14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

DOI: 10.1111/joes.12530

JOURNAL OF ECONOMIC SURVEYS WILEY

ARTICLE

Cross-country data on skills and the quality of schooling: A selective survey

Ángel de la Fuente[1,2] | Rafael Doménech[3,4]

[1]FEDEA, Madrid, Spain

[2]Instituto de Análisis Económico (CSIC), Bellaterra, Spain

[3]Head of Economic Analysis, BBVA Research, Madrid, Spain

[4]Department of Economic Analysis and IEI, University of Valencia, Valencia, Spain

Correspondence
Rafael Domenech, Head of Economic Analysis, BBVA Research - Ciudad BBVA, Calle Azul, 4. Edificio La Vela, planta 5, 28050 Madrid, Spain.
Email: r.domenech@bbva.com

Funding information
Generalitat Valenciana, Grant/Award Number: PROMETEO 2016-09; Ministerio de Ciencia e Innovación, Grant/Award Numbers: CICYT PID 2020-116242RB-I00, G

Abstract

Scores in standardized international student achievement tests and some recent adult literacy studies provide interesting data on the quality of educational outputs and on the skill level of the population that can be a useful complement to the data on the quantity of schooling which have been most commonly used in the growth literature. This paper describes the most recent available primary data on the subject, reviews different attempts to organize, standardize, and summarize them, and discusses the strengths and weaknesses of the existing indicators and their potential usefulness as explanatory variables in empirical analyses of the determinants of income and welfare levels and growth rates. A final section investigates the distribution of these indicators across a sample of 21 OECD countries.

KEYWORDS
adult skills, educational quality, human capital measurement, years of schooling

JEL CLASSIFICATION
O4, I2

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2022 The Authors. *Journal of Economic Surveys* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Most of the empirical literature on human capital and growth has relied on data on the quantity of education, often measured by the average number of years of schooling of the population. It is clear, however, that years of schooling can be at best an imperfect proxy for the stock of human capital, a limitation that can bias the estimation of its effects on economic and social progress (see, e.g., Breton, 2011; Folloni & Vittadini, 2010; Wößmann, 2003, or World Bank, 2018). As schooling is not the same as learning, as stated by Pritchett (2013), focusing solely on measures of schooling is not a good idea, as it implicitly assumes that school attendance guarantees a uniformly good education. In practice, however, knowledge and skill levels will vary across countries with similar school attainments if there are differences among them in the quality of their educational systems or in the extent to which skills are built up or maintained through other channels, such as various types of post-school training and on-the-job learning. In recent years, researchers have become more keenly aware of the limitations of quantity of education variables and have paid increasing attention to the quality of education and to direct indicators of the skills and competencies of the population.

This paper reviews the available cross-country data on skill levels and educational quality and analyzes their distribution across OECD countries and their strengths and limitations in comparison to years of schooling. Section 2 describes the available primary data from standardized international assessments of student and adult competencies. Section 3 deals with different attempts to organize, standardize, and summarize these data. Section 4 discusses the strengths and weaknesses of the existing indicators and their potential usefulness as explanatory variables in empirical analyses of the determinants of income and welfare levels and growth rates. Section 5 investigates the distribution of these indicators across a sample of 21 OECD countries for which the quality of the data and the number of observations available are greater than for developing countries. Finally, Section 6 presents the main conclusion of this selective survey.

2 | PRIMARY DATA ON STUDENT ACHIEVEMENT AND ON SKILL LEVELS FROM STANDARDIZED INTERNATIONAL TESTS

To approximate the quality of education in a country, researchers have generally relied on its performance in standardized international tests that measure the knowledge or competencies of the student population, although there have also been some studies that have used estimates of Mincerian returns to schooling as quality indicators (as, e.g., Hall & Jones, 1999 or Caselli, 2005). Student achievement tests come in two varieties. The first one measures the academic achievement of students at different stages of their primary and secondary education, focusing on their mastery of standard curricula. The second set of tests is also administered to students in mandatory education but focuses on the command of the basic and applied skills that can be identified with a broad concept of literacy (and numeracy), rather than on academic achievement in a strict sense. An interesting and more recent development has been the use of general literacy tests administered to adults rather than to students. As these tests provide a direct indicator of the basic skills and competencies of the entire adult or working-age population, regardless of how these may have been acquired, in principle they are likely to be a better indicator of the general stock of human capital than measures of student achievement or competencies at a certain age or data on years of schooling.

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 International tests of student achievement administered by the IEA

Years of dat collection	*Name*	*Subject*	*Population tested*	*No. of countries in OECD21*	*Scale*	*DCLT*
1964	FIMS	math	13 yrs, FS	10, 10	pc	No
1970-71	SRC	reading	Grade 4, 8, FS	8	pc	No
1970-71	FISS	science	10, 14, FS	9, 11, 11	pc	No
1980-82	SIMS	math	8th, FS	10, 8	pc	No
1983-84	SISS	science	5th, 9th, FS	9,10, 9	pc	No
1990-91	RLS	reading	9, 14	17	IRT	No
1994-95	TIMSS 95	math	4th, 8th, FSadv	11, 8, 13, 11	IRT	Yes
1994-95	TIMSS 95	science	4th, 8th, FSadv	10, 8, 13, 11	IRT	Yes
1999	TIMSS 99	math	8th grade	7	IRT	Yes
1999	TIMSS 99	science	8th grade	7	IRT	Yes
2001	PIRLS 01	reading	4th grade	10	IRT	Yes
2003	TIMSS 03	math	4th & 8th grades	10, 9	IRT	Yes
2003	TIMSS 03	science	4th & 8th grades	9, 8	IRT	Yes
2006	PIRLS 06	reading	4th grade	14	IRT	Yes
2007	TIMSS 07	math	4th & 8th grades	13, 8	IRT	Yes
2007	TIMSS 07	science	4th & 8th grades	12, 9	IRT	Yes
2008	TIMSS 08	math	FSadv	4	IRT	Yes
2008	TIMSS 08	science	FSadv	4	IRT	Yes
2011	PIRLS 11	reading	4th grade	14	IRT	Yes
2011	TIMSS 11	math	4th & 8th grades	18, 10	IRT	Yes
2011	TIMSS 11	science	4th & 8th grades	15, 10	IRT	Yes
2015	TIMSS 15	math	4th, 8th & FSadv	18, 10, 6	IRT	Yes
2015	TIMSS 15	science	4th, 8th & FSadv	18, 10, 6	IRT	Yes
2016	PIRLS 16	reading	4th grade	18	IRT	Yes
2019	TIMSS 19	math	4th & 8th grades	19, 13	IRT	Yes
2020-21	PIRLS 21	reading	4th grade	18	IRT	Yes

Abbreviations: FS = final year of secondary schooling; FSadv = advanced test, administered in the final year of secondary education. DCLT, directly comparable with latter tests; FIMS, first international mathematics study; FISS, First International Science Study; IRT, scoring based on item response theory; pc, percent correct; RLS, reading literacy study; SIMS, Second International Mathematics Study; SISS, Second International Science Study; SRC, study of reading comprehension.
Sources: Hanushek and Wößmann (2015), Altinok et al. (2018) and IEA (https://www.iea.nl/studies).

Tables 1–3 summarize the most relevant achievement and literacy tests that have been administered in relatively broad samples of countries during the last several decades. The three tables have a common structure. For each test wave and subject, the table indicates the population being tested as characterized either by its age or by the school grade they are in and the number of countries in our reference sample of 21 OECD[1] member states that participated in the test. The last two columns indicate the scoring scale being used (either percent correct or IRT) and whether or not scores are directly comparable with those in more recent waves of the same test *(DCLT)*. The relevant data are collected in an Excel file that is available with the paper.

