect educational quality and to understand how educational quality contributes to the accumulation of human capital in workers. We aim to collect more micro-level data and employ rigorous empirical methodologies, especially to identify the causal relationships between educational quality and cross-country variations in per-capita income or output.

This study provides new results regarding the contribution of educational quality to economic development. Discussions on improving human capital often focus on increasing school enrollment among young people and average years of schooling for adults; however, improving educational quality is also essential. Our panel data can be employed to directly appraise the role of policies that improve the quality of education and human capital in economic and social outcomes across countries and over time. We hope to contribute to future research by exploring the impact of such policies on income, human capital accumulation, and other economic and social factors.

References

Altinok, N., Angrist, N., & Patrinos, H. A. (2018). Global Data Set on Education Quality (1965–2015). Policy Research Working Paper; No. 8314, Washington, DC: World Bank,

Altinok, N., Diebolt, C., & Demeulemeester, J. L. (2014). A New International Database on Education Quality: 1965-2010. *Applied Economics,* 46 (11), 1212–47.

Altinok, N., & Murseli, H. (2007). International Database on Human Capital Quality. *Economics Letters,* 96 (2), 237–44.

Angrist, N., Djankov, S., Goldberg, P. K., & Patrinos, H. A. (2021). Measuring Human Capital Using Global Learning Data. *Nature,* 592 (7854), 403-408.

Athey, S., & Imbens, G. W. (2019). Machine Learning Methods Economists Should Know About. *Annual Review of Economics,* 11, 685–725.

Barro, R. J., & Lee, J. W. (2013). A New Data Set of Educational Attainment in the World, 1950-2010. *Journal of Development Economics,* 104, 184–98.

Barro, R. J., & Lee, J. W. (2015). *Education Matters: Global Schooling Gains from the 19th to the 21st Century*. New York: Oxford University Press.

Becker, G. S., & Chiswick, B. R. (1966). Education and the Distribution of Earnings. *American Economic Review,* 56 (1/2), 358–69.

Bertoli, S., & Stillman, S. (2019). All that Glitters is Not Gold: Wages and Education for US Immigrants. *Labour Economics,* 61, 101749.

Bils, M., & Klenow, P. J. (2000). "Does Schooling Cause Growth?." *American Economic Review,* 90 (5), 1160–83.

Borjas, G. J. (1999). The Economic Analysis of Immigration. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 3, pp. 1697-1760). Amsterdam: Elsevier.

Bratsberg, B., & Terrell, D. (2002). School Quality and Returns to Education of US Immigrants. *Economic Inquiry* 40 (2), 177–98.

Braun, H. I., & Holland, P. W. (1982). Observed-Score Test Equating: A Mathematical Analysis of Some ETS Equating Procedures. In P. W. Holland & D. B. Rubin (Eds.), *Test Equating* (pp. 9–49). New York: Academic Press.

Card, D. (1999). The Causal Effect of Education on Earnings. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 3, pp. 1801–1863). Amsterdam: Elsevier.

Card, D., & Krueger, A. B. (1992). Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy,* 100 (1), 1-40.

Caselli, F. (2005). Accounting for Cross-Country Income Differences. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of Economic Growth* (Vol. 1, pp. 679-741). Amsterdam: Elsevier.

Erosa, A., Koreshkova, T., & Restuccia, D. (2010). How Important is Human Capital? A Quantitative Theory Assessment of World Income Inequality. *Review of Economic Studies, 77* (4), 1421–49.

Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). The Next Generation of the Penn World Table. *American Economic Review* 105 (1), 3150–82, available for download at www.ggdc.net/pwt (last accessed: August 28, 2020).

Filmer, D., Rogers, H., Angrist, N., & Sabarwal, S. (2020). Learning-Adjusted Years of Schooling (LAYS): Defining a New Macro Measure of Education. *Economics of Education Review,* 77, 101971.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101(2), 171–191.

Gunderson, M., & Oreopolous, P. (2020). Returns to education in developed countries. In S. Bradley & C. Green (Eds.), *The Economics of Education (Second Edition)* (pp. 39–51). Amsterdam: Academic Press.

Hall, R. E., & Jones, C. I. (1999). Why do Some Countries Produce So Much More Output Per Worker Than Others?. *Quarterly Journal of Economics,* 114 (1), 83–116.

Hanushek, E. A., & Kimko, D. D. (2000). Schooling, Labor-Force Quality, and the Growth of Nations. *American Economic Review,* 90 (5), 1184–1208.

Hanushek, E. A., Ruhose, J., & Woessmann, L. (2017). Knowledge Capital and Aggregate Income Differences: Development Accounting for US States. *American Economic Journal: Macroeconomics*: 9 (4), 184–224.

Hanushek, E. A., Schwerdt, G., Wiederhold, S., & Woessmann, L. (2015). Returns to Skills Around the World: Evidence from PIAAC. *European Economic Review,* 73, 103–30.

Hanushek, E. A., & Woessmann, L. (2008). The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature,* 46 (3), 607–68.

Hanushek, E. A., & Woessmann, L. (2011). The economics of international differences in educational achievement. In E. A. Hanushek, S. Machin & L. Woessmann (Eds.), *Handbook of the Economics of Education* (Vol 3, 89-200). Amsterdam: North Holland.

Hanushek, E. A., & Woessmann, L. (2012). Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation. *Journal of Economic Growth,* 17 (4), 267–321.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica,* 47, 153–61.

Hendricks, L. (2002). How Important is Human Capital for Development? Evidence from Immigrant Earnings. *American Economic Review,* 92 (1), 198–219.

Hendricks, L., & Schoellman, T. (2018). Human Capital and Development Accounting: New Evidence from Wage Gains at Migration. *Quarterly Journal of Economics,* 133 (2), 665–700.

Hoekstra, M. (2020). Returns to education quality. In S. Bradley & C. Green (Eds.), *The Economics of Education (Second Edition)* (pp. 65–73). Amsterdam: Academic Press.

Islam, A., Jedwab, R., Romer, P., & Pereira, D. (2019). "Returns to Experience and the Sectoral and Spatial Allocation of Labor." https://files.webservices.illinois.edu/6984/returnstoexperienceandthesectoralallocationoflabor.pdf

Jedwab, R., Islam, A., Romer, P., & Samaniego, R. (2022). Human Capital Accumulation at Work: Estimates for the World and Implications for Development. *American Economic Journal: Macroeconomics*, forthcoming.

Jones, B. F. (2014). The Human Capital Stock: A Generalized Approach. *American Economic Review,* 104 (11), 3752–77.

Kaarsen, N. (2014). Cross-Country Differences in the Quality of Schooling. *Journal of Development Economics,* 107, 215–24.

Klenow, P. J., & Rodriguez-Clare, A. (1997). The Neoclassical Revival in Growth Economics: Has it Gone Too Far?. *NBER Macroeconomics Annual,* 12, 73–103.

Lagakos, D., Moll, B., Porzio, T., Qian, N., & Schoellman, T. (2018a). Life-Cycle Human Capital Accumulation Across Countries: Lessons from US Immigrants. *Journal of Human Capital,* 12 (2), 305–42.

Lagakos, D., Moll, B., Porzio, T., Qian, N., & Schoellman, T. (2018b). Life Cycle Wage Growth Across Countries. *Journal of Political Economy,* 126 (2), 797–849.

Lee, J. W., & Barro, R. J. (2001). Schooling Quality in a Cross-Section of Countries. *Economica,* 68 (272), 465–88.

Lee, J. W., & Lee, H. (2016). Human Capital in the Long Run. *Journal of Development Economics,* 122, 147–69.

Li, Q., & Sweetman, A. (2014). The Quality of Immigrant Source Country Educational Outcomes: Do They Matter in the Receiving Country?. *Labour Economic*s, 26, 81–93.

Lynn, R. (1982). IQ in Japan and the United States shows a growing disparity. *Nature,* 297(5863), 222-223.

Manuelli, R. E., & Seshadri, A. (2014). Human Capital and the Wealth of Nations. *American Economic Review,* 104 (9), 2736–62.

Mincer, J. A. (1974). *Schooling, Experience, and Earnings*. New York: Columbia University Press.

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives,* 31 (2), 87–106.

Mulligan, C. B. (1999). Galton versus the Human Capital Approach to Inheritance. *Journal of Political Economy,* 107 (S6), S184–S224.

OECD. (2013). *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*. Paris: OECD Publishing.

Patel, D. & Sandefur, J. (2020). A Rosetta Stone for Human Capital. CGD Working Paper 550. Washington, DC: Center for Global Development. https://www.cgdev.org/publication/rosetta-stone-human-capital

Provasnik, S., Kastberg, D., Ferraro, D., Lemanski, N., Roey, S., and Jenkins, F. (2012). Highlights From TIMSS 2011: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context (NCES 2013-009). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., & Sobek, M. (2020). IPUMS USA: Version 10.0 [dataset]. Minneapolis, MN: IPUMS, https://doi.org/10.18128/D010.V10.0 (last accessed: July 20, 2020).

Schoellman, T. (2012). Education Quality and Development Accounting. *The Review of Economic Studies,* 79 (1), 388–417.

Speakman, R., & Welch, F. (2006). Chapter 13 Using Wages to Infer School Quality. In E. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education* (Vol. 2, pp. 813-864). Amsterdam: Elsevier.

Online Appendix

Appendix A. Test Score Linking

Among several linking methods, we apply presmoothing in equipercentile linking, developed by Braun and Holland (1982).[28] This method determines a nonlinear relationship between score scales by setting equal cumulative distribution functions for two different assessments X and Y: $F = G(y)$. When the cumulative distribution functions are continuous and strictly increasing, the equipercentile linking function is solved as follows:

$$equip_Y(x) = G^{-1}[F(x)]$$

G^{-1} is the inverse of the cumulative distribution $G(y)$.

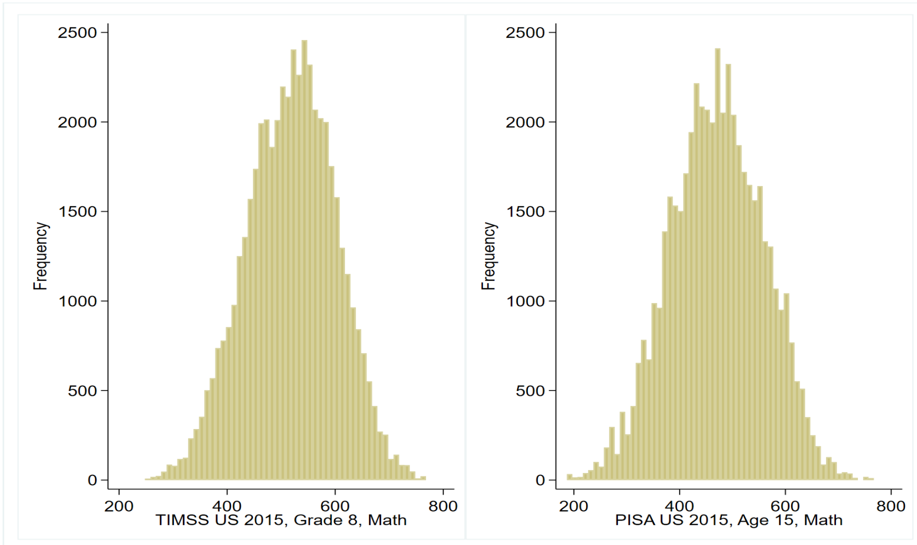
We can transform test scores on assessment X to the scores on assessment Y. For example, a student's score is 500 and in the 50th percentile on assessment X. If 550 is a student's score in the 50th percentile on assessment Y, it is equivalent to 500 on assessment X.[29]

We adopt the presmoothing method to deal with a limitation of simple equipercentile linking. When test score scales are discrete, we are unable to determine exactly matching test scores or percentiles in the sample. Using percentile ranks can deal with this limitation to some extent but not with adequate precision. Presmoothing is one of the solutions to handle this issue and allows us to interpolate test scores. It smooths the score distributions using polynomial loglinear function.

The following example shows the construction of TIMSS-equivalent PISA mathematics scores of eighth graders in the US. Figure A shows the distributions of these eighth graders' TIMSS and PISA mathematics scores. We apply "senate weights" to make the sample size equal to 50,000. The average TIMSS and PISA scores are 518.3 and 470.3. Standard deviation of PISA scores is larger than that of TIMSS scores.

[28] Kolen and Brennan (2014) and Altinok *et al*. (2018) provide details on the linking methods.
[29] The equipercentile linking methodology assumes that TIMSS and PISA measure similar domains of academic proficiency and the underlying populations are similar.

Figure A. Distributions of Eighth Graders' TIMSS and PISA Mathematics Scores in US



We apply the presmoothing method. We fit the raw score distributions using n-th order polynomials. To determine the best model, we select the model with the smallest value of Akaike information criterion (AIC). The ninth- or tenth-order polynomials are considered in our example because it gives the smallest AIC value. Table A.1 shows moments of PISA, TIMSS, and TIMSS-equivalents scores. The TIMSS-equivalent score achieves the same moments of TIMSS.

Table A.1. Moments of PISA, TIMSS, and TIMSS-equivalent Scores in US in 2015

	Mean	**Std. Dev.**	**Skewness**	**Kurtosis**	**Min.**	**Max.**
PISA	470.3	88.8	-0.03	2.75	248	762
TIMSS	518.3	82.5	-0.11	2.75	187	767
TIMSS-equivalent	518.3	82.5	-0.12	2.75	256	766

Source: Authors' calculation

Using the presmoothing result, we can determine TIMSS-equivalent PISA mathematics scores of the US eighth graders. Table A.2 shows a student with 500 points in PISA mathematics corresponds to 548 points in TIMSS.

Table A.2. TIMSS-Equivalent PISA Mathematics Scores of Eighth Graders in US

PISA	TIMSS-equivalent score
300	356.8
350	403.8
400	452.7
450	502.0
500	548.2
550	591.1
600	635.3

Source: Authors' calculation

We apply the equipercentile linking method to construct TIMSS-equivalent PISA mathematics and science scores by using TIMSS corresponding to each PISA survey year. We construct the TIMSS-equivalent PISA 2009 scores by linking PISA 2009 survey to TIMSS 2007 and 2011 surveys and combining the estimated TIMSS-equivalent scores together.

Appendix B. Use of Machine Learning Technique to Fill in Unobserved Test Scores

Machine learning technique can improve out-of-sample predictions' accuracy through flexible models and many predictors (input variables) (Mullainathan & Spiess, 2017; Athey & Imbens, 2019). It is well suitable for models with a high degree of multicollinearity. We predict test scores using a machine that utilizes information from different countries with different environments.

We use a large set of predictors to train our machine learning algorithm.[1] Among the 3,442 country-level predictors in the World Bank's World Development Indicators (World Bank, 2021a) and Education Statistics (World Bank, 2021b), we choose 769 predictors with complete data for our 92 sample countries.[2] We also use information from available actual test scores to increase prediction accuracy, following the literature on machine learning-based prediction models (Che et al., 2018; Little & Rubin, 2019). We include country-level average test score and nearest available test score[3] in a set of predictors.[4] We then implement the least absolute shrinkage and selection operator (LASSO) regression that performs both feature selection and regularization to improve the predictive accuracy of the resulting statistical model. (Tibshirani, 1996).[5] The LASSO regression has a powerful in-built feature selection capability in that the features with zero coefficient are completely disregarded for prediction. The results of the LASSO regression show that 15 features have non-zero coefficients. The selected predictors include country-level average test score, nearest available test score, percentage of male population ages 0-14 in the male population, five educational variables (percentage of female students in lower secondary education, percentage of female students in upper secondary education, percentage of students in

[1] Over the past decade, education and computer engineering researchers have adopted machine learning techniques to investigate factors predicting academic performance using large-scale student assessment data (Masci et al., 2018; Lee & Lee, 2021; Wang et al, 2022).

[2] We fill in missing observations using the average values of the available observations during that year and the past three years.

[3] The nearest available test score for the year *t* is the actual test score from its closest year. If two observations (i.e., *t-a* and *t+a*) exist, the test score from the earlier year (*t-a*) is used.