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 International tests of student literacy administered by the OECD (PISA)

Years of data collection	*Name*	*Subject*	*Population tested*	*No. of countries in OECD21*	*Scale*	*DCLT*
2000-02	PISA 2000	Reading	15 yrs	19	IRT	yes
2000-02	PISA 2000	math	15 yrs	20	IRT	no
2000-02	PISA 2000	Science	15 yrs	20	IRT	no
2003	PISA 2003	reading & math	15 yrs	20	IRT	yes
2003	PISA 2003	Science	15 yrs	20	IRT	no
2006	PISA 2006	reading, math & science	15 yrs	20	IRT	yes
2009	PISA 2009*	reading, math & science	15 yrs	20	IRT	yes
2012	PISA 2012	reading, math & science	15 yrs	21	IRT	yes
2015	PISA 2015	reading, math & science	15 yrs	21	IRT	yes
2018	PISA 2018	reading, math & science	15 yrs	20, 21, 21	IRT	yes

Sources: OECD, https://www.oecd.org/pisa/. *Ten additional emerging economies participated in the PISA 2009 study on a reduced and delayed timeline, known as the PISA 2009+ project, which was administered in 2010. Their results were published by Walker (2011).

TABLE 3 International adult literacy tests administered by the OECD

Years of data collection	*Name*	*Subject*	*Population tested*	*No. of countries in OECD21*	*Scale*	*DCLT*
1994-1998	IALS	reading literacy	16-65	15	IRT	yes
1994-1998	IALS	quantitative literacy	16-65	15	IRT	no
2003-08	ALLS	reading literacy	16-65	8	IRT	yes
2003-08	ALLS	numeracy	16-65	8	IRT	yes
2013, 14-15	PIACC	reading literacy	16-65	17 + 2	IRT	yes
2013, 14-15	PIACC	numeracy	16-65	17 + 2	IRT	yes
2014-15	PIACC	ICT skills	16-65	14 +2	IRT	

Abbreviations: *ALLS*, Adult Literacy and Lifeskills Survey; final sec. = final year of upper secondary schooling; *IALS*, International Adult Literacy Survey; *PIACC* = Program for the International Assessment of Adult Competencies.
Source: OECD, http://www.oecd.org/.

2.1 | Tests of student academic achievement

Table 1 lists the tests of student academic achievement that have been conducted by the International Association for the Evaluation of Educational Achievement (IEA). These tests seek to measure student achievement in three key areas (mathematics, science, and reading) at up to three different stages in their education: primary school and lower and upper secondary school (4th and 8th grade and the final year of upper secondary education, *FS*). The last of these tests has

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

a special version, known as *TIMSS advanced* (*FSadv* in the table), that is administered to upper secondary students who are enrolled in advanced mathematics and physics programs or tracks.

The current generation of IEA's math and science tests goes under the name of TIMSS (for Trends in International Mathematics and Science Study). TIMSS started in 1995 (as the Third International Mathematics and Science Study) and has been administered every four years since then. As quite a few other international tests, TIMSS is scored using a methodology based on what is known as *Item Response Theory* (IRT) that takes into account the revealed difficulty of different test items.[2] For the first edition of TIMSS, the grading scale was normalized to have a mean of 500 and a standard deviation of 100, which corresponded to the overall achievement distribution across all countries that participated in the test, assigning an equal weight to all of them. To allow TIMSS scores to be comparable over time within each subject, subsequent editions of the test retain a sufficient number of items from previous waves and the grading scale remains "constant," not being renormalized to a mean of 500 each year.

The situation is the same for the current version of IEA's reading test, known as PIRLS (for Progress in International Reading Literacy Study), which is aimed at 4th grade students. PIRLS has been administered every 5 years starting in 2001. The grading scale is similar to the one used in TIMSS, with an original mean of 500 and standard deviation of 100 that correspond to the first edition and fully comparable scores for latter editions. Earlier versions of the IEA tests did not have a fixed periodicity but covered the same subjects as TIMSS and PIRLS at irregular intervals. With the exception of RLS (the Reading Literacy Study), these tests reported scores simply as the percentage of correct answers.

In addition to those run by IEA, there have been a number of other international assessments of student performance. The Early Grade Reading Assessment (EGRA) is a reading test developed by the Research Triangle Institute (RTI) that has been measuring literacy skills since 2006 in over 65 countries, most often in grades 2–4, through a 15-min individual oral assessment of five fundamental reading skills. The reading comprehension indicator in EGRA has been used by Angrist et al. (2021) to construct their database.

One of the most ambitious of these assessments has been the MLA project (*Monitoring Learning Achievement*), organized by Unesco and Unicef, which covered over 70 countries, mostly LDCs (see Chinapah, 2003). There have also been a number of regional assessments. Thirteen Latin American countries have joined UNESCO's *Laboratorio Latinoamericano para la Evaluación de la Calidad de la Educación* (*LLECE*), conducted in three waves in 1997, 2006, and 2013 (see UNESCO-OREALC, 2016). Similarly, 14 Anglophone countries in Africa have participated in the *South and Eastern African Consortium for Monitoring* Educational Quality (SACMEQ, http://www.sacmeq.org), with waves in 1995, 2000, and 2007, and 22 Francophone countries in Africa and Asia in at least one of the waves of the Programme *d'Analyse des Systèmes Educatifs* de la CONFEMEN (*PASEC*, http://www.pasec.confemen.org), in which started in 1995 (see Michaelowa, 2001).

2.2 | Tests of student literacy and numeracy

Table 2 lists another family of tests, known as the PISA studies, that have been administered by the OECD every three years starting in 2000. PISA tests 15-year-old students just prior to the completion of mandatory schooling. The focus is not so much on academic achievement per se as on the command of the basic and applied skills that can be identified with a broad concept of literacy in the same three areas of interest as IEA tests (math, reading, and science). In each wave of PISA, one of these three subjects is selected for a more in-depth analysis.

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

All PISA tests are scored using an IRT proficiency scale. Not all of them are directly comparable with latter studies, however. In particular, the reference scale for each subject has been set the first time the subject was the main focus of study. Hence, all reading tests are comparable because reading was the main subject of the first edition of PISA, but math scores are only directly comparable from 2003 onward and science scores from 2006 onward, after the first full assessment of each subject.

2.3 | Tests of adult literacy

The final group of tests, listed in Table 3, are also literacy tests conducted by the OECD, but aimed now at the entire working-age population rather than at young people currently enrolled in school. Three successive studies (IALS, ALLS, and PIAAC) have been conducted by the OECD until now. All of them have tested reading and quantitative literacy while ALLS and PIACC also try to measure problem-solving abilities (not shown in the table). All three tests are scored using an IRT proficiency scale with a range from 0 to 500. As noted in OECD (2009), some results are comparable across tests. In particular, reading or literacy scores are directly comparable across all three of the surveys (after averaging prose and document literacy to obtain a single literacy score in the case of IALS and ALLS). Quantitative literacy scores from IALS are not directly comparable with numeracy scores from ALL and PIAAC (which are, however, directly comparable with each other) because the concept of numeracy used in the two more recent surveys is broader than the concept of quantitative literacy used in the earlier one. The last wave of PIAAC, finally, incorporates a section on ICT skills.

The World Bank's Skills Towards Employment and Productivity (STEP) program provides measures of reading skills on a similar scale as PIAAC, but for 17 non-OECD low and middle-income countries between 2012 and 2017 (see Gaëlle et al., 2014). As shown by Keslair and Paccagnella (2020), however, there are important differences between PIAAC and STEP in the way data are collected and proficiency is measured. Despite the many similarities between the two studies, these differences limit their usefulness for cross-country comparisons.