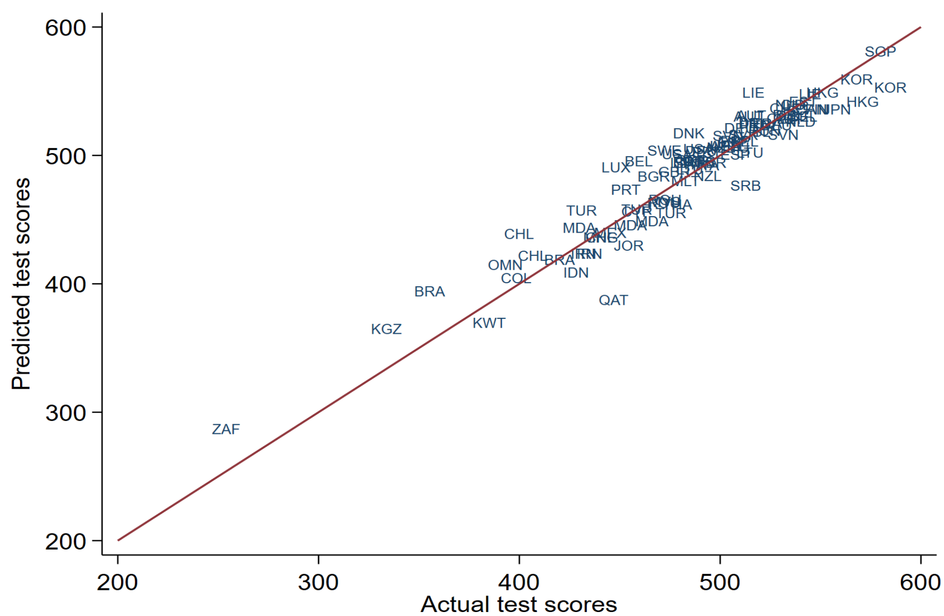
[4] We compare the predictive performance of different predictor sets and find that a set of predictors with both country-level average test score and nearest available test score performs best in terms of prediction accuracy. Even when the set of predictors include either of the two test score variables, our results reported in Section III and IV remain robust.

[5] Other machine learning techniques, such as Ridge, Elastic Net, Random Forest, and XGBoost are comparable in terms of predictive performance.

upper secondary education enrolled in vocational programmes, enrolment in tertiary education, and female enrolment in tertiary education), and seven economic indicators (industry value-added per worker, access to electricity, fixed telephone subscriptions, percentage of gross national expenditure in GDP, goods imports, CO_2 emissions from gaseous fuel consumption, and CO_2 emissions from liquid fuel consumption).

We use the LASSO regression and the 15 selected predictors to predict unobserved test scores for individual countries. We randomly split our data into training (80%) and test sets (20%). In the training set, we perform a grid search procedure using tenfold cross-validation for hyperparameter optimization and choose hyperparameters that provide the smallest Root Mean Squared Error (RMSE). Finally, we use a full dataset that includes both the training and test sets to train our machine and predict unobserved test scores. Figure B plots the actual test scores against those predicted by the LASSO regression for the observations in the test set. Two test scores are highly correlated (correlation coefficient = 0.948). The measures of out-of-sample predictive performance by the LASSO regression are satisfactory: the RMSE and R-squared (R^2) of the test set are 13.5 and 0.936, respectively. Our model must be effective for predicting unobserved test scores.

Figure B. Comparison of Actual and Predicted Scores for the Observations in the Test Set



Note: The predicted scores are estimated by the LASSO algorithm.

Appendix C. Additional Figures and Tables

Table C.1. List of 92 Countries in The Sample of Educational Quality

Country	Country code	Human capital sample (83)	US immigrants sample (76)	PIAAC sample (23)	Development and growth accounting sample (83)
OECD (34)					
Australia	AUS	O	O		O
Austria	AUT	O	O		O
Belgium	BEL	O	O	O	O
Canada	CAN	O	O		O
Chile	CHL	O	O	O	O
Czech Republic	CZE	O	O	O	O
Denmark	DNK	O	O	O	O
Estonia	EST	O			O
Finland	FIN	O	O	O	O
France	FRA	O	O	O	O
Germany	DEU	O	O		O
Greece	GRC	O	O	O	O
Hungary	HUN	O	O		O
Iceland	ISL	O	O		O
Ireland	IRL	O	O	O	O
Israel	ISR	O		O	O
Italy	ITA	O	O	O	O
Japan	JPN	O	O	O	O
Korea, Republic of	KOR	O	O	O	O
Luxembourg	LUX	O			O
Mexico	MEX	O	O	O	O
Netherlands	NLD	O	O	O	O
New Zealand	NZL	O	O		O
Norway	NOR	O	O	O	O
Poland	POL	O	O	O	O
Portugal	PRT	O	O		O
Slovak Republic	SVK	O	O	O	O
Slovenia	SVN	O		O	O
Spain	ESP	O	O	O	O
Sweden	SWE	O	O		O
Switzerland	CHE	O	O		O
Turkey	TUR	O	O		O
United Kingdom	GBR	O	O	O	O
United States	USA	O		O	O

Non-OECD (67)					
Albania	ALB	O	O		O
Algeria	DZA	O	O		O
Argentina	ARG	O	O		O
Armenia	ARM	O	O		O
Azerbaijan	AZE		O		
Bahrain	BHR	O			O
Bosnia and Herzegovina	BIH		O		
Botswana	BWA	O			O
Brazil	BRA	O	O		O
Bulgaria	BGR	O	O		O
Colombia	COL	O	O		O
Costa Rica	CRI	O	O		O
Croatia	HRV	O	O		O
Cyprus	CYP	O		O	O
Egypt, Arab Rep.	EGY	O	O		O
El Salvador	SLV	O	O		O
Georgia	GEO		O		
Ghana	GHA	O	O		O
Hong Kong SAR, China	HKG	O	O		O
Indonesia	IDN	O	O		O
Iran, Islamic Rep.	IRN	O	O		O
Jordan	JOR	O	O		O
Kazakhstan	KAZ	O	O		O
Kuwait	KWT	O	O		O
Kyrgyz Republic	KGZ	O			O
Latvia	LVA	O	O		O
Lebanon	LBN		O		
Liechtenstein	LIE				
Lithuania	LTU	O	O	O	O
Macao SAR, China	MAC	O			O
Malaysia	MYS	O	O		O
Malta	MLT	O			O
Mauritius	MUS	O			O
Moldova	MDA	O	O		O
Mongolia	MNG	O	O		O
Montenegro	MNE		O		
Morocco	MAR	O	O		O
North Macedonia	MKD		O		
Oman	OMN				
Panama	PAN	O	O		O
Peru	PER	O	O		O
Philippines	PHL	O	O		O
Qatar	QAT	O			O

Romania	ROU	O	O	O
Russian Federation	RUS	O	O	O
Saudi Arabia	SAU	O	O	O
Serbia	SRB	O	O	O
Singapore	SGP	O	O	O
South Africa, Republic of	ZAF	O	O	O
Syrian Arab Republic	SYR	O	O	O
Thailand	THA	O	O	O
Trinidad and Tobago	TTO	O	O	O
Tunisia	TUN	O	O	O
Ukraine	UKR	O	O	O
United Arab Emirates	ARE	O	O	O
Uruguay	URY	O	O	O
Vietnam	VNM	O	O	O
West Bank and Gaza	PSE			

Notes: This list displays 92 countries (regions) that have complete estimates of their educational quality.