While these data are of considerable interest because they provide the only available cross-country information on the skill level of the adult population, the short history of adult skill assessments is an important drawback. On the other hand, PIAAC sample sizes (over 5000 per country) are sufficiently large to allow us to disaggregate the results by age group with some guarantee of representativeness and may therefore be used to construct synthetic time series of scores, as has been done by Coulombe and Tremblay (2006) using IALS data and by Schwerdt and Wiederhold (2018) with PIAAC.

3 | SUMMARY MEASURES OF SCHOOLING QUALITY AND TOTAL HUMAN CAPITAL

Several groups of researchers have collected and homogenized the results of international student tests and have used them to construct summary performance measures that are usually interpreted as indicators of the quality of national educational systems or the level of skill of the labor force. Among the most influential studies in this line of work are those of Eric Hanushek, Ludger Wößmann and different coauthors, and those of Nadir Altinok, Noam Angrist and other researchers linked to the World Bank. Going one step further, some recent papers have begun

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

to explore ways to construct indicators of the stock of human capital that combine quantity and quality variables.

Hanushek and Kimko (H&K 2000) construct an indicator of labor force quality for a sample of 31 (mostly advanced) countries using mean national scores in a number of international achievement tests in mathematics and science spread over several decades.[3] To approximate the average quality of the labor force (rather than that of current students), H&K combine all the scores available for each country up until 1991 into a single cross-section indicator that is constructed as a weighted average of the normalized values of such scores (where the weights are based on the inverses of the country specific standard errors of the scores). They use two alternative normalization procedures to produce two different (but highly correlated) measures of labor force quality that they denote by *QL1* and *QL2*. In the first case *(QL1)*, the average world score in each year (measured by the percentage of correct answers) is normalized to 50. This procedure implicitly assumes that average performance does not vary over time. In the second case *(QL2)*, they allow average performance to drift over time reflecting average US scores in a different but comparable set of national tests (NAEP).[4]

Hanushek and Wößmann (H&W, 2012 and 2015) construct a refined version of *QL2* for a sample of 64 countries (extended to 77 in the second study) using data from different tests of math and science conducted between 1964 and 2003 that include the first two waves of PISA. They standardize test scores prior to averaging them in order to put them all on the same distribution as the 2000 PISA test, with an overall mean of 500 for the OECD and an individual-level standard deviation of 100 for the same sample. In addition to average standardized scores (across assessments) for each country, they also report data on the average share of students that reach the thresholds for "basic" and "superior" performance, set at one standard deviation above and below the OECD average.

Standardized test scores are constructed as follows. First, the authors reconstruct the time path of absolute US performance starting from this country's results in PISA 2000 and going backward with the help of NAEP data. For each test conducted at time t, on subject s for age group a, an absolute normalized score for the United States is calculated as

$$I_{ast}^{US} = O_{s,2000}^{US,PISA} + \frac{NAEP_{ast}^{US} - NAEP_{as1999}^{US}}{\overline{SD}_{as}^{US,NAEP}} SD_{s,2000}^{US,PISA} \tag{1}$$

where $O_{s,2000}^{US,PISA}$ is the original score of the United States in PISA 2000 in subject s, NAEP is the age-, subject-, and time-specific NAEP test score,[5] $\overline{SD}_{as}^{US,NAEP}$ is the age- and subject specific standard deviation of the US NAEP test scores across individuals, calculated by averaging the available observations on standard deviations during the relevant period, and $SD_{s,2000}^{US,PISA}$ is the subject-specific standard deviation of US students on the PISA 2000 test. Hence, changes in NAEP scores over time, relative to the 1999 edition of the test (the one closest to the PISA test used as a benchmark) are scaled up or down taking into account the difference in standard deviations of US individual scores across the two tests.

Next, other countries' normalized scores are calculated, taking into account their respective positions in relation to the United States. For each country i, we have:

$$I_{ast}^{i} = I_{ast}^{US} + \frac{O_{ast}^{i} - O_{ast}^{US}}{SD_{ast}^{OSG}} SD_{s,2000}^{OSG,PISA} \tag{2}$$

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where O^i_{ast} denotes country i's original score for each subject and age group in the test conducted at time t. Differences in original scores between each country and the United States are adjusted taking into account the cross-country variance in mean scores across a group of 13 advanced OECD countries (labeled by H&W as the OECD standardization group, OSG) that have participated in many of the relevant tests.

In a series of papers, Nadir Altinok, Noam Angrist, and various coauthors also construct and extend a database of standardized student achievement measures for a large number of countries. They augment H&W's sample by incorporating data on reading assessments that H&W's main indicator of cognitive skills disregards and information from other sources, such as regional achievement studies for countries that do not participate in global achievement studies and reading assessments. One of the latest versions of this database (AAP, 2018) provides data for 163 countries covering (unevenly) the period 1965–2015 at 5-year intervals. Angrist et al (ADGP, 2021) add an additional country to the sample and change the sample period to 2000–2017.[6]

Unlike H&K or H&W, AAP provide panel data with several observations for most countries that correspond to what they call *harmonized learning outcomes* (HLOs). HLOs are constructed as averages taken over different tests administered in the same or nearby years, after adjusting their results for differences in difficulty. In addition to mean scores (overall and disaggregated by educational level, subject, gender, and other characteristics), they also report data on the percentage of students who reach three different benchmark levels (minimum, intermediate, and advanced), thus providing useful information on the distribution of skills. Average country scores in different assessments at each point in time are standardized and brought into a common scale by using a procedure the authors refer to as *pseudo-linear linking*. A similar procedure is used to homogenize results over time, using NAEP data for the United States as an anchor, as in H&W.

The standardization procedure essentially involves using the average scores obtained in each test by the set of countries that participate in both of them in order to calculate an "exchange rate" that can be used to adjust for differences in difficulty and grading scales. That is, given two tests X and Y, the score of country i in test X, χ_i is converted to the scale of test Y using

$$y_i = x_i e \tag{3}$$

with

$$e = \frac{\mu(y)}{\mu(x)} = \frac{\frac{1}{n}\sum_{i\in X\cap Y} y_i}{\frac{1}{n}\sum_{i\in X\cap Y} x_i} \tag{4}$$

where the average scores for the two tests, $\mu(x)$ and $\mu(y)$, are calculated over the n countries that have participated in both of them (i.e., over all $i \in X \cap Y$). When correcting for differences in difficulty over time, the exchange rate is based on US performance in international assessments and on NAEP.[7]

In the last few years, researchers have begun to work on ways to combine quantity and quality data to construct indicators of the "total" stock of human capital. Filmer et al. (2020) propose combining HLO's with data on average years of schooling to construct quality-adjusted or learning-adjusted years of schooling *(LAYS)*. In particular, *LAYS* for country i are constructed as

$$LAYS_i = YRSCH_i \times Q_i^b \tag{5}$$

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where *YRSCH* is the average years of schooling of the population cohort and Q_i^b a measure of quality, relative to a benchmark level b.[8] This benchmark may correspond to the top performing country or group of countries or to some other convenient reference level for good performance, for example, a TIMSS score of 625 which corresponds to the threshold for advanced attainment set by TIMSS. While this measure depends in principle on the details of the grading scale, the specific test and subject chosen and the choice of benchmark level, the authors check that in practice the results do not change qualitatively with these factors. They report that correcting for quality increases cross-country differences, as countries where attainment levels are low also tend to display poor performance in international assessments of student achievement. Glawe and Wagner (2022) have constructed an extended database of *LAYS* covering the period 1995–2015 for 65 countries, using quinquennial data on average years of schooling of 25–29-year olds from Barro and Lee (2013) and the Cohen-Soto-Leker dataset, and the Grade 8 TIMSS mathematics learning assessment results. As shown by the authors, the correlation between LAYS and mean school years of the 25–29 year olds is 0.93. Similarly, Kaarsen (2014) uses the variation in the results of achievement tests associated with an additional year of schooling in each country to adjust years of schooling for quality.

Égert et al. (2022) construct a new measure of human capital by combining past student achievement scores with data on the average years of schooling of the current population. Past student scores, measured in logs, are weighted by the share of the corresponding cohorts in the current population to construct a quality indicator which is then combined with years of schooling in a Cobb–Douglas function. The elasticities of the quality and quantity variables in this function are estimated with PIAAC data disaggregated by country and cohort through an auxiliary regression that relates average PIAAC scores in each cohort-country pair on the corresponding average PISA score and average schooling (apparently at the country level for all cohorts). Due to the limited availability of student achievement scores (even after projecting PISA scores backward with data from other assessments), estimates of the human capital stock can only be constructed for 15 OECD countries starting in the mid-2000s. For a larger sample of 54 countries it is possible to approximate the stock of human capital only for the younger subset of the working age population, in particular those with ages between 16 and 39 years.

Taking a different route, Schoellman (2012) and Botev et al (2019) correct for quality using estimates of mincerian returns to schooling, that is, the average wage increase linked to an additional year of education. The first of these studies uses data on immigrants to the United States who have been educated in their country of origin, while the second one uses estimates of standard wage equations with data for the resident population, including migrants. An alternative proposal is the one used in the Penn World Table since its version 8 (see Feenstra et al., 2015). The human capital index is computed using the average years of schooling (from Barro & Lee, 2013; Cohen & Leker, 2014; de la Fuente & Doménech, 2006, 2015; Lee & Barro, 2001) and an estimated rate of return to education, based on Mincer equation estimates around the world (from Psacharopoulos, 1994). An extension of this approach has been used by Angrist et al. (2019), who also take into account an indicator of student learning or the quality of schooling.[9]

4 | INDICATORS OF EDUCATIONAL QUALITY: POTENTIAL USES AND LIMITATIONS IN GROWTH STUDIES

The use of data from international achievement test to construct indicators of educational quality or cognitive skills is certainly a relevant development that can help give us a better picture of cross-country stocks of human capital. Such data, however, have important limitations that should be kept in mind. An obvious one is the lack of quality indicators for tertiary education.[10] Even more important is the relative scarcity of comparable cross-country data on student performance. Many countries have participated only in one or a few international assessments, mostly in recent years. As a result, there are only a few countries for which we have relatively long time series. The problem is compounded by the fact that most of these assessments measure the performance of students enrolled in primary or secondary schools, who have not yet entered the labor market. Hence, we are quite far from having the information that would be necessary to approximate, working with cohort data, the average quality of the human capital embodied in the labor force of most countries – even for recent years, and much more so as we go back in time. An additional worry that arises when we try to go beyond the richest countries is that, certainly in past decades but even today, schooling is far from being universal in many countries. For countries with low enrollment rates, the results of the student assessments described above measure the knowledge and competencies of only a relatively small part of each cohort whose weight has likely been rising over time.

The following calculations may help highlight the importance of the problems posed by the scarcity of student test data for their use in empirical studies of growth performance.[11] Both PISA and IEA assessments are generally conducted with students between 10 and 16 years of age. Assuming the average individual remains in the labor force for 45 years, between ages 20 and 65, in order to approximate the average quality of the labor force in 2020, we would need test data for all those who entered the labor force during the previous 45 years, that is, between 1975 and 2020. These cohorts were tested between 5 and 10 years earlier, that is, between 1965 and 1970. Since the earliest assessments we have were conducted in those years (in a handful of countries), all the test data we have accumulated to date would only allow us to approximate the skill level of today's labor force in a few countries – but certainly not its average quality over the last several decades, which is the variable that should be included in many of the growth regressions that have been run in the literature.

Hence, the available data on student performance is clearly insufficient to construct time series of stock measures of average skill for the labor force, but they can still be quite useful as a flow measure of investment in quality at each point in time. To exploit these data in growth studies, we need to use empirical specifications that are suitable for flow data. One possibility that has been used in the literature to get around similar problems regarding other growth determinants is the specification developed by Mankiw et al. (MRW, 1992) as a log-linear approximation around the steady state of a generalized Solow model. This approach may be particularly useful in combination with pooled data at relatively high frequencies as a way to exploit the time variation in the data.

Another important limitation of using student test data in empirical growth equations is the high potential for severe endogeneity and reverse causation problems. Economic growth generates increased public and private resources that may be used by governments and families to finance higher quality educational systems that yield improvements in student performance, as well as more years of schooling. If we focus on flow measures, such as test scores (or enrollment rates), the feedback effect from growth to education can be quite rapid. However, higher test scores will only affect growth much further into the future when today's students enter the labor market and become employed. As a result, reverse causation is likely to be an important problem even in data at relatively high frequencies. On the other hand, when we rely on data on average years of schooling of the adult population, the direct effect of schooling on growth should be immediate, while feedback effects from growth to increased average schooling will involve much longer lags.

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

As a result, average schooling levels can be considered as a predetermined variable (except over rather long periods) and reverse causation should be much less of a problem.

The problems discussed above do not arise in literacy assessments of the entire adult population, such as PIAAC, but in this case we only have very recent results for relatively few countries. As we have seen, a possible way to mitigate the problem this represents is to construct synthetic time series of adult competencies using the age distribution of the microdata of these tests, as has been done by Coulombe and Tremblay (2006) using IALS data and Schwerdt and Wiederhold (2018) with PIAAC. In this regard, Castelló-Climent (2019) has shown that the largest effects on economic growth are linked to the human capital stock of the population aged 40–49 years for a large sample of 146 countries. That is, the population in the middle of its professional career is the most representative cohort of the working-age population in terms of productivity. This result implies that the most relevant variable for explaining today's growth is the quality of schooling between 30 and 35 years ago, when this central cohort was in primary and secondary school.

Synthetic time series, however, raise several complications that have to do with the fact that the scores of the different cohorts are measured at a single point in time, which corresponds to a different age for each of them. This may introduce a bias if, for instance, skills change significantly over time (due for instance to the accumulation of experience and then to aging and depreciation), or if survival or migration rates are correlated with skill levels, as seems likely.

More generally, none of the existing assessments can cover and measure correctly all the competencies and skills that determine the productivity of a country's labor force or its capacity to innovate. These include those acquired in universities and workplaces or the specialized and highly complex knowledge and skills of scientists and high-level technicians that go well beyond what these tests measure. For all these reasons, it is important to exercise caution when using and interpreting existing indicators of educational quality and of the skill level of the overall population. In countries with high enrollment rates where we have relatively long series of test results and these remain stable over time, we may perhaps be somewhat confident that student achievement data can give us some idea of the average quality of the school system, which is certainly an important input for growth, but not the only one. In countries with low enrolment rates, shorter series or where test results have changed significantly over time, we have to be even more cautious. As for adult literacy data, we must keep in mind that they pick up only a (possibly small) fraction of the relevant knowledge and skills.