Table C.2. Descriptive Statistics of US Immigrants' Data by Country of Origin

Country	Median Hourly wage	Average years of schooling	Average test score	N
Pooled	21.3	14.9	460.9	325177
Albania	16.2	13.6	414.7	1143
Algeria	17.3	14.9	397.5	331
Argentina	19.1	14.7	424.7	2069
Armenia	17.4	14.9	469.7	561
Australia	34.1	15.9	510.3	1661
Austria	28.2	16.6	532.7	340
Azerbaijan	20.6	15.5	458.8	182
Belgium	35.2	16.7	528.4	441
Bosnia and Herzegovina	15.3	12.4	461.0	1920
Brazil	19.0	14.4	368.0	5201
Bulgaria	20.3	15.8	512.9	1219
Canada	32.2	15.9	519.6	8137
Chile	18.0	14.7	406.5	861
Colombia	15.3	13.9	351.2	7016
Costa Rica	14.6	12.7	446.4	826
Croatia	19.6	14.0	503.0	394
Czech Republic	22.0	15.3	548.2	444
Denmark	36.1	16.3	488.1	389
Egypt, Arab Rep.	18.0	15.6	413.4	1758
El Salvador	11.2	8.7	363.5	19535
Finland	36.1	17.0	517.8	237
France	34.9	17.2	518.9	2720
Georgia	19.2	15.7	415.5	185
Germany	24.8	15.8	505.4	5318
Ghana	15.7	13.8	265.8	2152
Greece	22.3	15.9	482.6	471
Hong Kong SAR, China	26.2	14.9	506.7	1204
Hungary	21.7	15.1	550.7	506
Iceland	22.6	15.8	486.6	41
Indonesia	17.1	14.9	418.9	1080
Iran, Islamic Rep.	22.6	15.7	428.3	2330
Ireland	33.0	15.1	499.7	1714
Italy	26.6	16.0	496.5	1803
Japan	23.8	15.8	557.3	4240
Jordan	16.3	14.7	414.4	554
Kazakhstan	21.3	15.2	428.1	253
Korea, Republic of	21.2	15.9	542.5	6518
Kuwait	22.5	15.3	385.1	306
Latvia	19.9	15.1	484.8	130
Lebanon	21.4	15.0	413.1	903

Lithuania	19.0	15.4	470.6	339
Malaysia	26.3	15.0	505.4	731
Mexico	10.8	8.9	438.7	156863
Moldova	17.0	14.9	462.7	438
Mongolia	20.4	15.1	440.5	37
Montenegro	20.9	13.0	433.3	118
Morocco	15.7	13.9	333.1	1145
Netherlands	35.4	16.8	538.0	917
New Zealand	30.9	15.3	496.6	461
North Macedonia	17.2	13.4	451.8	300
Norway	36.2	16.5	505.0	227
Panama	14.9	13.8	391.1	801
Peru	14.0	13.6	354.5	4744
Philippines	19.2	14.9	373.8	25261
Poland	20.4	14.3	506.2	3961
Portugal	19.4	11.1	458.3	874
Romania	22.9	15.5	472.4	2333
Russian Federation	24.4	16.2	542.3	3639
Saudi Arabia	21.3	15.7	365.1	253
Serbia	21.8	15.4	502.8	382
Singapore	31.7	16.2	531.4	362
Slovak Republic	20.9	15.5	532.2	312
South Africa, Republic of	30.2	15.4	268.6	1403
Spain	26.7	17.0	505.2	1409
Sweden	36.2	16.6	519.1	666
Switzerland	32.7	16.8	552.4	450
Syrian Arab Republic	18.9	14.3	423.5	577
Thailand	13.8	13.2	485.7	1632
Trinidad and Tobago	17.5	13.2	434.8	1805
Tunisia	21.8	15.7	436.7	28
Turkey	22.6	15.7	432.4	1400
Ukraine	20.6	15.1	473.5	3130
United Arab Emirates	27.1	16.2	453.0	100
United Kingdom	34.1	15.8	503.4	8180
Uruguay	14.5	12.8	458.3	570
Vietnam	12.8	10.9	538.8	12236

Note: The values are the sample statistics of US immigrants by country of origin for wage, years of schooling, and average test score.

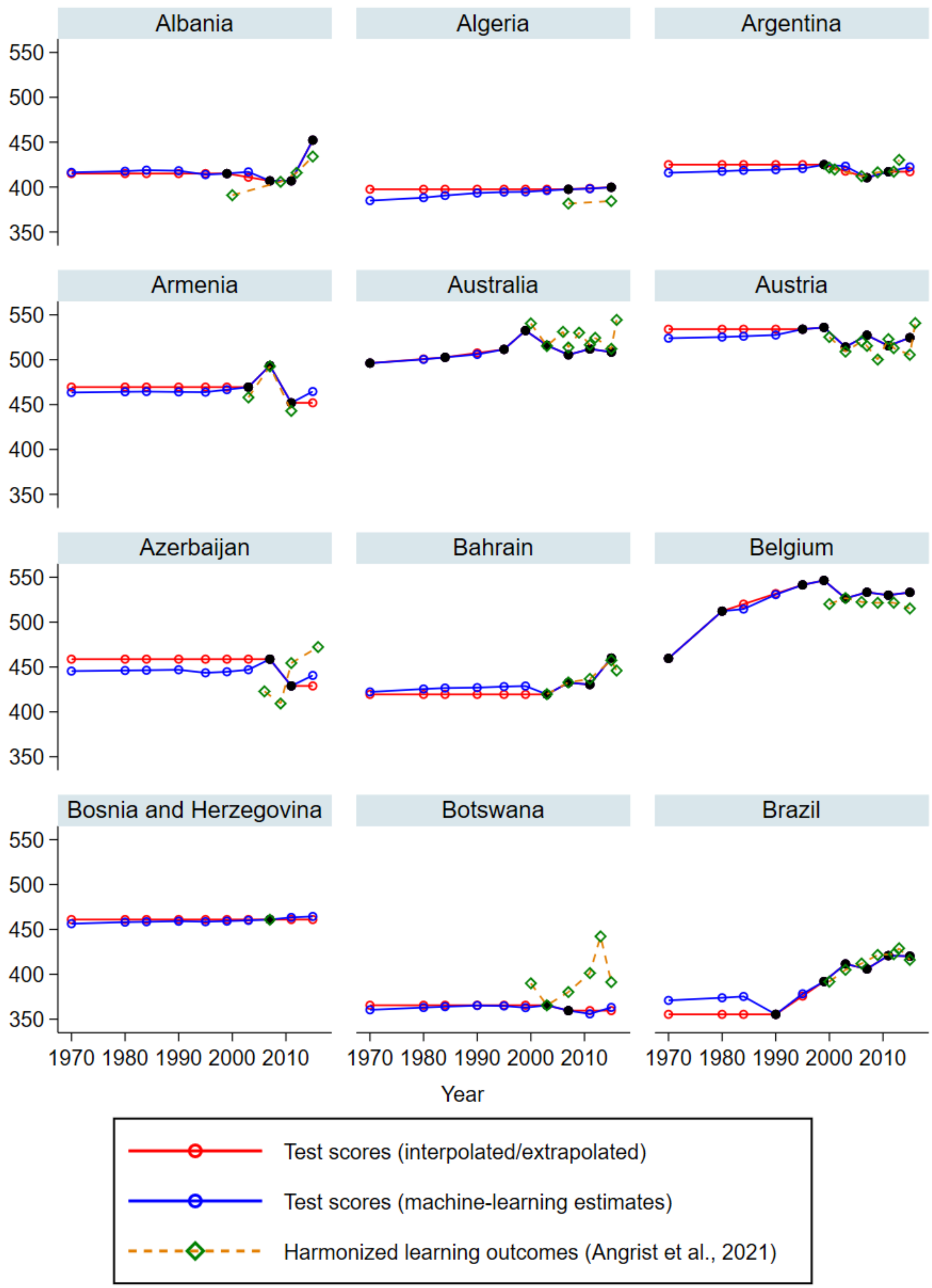
Source: Authors' calculation from the American Community Survey 2009-2013 and 2014–2018 5-year samples (Ruggles et al., 2020).