Finally, it seems obvious that both the quantity and the quality of education must be taken into account in order to correctly approximate the stock of human capital, as both are essential inputs in its production. Keeping a large fraction of the population in school for many years may be a bad investment if the poor quality of education prevents students from acquiring the skills the productive system demands. On the contrary, an excellent educational system will have only a limited effect on productivity if it excludes most of the population. Unless we have good direct measures of the relevant knowledge and skills of the entire population – which we surely do not, except possibly for very recent years – we need to measure as well as possible both dimensions of the stock of educational capital and use them jointly in empirical analysis. A promising possibility in this line consists in adjusting years of schooling for quality, as has been done in some recent studies using alternative procedures (see for instance, Angrist et al., 2019; Filmer et al., 2020; Glawe & Wagner, 2022, or Égert et al., 2022), although the scarcity of data implies that stock quality measures will generally have no time variation, which will limit their usefulness in empirical analyses and other applications.

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 Selected normalized education and income indicators around 2010

	[1]	[2] 2012	[3] 2011	[4] 2010	[5]	[6]	[7] 2010	[8] 2010	[9] 2005–2015
	Adult skills	Student skills	Student achiev	HLO	H&W	HLO avge till 2005	Years of schooling	ypc 15-64	avge welfare
Australia	102.1	101.6	98.4	98.7	102.5	95.3	105.9	110.7	109.1
Austria	101.4	99.2	100.1	99.2	102.4	105.4	101.4	105.1	112.1
Belgium	103.5	101.0	99.8	103.4	101.5	107.0	95.9	101.1	103.9
Canada	100.4	103.5	100.8	100.4	101.4	101.4	112.9	103.5	104.5
Denmark	102.3	98.7	103.3	101.9	99.9	106.2	103.1	99.8	101.7
Finland	106.1	104.9	101.4	102.1	103.2	100.7	102.6	96.9	102.5
France	96.2	99.0	99.6	100.0	101.4	98.0	101.0	92.2	106.4
Germany	100.8	102.2	101.9	102.4	99.7	95.4	103.8	102.0	103.0
Greece	94.2	92.3	96.5	93.3	92.7	94.8	86.0	72.9	69.1
Ireland	97.2	102.2	99.9	100.3	100.5	102.9	98.5	106.2	84.1
Italy	92.7	97.1	98.5	96.9	95.8	94.1	84.9	83.4	98.0
Japan	108.8	107.1	108.8	111.4	106.9	112.8	105.6	98.5	88.5
Netherlands	105.1	102.8	103.2	104.7	102.9	97.9	105.1	110.8	109.7
New Zealand	102.8	101.0	96.6	94.6	100.2	91.5	96.1	76.9	77.7
Norway	103.7	98.3	94.4	93.0	97.2	98.3	111.4	144.3	124.9
Portugal	84.3	96.7	100.9	99.7	91.9	90.6	72.2	66.8	65.7
Spain	92.7	97.0	95.3	94.9	97.2	98.9	81.9	80.0	89.6
Sweden	104.0	95.5	98.5	95.7	100.9	95.1	113.9	106.3	117.8
Switzerland	105.1	102.7	97.8	106.9	103.5	110.6	104.9	114.1	122.0
United Kingdom	99.5	99.6	102.0	99.4	99.6	104.8	98.5	100.4	98.4
United States	97.3	97.5	102.4	101.3	98.7	98.4	114.3	128.0	111.3
average	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Notes: Adult skills = average of literacy and numeracy scores from PIAAC, except for Switzerland and Portugal, where ALL and IALS are respectively used. Sudent skills = average PISA scores in 2012. Student achiev = average scores in the 2011 round of IEA tests (TIMSS and PIRLS). HLO = Average harmonized learning outcomes (HLOs) in 2010, taken from AAP (2018). H&W = Hanushek & Woessmann's (2015) indicator of population cognitive skills. HLO avge. till 2005 = cumulative average of available HLO's until 2005. Years of schooling = average years of schooling in 2010 from de la Fuente and Doménech (2015). ypc 15-64 = GDP per working-age person in 2010. Constructed using OECD data.
welfare = average social welfare from 2005 to 2015. Constructed using Jones and Klenowt's (2016) methodology.

5 | A QUICK LOOK AT THE DATA

Table 4 collects some educational indicators of interest, mostly referring to years around 2010, for a sample of 21 OECD countries. All variables are normalized, with their unweighted cross-country averages set to 100. Column [1] shows an indicator of *adult skills*, constructed as the average of literacy and numeracy scores. For most countries, the data come from PIAAC. For those that did not participate in this assessment, we use the results of the most recent similar test that is available, that is, ALL in Switzerland and IALS in Portugal. Column [2] shows average PISA scores in 2012 and column [3] average scores (across available subjects and grades in each country) in the 2011 round of IEA tests (TIMSS and PIRLS). The following three columns contain summary

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 5 Correlations between pairs of indicators

Correlation with:	*Student skills*	*Student achiev*	*HLO*	*H&W*	*HLO avge till 2005*	*Years of schooling*	*Ypc 15-64*	*Welfare*
adult skills (PIAAC)	0.641	0.295	0.448	**0.839**	0.545	**0.754**	**0.559**	**0.570**
student skills (PISA)	1.000	0.548	**0.743**	0.812	0.525	0.407	0.265	0.206
student achievement (IAE)		1.000	**0.812**	0.481	0.463	0.241	0.047	−0.049
HLO			1.000	0.669	**0.686**	0.270	0.177	0.160
H&W				1.000	**0.640**	**0.658**	0.418	0.502
HLO avge until 2005					1.000	0.329	0.339	0.310
years of schooling						1.000	**0.816**	**0.755**
ypc 15-64							1.000	**0.826**

Notes: Adult skills = average of literacy and numeracy scores from PIAAC, except for Switzerland and Portugal, where ALL and IALS are respectively used. Student skills = average PISA scores in 2012. Student achievement = average scores in the 2011 round of IEA tests (TIMSS and PIRLS) .

HLO = Average harmonized learning outcomes (HLOs) in 2010, taken from AAP (2018). H&W = Hanushek & Woessmann's (2015) indicator of population cognitive skills. HLO avge. till 2005 = cumulative average of available HLO's until 2005. Years of schooling = average years of schooling in 2010 from de la Fuente and Doménech (2015). ypc 15-64 = GDP per working-age person in 2010. Constructed using OECD data. .

welfare = average social welfare from 2005 to 2015. Constructed using Jones and Klenowt's (2016) methodology.

indicators based on student achievement and skills tests. Average HLOs in 2010, taken from AAP (2018) and shown in column [4], contain information on both student achievement and student skills. Column [5] shows H&Wt's (2015) indicator of population cognitive skills, which is constructed by averaging student achievement and literacy tests on math and science (but not reading) over several decades. Column [6] contains a similar "stock" indicator of educational quality, the cumulative average of available HLO's until 2005, which also incorporates reading results. Column [7] shows our estimate of average years of schooling in 2010, taken from de la Fuente and Doménech (D&D, 2015).[12] Column [8] displays relative real GDP per working-age person (relative income per capita, for short, from now on), taken from an updated version of the data set used in D&D (2006).[13] Finally, column [9] shows the average value between 2005 and 2015 of the measure of social welfare proposed by Jones and Klenow (2016), which is computed by aggregating (with the appropriate utility weights) private and public consumption per capita, an indicator of income equality, hours worked and life expectancy, all of them potentially related to human capital.[14]

Table 5 displays pairwise correlations for the variables shown in Table 4. It should be noted that correlations between quality variables, while always positive, are often fairly low, suggesting that it may be difficult to construct a single indicator that adequately summarizes educational quality. We can classify the educational indicators we have gathered into two groups: *flow* measures of student performance (measuring the academic achievement or basic skills of each young cohort), which can be seen as indicators of the quality of education at a given point in time, and *stock* measures of the quantity or quality of schooling for the entire adult or working-age population, which are sometimes constructed by averaging flow indicators over long periods. Correlations tend to be higher within each of these groups than across them, but there are some exceptions. As should