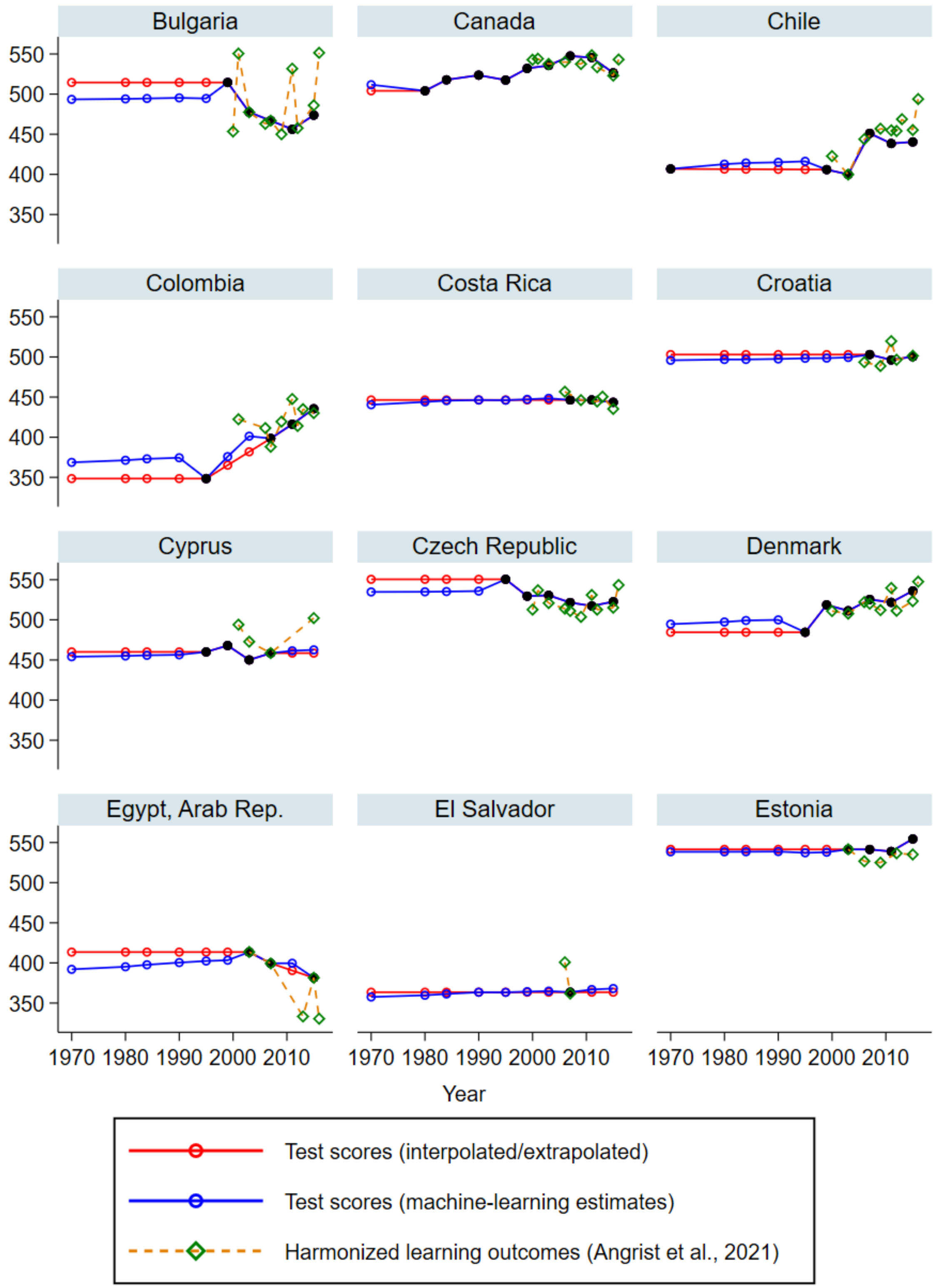
Table C.3. Wage Regression: US Immigrants Sample with All Test Scores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Baseline	Alternative sample						Additional control		Control for non-labor market participation
		Including US natives	Without Mexico	Including self-employed	Non-naturalized only	3-year buffer	Excluding imputed wages	A quadratic polynomial in home-country potential work experience	5-year cohort fixed effects	Heckman selection model
Years of schooling	0.076***	0.077***	0.077***	0.077***	0.073***	0.072***	0.079***	0.076***	0.077***	0.075***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Experience	0.035***	0.035***	0.036***	0.032***	0.038***	0.028***	0.039***	0.045***	0.025***	0.042***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)
Experience square	-0.049***	-0.048***	-0.051***	-0.042***	-0.057***	-0.033***	-0.060***	-0.061***	-0.024***	-0.064***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.006)	(0.006)	(0.005)	(0.006)	(0.005)
Average test score	0.085***	0.083***	0.085***	0.086***	0.080***	0.078***	0.074***	0.087***	0.078***	0.070***
	(0.021)	(0.021)	(0.021)	(0.022)	(0.025)	(0.024)	(0.023)	(0.000)	(0.021)	(0.021)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	325,177	4,335,027	168,313	340,094	228,332	254,991	248,081	325,177	325,177	532,685
R^2	0.322	0.322	0.318	0.292	0.333	0.332	0.361	0.323	0.323	

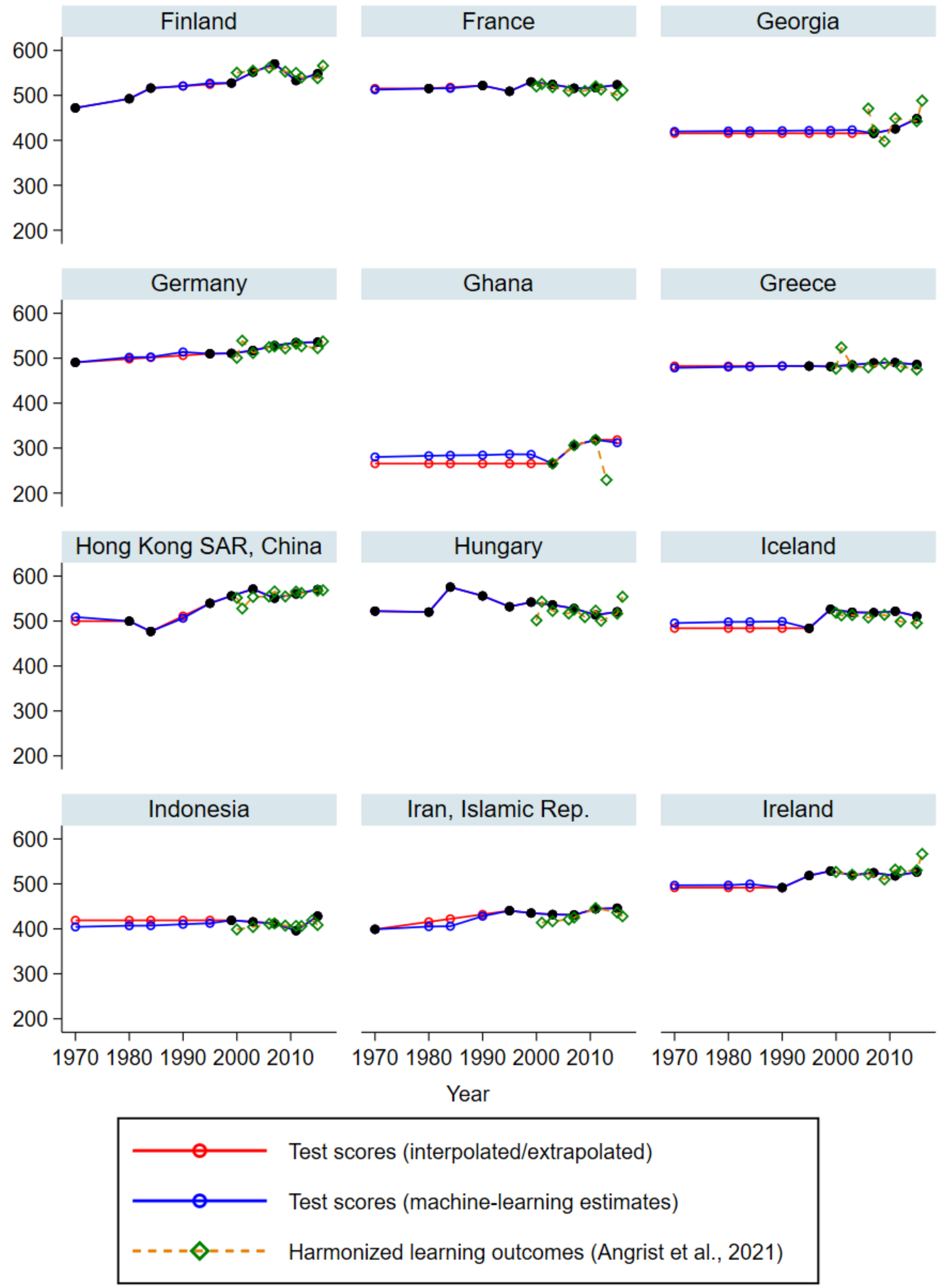
Notes: All regressions control for region of workplace, gender, disability status, workplace in a metropolitan area, self-reported English language proficiency, and year of immigration. The regressions also include country-of-origin and survey-year fixed effects. The three-year buffer excludes immigrants who entered the US less than three years after their expected graduation date. The exclusion restriction in the selection equation in the Heckman model is a variable indicating the number of children of an individual. Robust standard errors are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

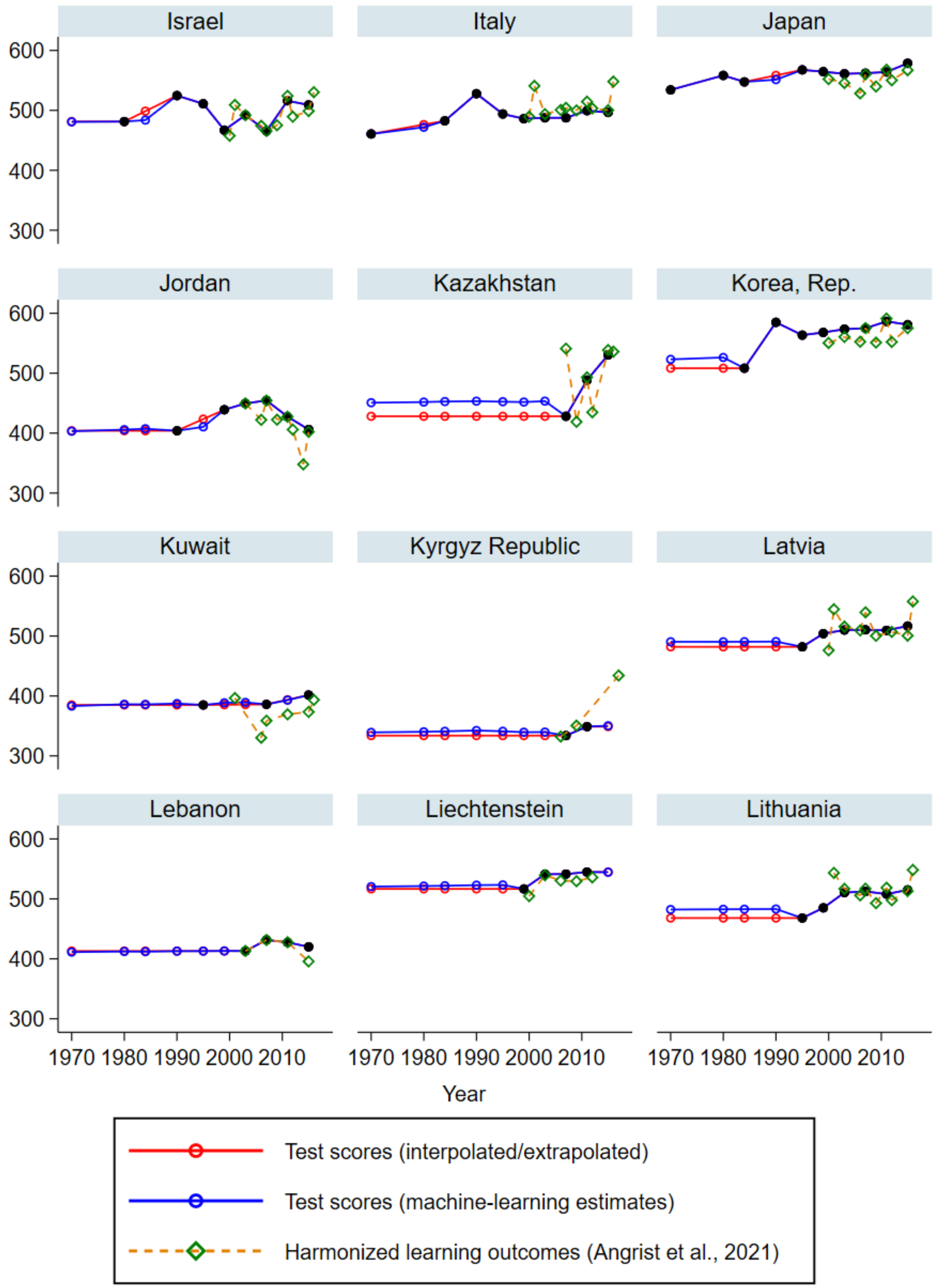
Figure C. Comparison of Educational Quality Estimates by Country, 1970–2015