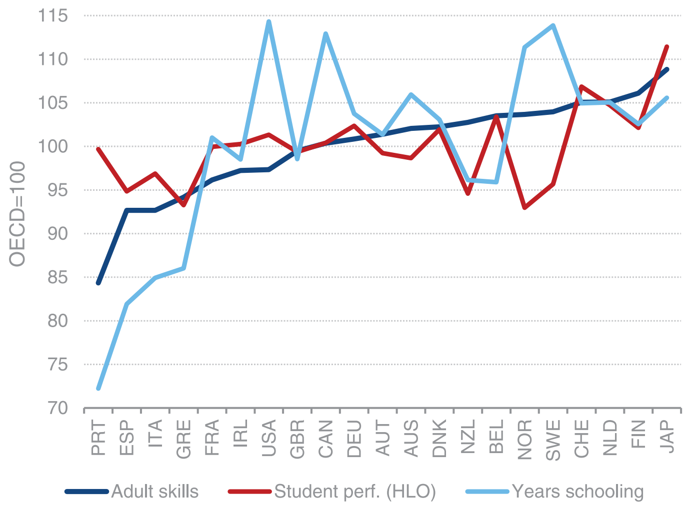

FIGURE 1 Selected human capital indicators around 2010 Unweighted sample average = 100 [Colour figure can be viewed at wileyonlinelibrary.com]

be expected, stock measures are more highly correlated with income per capita than flow measures. As a summary indicator of student performance, we will use AAP's HLOs, which combine information on all PISA and IEA scores and display a fairly high correlation with both of these variables (0.743 and 0.812, respectively). As for the stock measures, we will focus on adult skills and years of schooling as direct measures of quality and quantity, and retain for some purposes the other two variables, H&W's indicator of cognitive skills and the average value of available HLOs until 2005.

Figure 1 displays the cross-section profile of the three main indicators we have selected, those that measure adult skills, student performance and average years of schooling of the adult population, with countries ordered by the adult skills indicator. In terms of this variable, Southern European countries display the lowest scores, followed by the Anglo-Saxon and Central European nations, while Northern Europe and Japan perform best. Country performance in terms of student competencies and average years of schooling, however, often deviates markedly from this pattern. For instance, the United States, Canada, Norway, and Sweden do much better in terms of years of schooling than in adult skills, while the opposite is true of Southern Europe. Roughly speaking, student performance measures tend to lie above adult skills for lower values of the latter variable and below them in the upper half of the distribution.

Table 6 shows country rankings according to the same three indicators, together with each country's average rank and its rank range, defined as the difference between its highest and lowest rankings. Looking at the table, it is clear that the three indicators generate rather different rankings. In some cases, the differences across indicators for a given country are quite striking. For instance, Sweden and Norway do quite well in terms of adult skills (where they rank in positions 5 and 6) but very poorly in terms of student performance in standardized tests (where they drop to positions 17 and 21, respectively), and the United States goes from the first position in terms of years of schooling to the 15th when we consider adult skills. Japan, the Netherlands, Switzerland, and Finland are well ranked in terms of adult and student performance, but not so much when it comes to years of schooling, and Southern Europe displays consistently poor performance in terms of all indicators, with the partial exception of Portugal in the case of student performance.

As we have already indicated, we expect that both years of schooling and educational quality should contribute positively to adult skills. As a very rough test of this hypothesis,

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 6 Country rankings around 2010

	Adult skills	*Student perf. HLOs*	*Years of schooling*	*Average rank*	*Range max – min rank*
Japan	1	1	6	2.7	5
Netherlands	3	3	7	4.3	4
Switzerland	4	2	8	4.7	6
Finland	2	6	11	6.3	9
Sweden	5	17	2	8.0	15
United States	15	8	1	8.0	14
Canada	13	9	3	8.3	10
Denmark	9	7	10	8.7	3
Germany	12	5	9	8.7	7
Belgium	7	4	17	9.3	13
Australia	10	15	5	10.0	10
Norway	6	21	4	10.3	17
Austria	11	14	12	12.3	3
France	17	11	13	13.7	6
Ireland	16	10	15	13.7	6
United Kingdom	14	13	14	13.7	1
New Zealand	8	19	16	14.3	11
Italy	19	16	19	18.0	3
Portugal	21	12	21	18.0	9
Greece	18	20	18	18.7	2
Spain	20	18	20	19.3	2

Note: Adult skills, average of literacy and numeracy scores from PIAAC, except for Switzerland and Portugal, where ALL and IALS are respectively used. Student perf. HLOs = Average harmonized learning outcomes (HLOs) in 2010, taken from AAP (2018). Years of schooling = average years of schooling in 2010 from de la Fuente and Doménech (2015).

we can regress the PIAAC-based indicator of adult skills, which a priori would seem to be the best available proxy for this variable, on years of schooling (*yrsch*) and one of the two stock quality indicators we have selected (*h&w* and *hlo_at05*), with all variables measured in logs.

As can be seen in Table 7, the results are consistent with our hypothesis that both quantity and quality matter. Years of schooling and educational quality are always significant, whether entered alone or jointly in the equation. H&W's science and math-based cognitive skills indicator performs better than the cumulative average of HLOs, but in both cases the strategy of averaging flow performance measures over several decades seems to be successful at producing a stock indicator of quality that helps explain average levels of adult skills. Incidentally, the high R-squared of these regressions suggest that we are likely to run into severe multicollinearity problems if we try to use several educational indicators as explanatory variables for income or welfare levels or growth rates in the same equation. One way to mitigate this problem may be to use quality-adjusted years of schooling. With the variables in logs, this basically involves adding up the quantity and quality variables to leave a single regressor, a procedure that would only be justified if we cannot reject the hypothesis that the coefficients of the two variables in the regression are equal. Looking at Equations (2) and (3), the relevant coefficients are similar in the case of *hlo*, but not when we use

TABLE 7 Determinants of adult skills around 2010

	[1]	*[2]*	*[3]*	*[4]*	*[5]*	*[6]*
Lyrsch	0.397	0.343	0.201			
	(5.66)	(4.95)	(2.69)			
lhlo_at05		0.290			0.536	
		(2.14)			(2.85)	
lh&w			0.920	1.374		
			(3.79)	(6.81)		
Lqayrsch						0.328
						(6.67)
R^2	0.6275	0.7031	0.7927	0.7092	0.2991	0.7010

Notes: the dependent variable, *ladult*, is the log of the adult skills indicator, an average of literacy and numeracy scores from PIAAC, except for Switzerland and Portugal.
lyrsch = average years of schooling in 2010 from de la Fuente and Doménech (2015).
lhlo_at05 = cumulative average of available HLO's until 2005.
lh&w = Hanushek and Woessmannt's (2015) indicator of population cognitive skills .
lqayrsch = quality adjustment based on *hlo_at05*, using as a benchmark the sample average of this indicator.
All equations include a constant that is not reported. All regressors are measured in logs.

h&w as a quality indicator, suggesting that quality adjustment would be acceptable for the first variable but not for the second. In column (6) we impose the equality restriction and check that HLO-based quality-adjusted years of schooling (*qayrsch*) performs rather well – although slightly less so than *lh&w* alone.

Table 8 displays the results of separate regressions of individual PIAAC scores on years of schooling; time elapsed since the completion of schooling and the square of this last variable for those countries for which all the required data are available. The correlation between years of schooling and PIAAC scores at the individual level is very strong and statistically significant in all countries, suggesting that, as may be expected, skills are gradually acquired over time in school. The expected contribution of a year of schooling to the average PIAAC score ranges between 4.77 points in Italy to 8.36 points in Germany, but it is not clear that the value of this coefficient can be interpreted as an indicator of school quality, as intercept coefficients also vary widely within the sample and do so in a way that tends to offset slope differences. Figure 2a shows the estimated effect of years of schooling on adult skills. On average, each additional year of schooling increases PIAAC scores by approximately 6 points.

Except for Greece and the United Kingdom, PIAAC scores are lower for individuals who left school a long time ago than for more recent graduates with the same level of schooling as shown in Figure 2b. This pattern suggests that school-acquired knowledge and competencies depreciate over time, possibly as a result of aging and obsolescence, but may also reflect an increase in the quality of schooling over time. There are, however, some exceptions and significant differences across countries in the rate at which skills seem to depreciate over time. In any case, the effect of time elapsed since graduation is relatively small: on average adult skills scores are 20 points lower after 45 years since graduation.

Switching from the cross-section to the time-series dimension, we are interested in the stability of flow measures of educational performance over time. The assumption that quality levels do not change much over time has been made in many studies (see, e.g., H&K) as a convenient way to get around limited data availability, for it allows us to approximate the skill level of the entire adult

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 8 PIAAC average score in math and reading as a function of years of schooling and time elapsed since completion of studies individual data by country

	yrs school	Time since grad	Time since grad sq	constant	*N* obs	Rsq
Germany	8.36	−0.91	0.00	176.70	5.088	0.322
	(37.48)	(6.09)	(0.16)	(54.64)		
Belgium	7.39	−0.85	0.00	203.50	4.910	0.324
	(35.20)	(6.33)	(1.16)	(71.25)		
Denmark	6.85	−0.46	−0.01	200.00	7.167	0.234
	(32.75)	(3.54)	(1.74)	(73.29)		
Spain	6.20	−0.45	0.00	191.90	5.689	0.333
	(38.48)	(3.68)	(1.06)	(82.93)		
Finland	5.94	−0.92	0.00	227.70	5.420	0.259
	(23.11)	(5.76)	(0.80)	(69.70)		
France	7.30	−0.88	0.01	189.40	6.617	0.355
	(44.49)	(7.49)	(3.72)	(82.28)		
Ireland	6.53	−0.33	0.00	171.30	5.890	0.239
	(27.10)	(2.19)	(0.86)	(41.95)		
Italy	4.77	−0.38	0.00	207.00	4.506	0.237
	(22.82)	(2.34)	(0.40)	(57.29)		
Japan	6.53	**+0.46**	**−0.02**	213.80	5.147	0.301
	(31.57)	(3.75)	(8.83)	(72.38)		
Norway	7.05	−0.16	−0.01	186.70	4.888	0.196
	(23.43)	(0.99)	(2.14)	(44.84)		
Netherlands	6.50	−0.34	−0.01	210.00	4.968	0.269
	(24.71)	(2.33)	(3.04)	(55.90)		
UK	7.53	**+0.56**	**−0.01**	170.20	7.549	0.145
	(23.10)	(3.04)	(2.38)	(37.40)		
Sweden	7.95	−0.87	0.01	195.00	4.363	0.207
	(25.80)	(5.16)	(2.43)	(49.74)		

Note: *t* statistics in parentheses below estimated coefficients. The dependent variable is the individual's PIAAC score, measured as the average value of the math and reading scores. The regressors are the number of years of schooling, the time elapsed since the completion of schooling and the square of this last variable.

population using data on the educational performance of current and recent cohorts of students. It is not clear, however, that this is indeed the case.

As shown in Figure 3, a first look at HLO scores, the indicator with more observations in our sample, shows a lot of variation over time, some strange patterns and positive trends for many countries and for the sample average. Thus, the average score for the 21 countries has increased steadily, except in 1980, between 1970 and 2015, rising from 467 to 522, with a 11.6% increase in educational performance. The improvement in scores has been significant in the case of Portugal: in 1990 it was the country with the worst performance in the sample, with a HLO score of 397, but 25 years later it had become the 7th country with better performance, placed between the United States and Germany, after having registered a 32% increase in scores. At the same time, there are

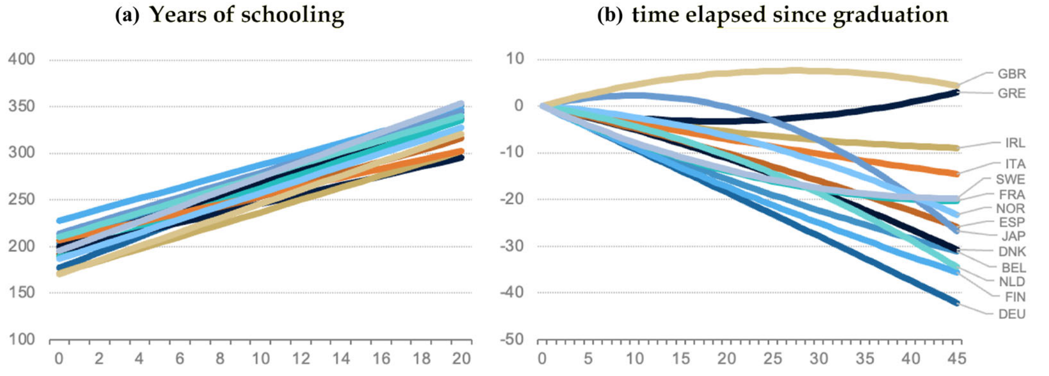

FIGURE 2 Predicted PIAAC score by country as a function of (a) years of schooling and (b) time elapsed since graduation [Colour figure can be viewed at wileyonlinelibrary.com]

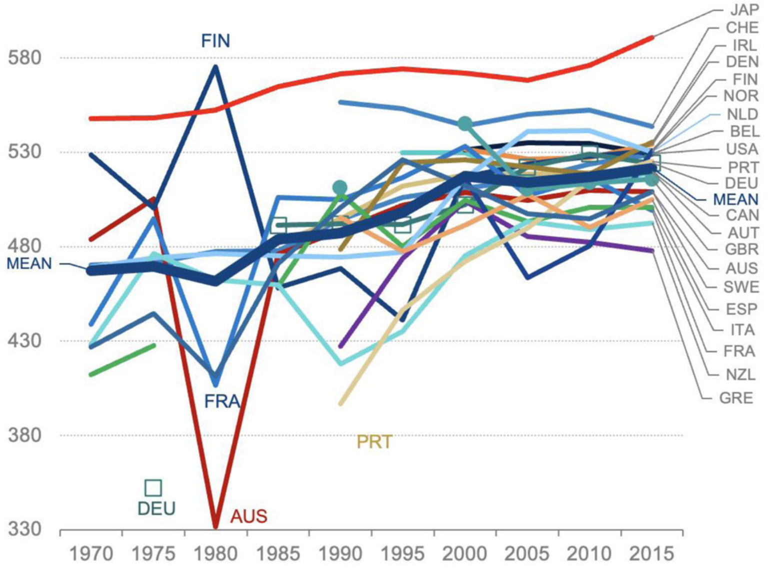

FIGURE 3 HLO scores over time, 1970–2015, 21 OECD countries [Colour figure can be viewed at wileyonlinelibrary.com]

some surprising observations as, for example, Germany in 1975 or Australia, Finland, and France in 1980.