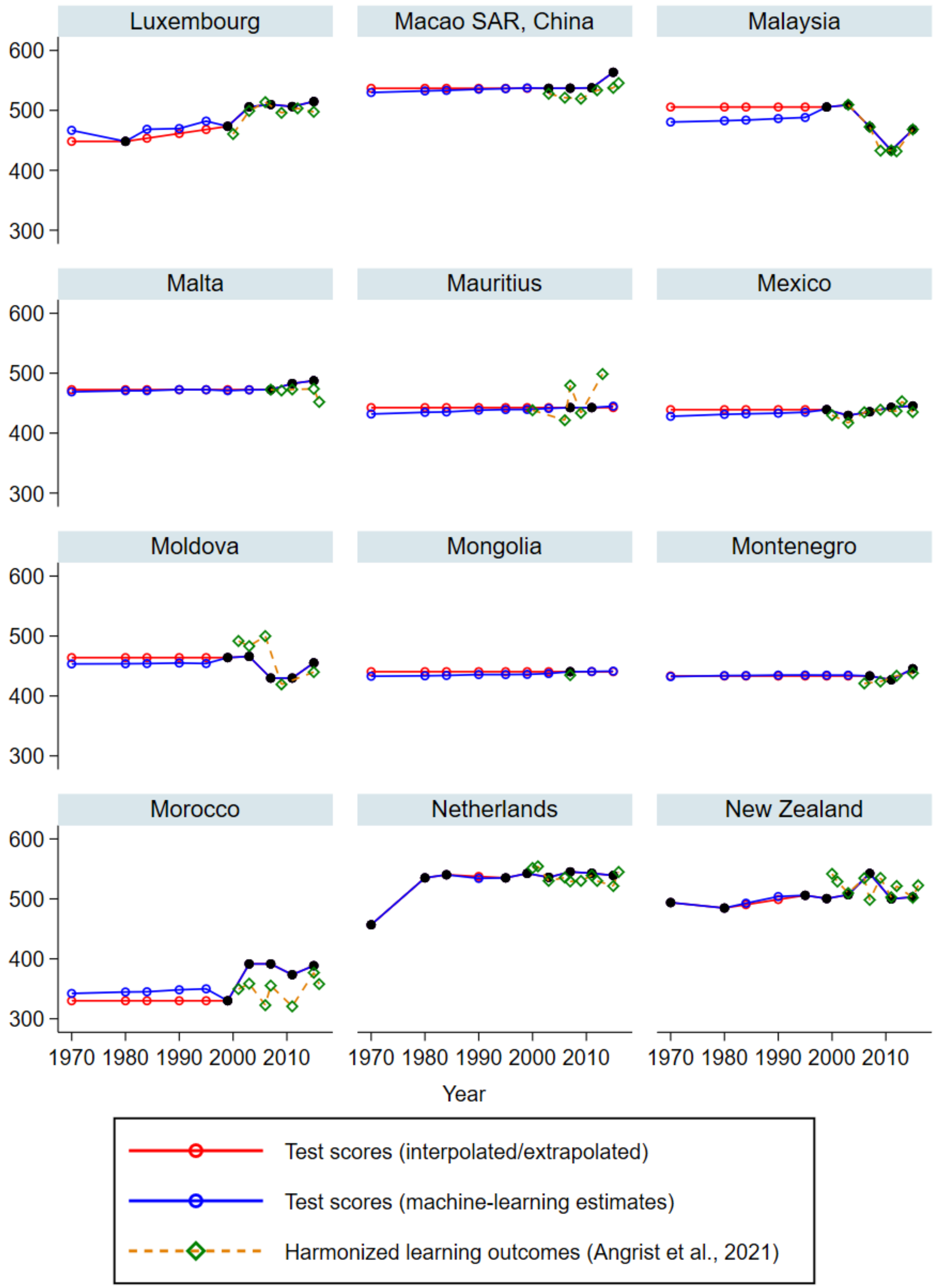


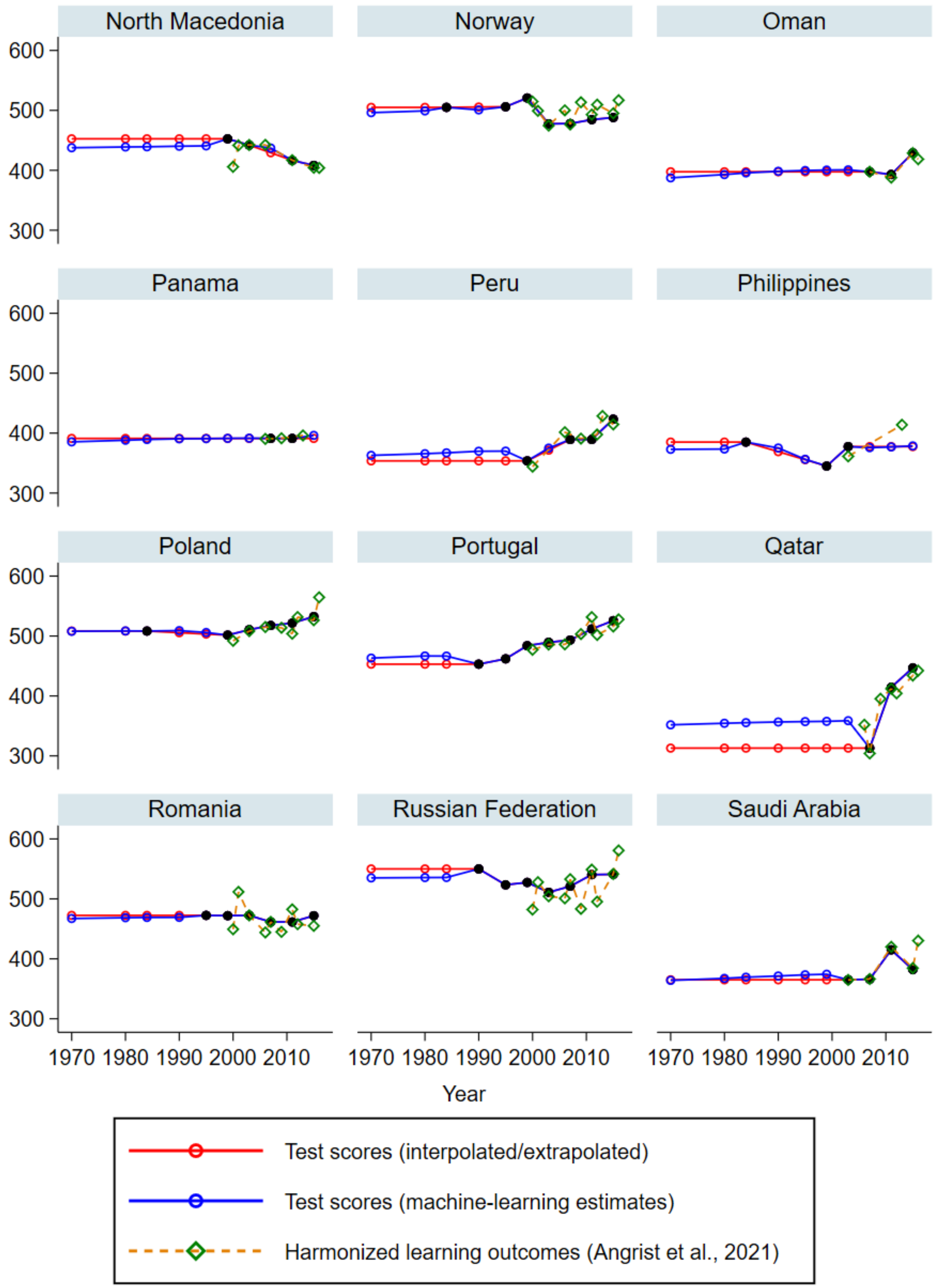
Bulgaria
Canada
Chile
Colombia
Costa Rica
Croatia
Cyprus
Czech Republic
Denmark
Egypt, Arab Rep.
El Salvador
Estonia
550
500
450
400
350
1970 1980 1990 2000 2010
Year
Test scores (interpolated/extrapolated)
Test scores (machine-learning estimates)
Harmonized learning outcomes (Angrist et al., 2021)

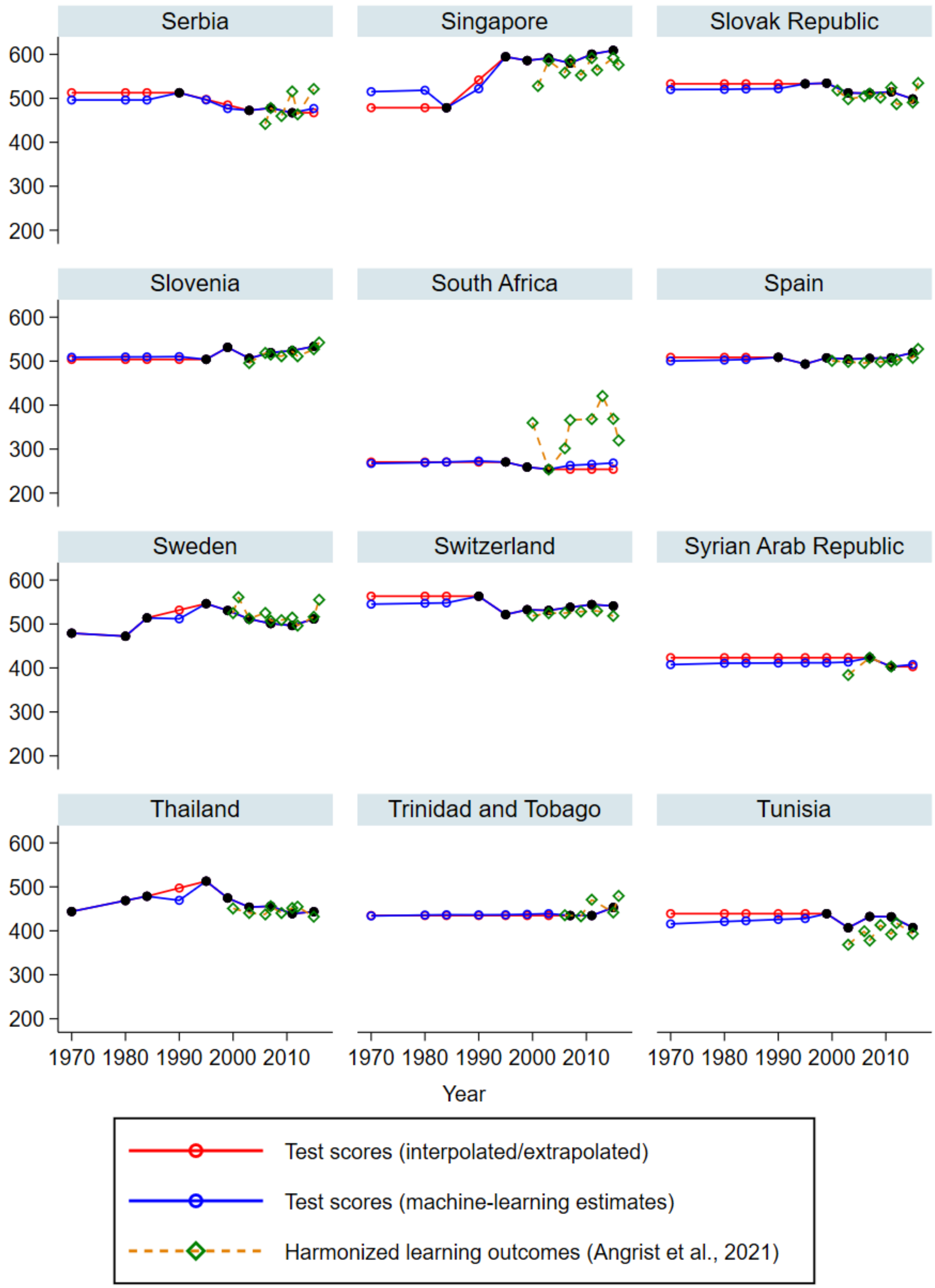

Finland
France
Georgia
Germany
Ghana
Greece
Hong Kong SAR, China
Hungary
Iceland
Indonesia
Iran, Islamic Rep.
Ireland
600
500
400
300
200
1970 1980 1990 2000 2010
Year
Test scores (interpolated/extrapolated)
Test scores (machine-learning estimates)
Harmonized learning outcomes (Angrist et al., 2021)

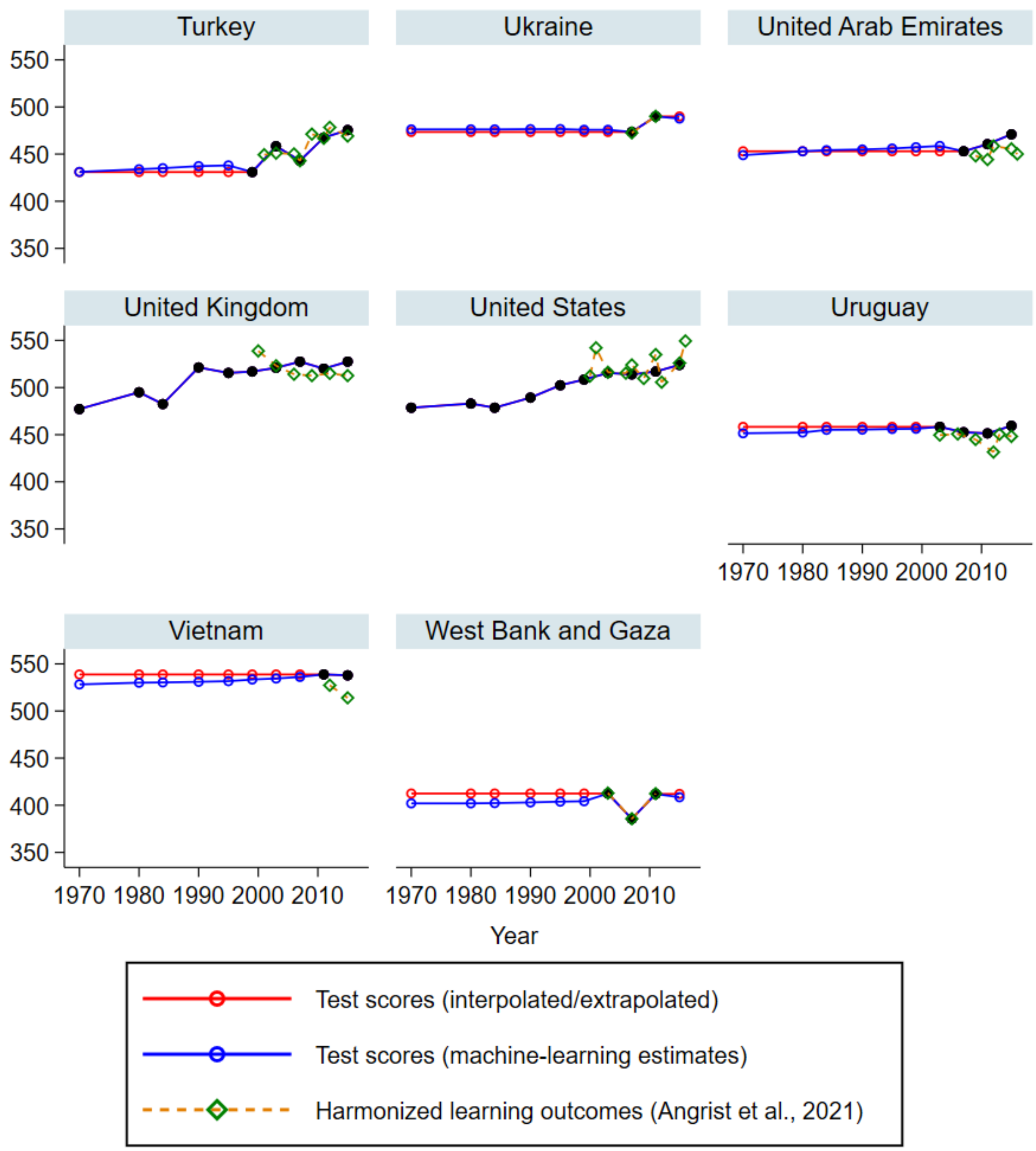


Israel
Italy
Japan
Jordan
Kazakhstan
Korea, Rep.
Kuwait
Kyrgyz Republic
Latvia
Lebanon
Liechtenstein
Lithuania
600
500
400
300
1970 1980 1990 2000 2010
Year
Test scores (interpolated/extrapolated)
Test scores (machine-learning estimates)
Harmonized learning outcomes (Angrist et al., 2021)


Luxembourg
Macao SAR, China
Malaysia
Malta
Mauritius
Mexico
Moldova
Mongolia
Montenegro
Morocco
Netherlands
New Zealand
600
500
400
300
1970 1980 1990 2000 2010
Year
Test scores (interpolated/extrapolated)
Test scores (machine-learning estimates)
Harmonized learning outcomes (Angrist et al., 2021)


North Macedonia
Norway
Oman
Panama
Peru
Philippines
Poland
Portugal
Qatar
Romania
Russian Federation
Saudi Arabia
600
500
400
300
1970 1980 1990 2000 2010
Year
Test scores (interpolated/extrapolated)
Test scores (machine-learning estimates)
Harmonized learning outcomes (Angrist et al., 2021)


Serbia
Singapore
Slovak Republic
Slovenia
South Africa
Spain
Sweden
Switzerland
Syrian Arab Republic
Thailand
Trinidad and Tobago
Tunisia
600
500
400
300
200
1970 1980 1990 2000 2010
Year
Test scores (interpolated/extrapolated)
Test scores (machine-learning estimates)
Harmonized learning outcomes (Angrist et al., 2021)



Notes: Angrist et al.'s (2021) estimates, denoted by a hollow diamond, are actual test scores in mathematics, science, and reading for both primary and secondary students. If multiple observations are available for the same year, a simple average is used. In our estimates, the actual test scores for secondary students are represented by black dots, whereas the imputed scores are depicted as hollow circles. The interpolated scores are shown in red and machine learning estimates are displayed in blue.

References

Altinok, N., Angrist, N., & Patrinos, H. A. (2018). Global Data Set on Education Quality (1965–2015). Policy Research Working Paper; No. 8314, Washington, DC: World Bank.

Athey, S., & Imbens, G. W. (2019). Machine Learning Methods Economists Should Know About. *Annual Review of Economics,* 11, 685–725.

Braun, H. I., & Holland, P. W. (1982). Observed-Score Test Equating: A Mathematical Analysis of Some ETS Equating Procedures." In P. W. Holland & D. B. Rubin (Eds.), *Test Equating* (pp. 9–49). New York: Academic Press.

Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2018). Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8(1), 1-12.

Kolen, M. J., Brennan, R. L., & Kolen, M. J. (2004). *Test Equating, Scaling, and Linking: Methods and practices*. New York: Springer Science+Business Media.

Lee, H., & Lee, J. W. (2021). Why East Asian Students Perform Better in Mathematics Than Their Peers: An Investigation Using a Machine Learning Approach. CAMA Working Paper No. 66/2021.

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (3rd edition). New York: John Wiley & Sons.

Masci, C., Johnes, G., & Agasisti, T. (2018). Student and School Performance across Countries: A Machine Learning Approach. *European Journal of Operational Research,* 269 (3), 1072-1085.

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives,* 31 (2), 87–106.

OECD. (2013). *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*. Paris: OECD Publishing.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., & Sobek, M. (2020). IPUMS USA: Version 10.0 [dataset]. Minneapolis, MN: IPUMS, https://doi.org/10.18128/D010.V10.0 (last accessed: July 20, 2020).

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological),* 58(1), 267–288.

Wang, F., King, R. B., & Leung, S. O. (2022). Why do East Asian students do so well in mathematics? A machine learning study. *International Journal of Science and Mathematics Education*, 1-21.

World Bank. (2021a). Education Statistics. (Washington, DC: World Bank). https://datacatalog.worldbank.org/dataset/education-statistics (last accessed: 16 October, 2021)

World Bank. (2021a). World Development Indicators. (Washington, DC: World Bank). https://datacatalog.worldbank.org/dataset/world-development-indicators. (last accessed: 1 April, 2021)