To gauge the degree of stability of country performance over time, we estimate country-specific trends as follows. Given an educational indicator, x, let

$$\Delta x_n = \frac{x_n - x_{n-1}}{t_n - t_{n-1}} \tag{6}$$

be its average annual variation between observations $n{-}1$ and n, dated at t_{n-1} and t_n, respectively. For each country, we estimate a regression of the form

$$\Delta x_n = g * x_{n-1} + e \tag{7}$$

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 9 Estimated country specific trends of some indicators of educational quality

	Student skills (PISA)		***Student achievement (IEA)***		***HLOs***	
	G	***(t)***	***g***	***(t)***	***g***	***(t)***
Australia	**−0.33%**	(4.25)	**+0.12%**	(2.23)	−0.13%	(0.12)
Austria	**−0.23%**	(1.07)	+0.04%	(0.13)	−0.10%	(0.65)
Belgium	−0.08%	(0.64)	−0.06%	(0.13)	−0.03%	(0.28)
Canada	**−0.16%**	(1.55)	+0.16%	(0.69)	**+0.24%**	(2.00)
Denmark	+0.04%	(0.29)	+0.13%	(0.38)	+0.03%	(0.21)
Finland	**−0.24%**	(1.13)	−0.44%	(0.72)	−0.11%	(0.14)
France	**−0.15%**	(0.94)	**−0.42%**	(2.02)	+0.16%	(0.22)
Germany	+0.14%	(0.45)	+0.16%	(0.42)	**+0.50%**	(1.18)
Greece	−0.09%	(0.44)	–	–	+0.37%	(0.72)
Ireland	−0.11%	(0.40)	+0.16%	(0.45)	**+0.42%**	(1.12)
Italy	+0.03%	(0.11)	−0.20%	(0.78)	**+0.35%**	(0.98)
Japan	−0.25%	(0.80)	**+0.12%**	(0.99)	**+0.17%**	(2.28)
Netherlands	**−0.29%**	(2.51)	+0.04%	(0.11)	**+0.27%**	(1.27)
New Zealand	**−0.31%**	(1.90)	−0.09%	(0.30)	+0.28%	(0.69)
Norway	−0.06%	(0.23)	+0.12%	(0.23)	+0.13%	(0.10)
Portugal	**+0.34%**	(0.95)	0.00%	(0.01)	**+1.05%**	(3.18)
Spain	−0.06%	(0.28)	+0.09%	(0.29)	+0.07%	(0.20)
Sweden	−0.12%	(0.45)	−0.21%	(0.45)	**+0.34%**	(0.86)
Switzerland	−0.09%	(0.48)			**−0.09%**	(0.87)
United Kingdom	−0.17%	(0.82)	**+0.15%**	(0.86)	−0.14%	(0.32)
United States	−0.05%	(0.18)	+0.04%	(0.36)	**+0.26%**	(3.85)
Corr. con PISA	1.000		0.124		0.665	
Max no. of obs.	6		6		9	

Notes: Student skills = average PISA scores in 2012; student achievement = average scores in the 2011 round of IEA tests (TIMSS and PIRLS) .
HLOs = Average harmonized learning outcomes (HLOs) in 2010, taken from AAP (2018).

where the constant has been suppressed and ε is a random disturbance.

The results, shown in Table 9, suggest that educational quality is not stable over time in many countries. At the standard confidence level of 95%, only between two and four countries display trends that are significantly different from zero depending on the specific variable we examine. On the other hand, this test may be too stringent given the small number of observations, which go from a minimum of 6 to a maximum of 9 per country, depending on the indicator. If we consider all estimates with a *t* ratio of roughly one or greater, the number of countries for which there are fairly clear indications of a positive or negative trend rises sharply, raising increasing doubts on the validity of the constant quality assumption that has often been used in the literature to justify the use as an explanatory variable for growth of average test scores computed over different time periods depending on data availability in each country. This result, however, does not raise doubts about the potential usefulness of such data in combination with more appropriate flow specifications.

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

6 | CONCLUSIONS

Scores in standardized international student achievement tests and some recent adult literacy studies provide interesting data on the quality of educational outputs and on the skill level of the population that can be a useful complement to the data on the quantity of schooling that have been most commonly used in the growth literature. The use of these data is likely to improve our ability to measure human capital accurately and to help us understand its contribution to output levels, economic growth, and social welfare. In this paper we have reviewed the main sources of primary data in this area and some recent efforts to systematize and standardize them with a view to constructing useful summary indicators of the quality of educational systems or the skill level of the adult population. We have used these data to look at cross-country educational performance in recent years within a sample of 21 OECD countries, for which the quality of the data and the number of observations available are greater than for developing countries, and discussed some of their limitations in terms of their potential use in empirical growth studies.

Accepting that quality matters in education does not mean that quantity should be ignored. The skill level of the labor force will surely depend on both the quantity and the quality of schooling. We have provided some preliminary evidence in favor of this view and argued that progress in this area is most likely to come from studies that try to combine both dimensions, such as those that construct measures of *LAYS*.

The scarcity of quality data, both across countries and over time, however, will be a serious handicap in this effort. Our series of student performance indicators are not long enough to allow us to construct good stock measures of quality by averaging scores over a sufficiently long period, but we may be able to get around this difficulty by using flow data on quality in MRW-type panel specifications. Another possibility may be to try to correct years of schooling for quality before using them to estimate standard growth or productivity equations with panel data, although the quality indicators required for the correction are likely to lack time variation. A third route relies on the construction of synthetic time series of adult skill indicators using the available information on the age distribution of PIAAC results, and possibly correcting for estimated depreciation over time. This approach is likely to become more productive as new waves of PIAAC become available.

ACKNOWLEDGMENTS

The authors thank the useful comments and suggestions by the co-editor-in-chief and three referees, and gratefully acknowledge financial support from the Spanish and Valencian governments through research projects CICYT PID 2020-116242RB-I00 and GVPROMETEO2020-083.

ORCID

Ángel de la Fuente https://orcid.org/0000-0002-5738-5162
Rafael Doménech https://orcid.org/0000-0001-9902-1060

ENDNOTES

[1] We will work with a sample comprised by the initial OECD countries except for a few small and somewhat atypical economies such as Luxembourg and Iceland. These countries are, in particular, Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, and United States.

[2] For a useful introduction to IRT proficiency scoring, see Annex B in OECD and Statistics Canada (2011).

[3] The authors use the results of six such tests that were conducted between 1965 and 1991 (four by IEA and two by IAEP (International Assessment of Educational Progress).

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[4]NAEP (National Assessment of Educational Progress) measures the performance of US students at different benchmark ages in around a dozen subjects. One strand of the test (the so-called long-term trend assessments) provides nationally representative scores for math and reading for 9-, 13-, and 17-year old students at intervals of 2–4 years since the early 1970s that are measured on a consistent scale and can therefore be compared over time.

[5]NAEP scores are available at 2- to 4-year intervals over the period; values for non-NAEP years are obtained by linear interpolation between available years.

[6]In addition to the papers cited in the text, see also Altinok and Murseli (2007), Angrist et al. (2013), Altinok et al. (2014), Altinok et al. (2019).

[7]In the most recent version of this database (Angrist et al., 2021, supplementary information p. 7) the standardization procedure is based on a regression of the form $y_i = a + bx_i + \varepsilon_i$. The equation is estimated using data for all countries that participate in both tests and is then used to estimate y for those countries that have only participated in X.

[8]In principle, they would like to base the adjustment on a measure of how much is learned on average during an additional year in school, but this is difficult to estimate without simplifying assumptions that essentially bring us back to observed relative scores in achievement tests.

[9]Hence, the indicator of quality-adjusted human capital would be of the form $H = exp(r*YRSCH+\omega*Q)$, where r and ω are estimates of the returns to the quantity and quality of schooling.

[10]The exception is the attempt by Demirgüç-Kunt and Torre (2020) to measure quality-adjusted years of tertiary education for 48 countries in Europe and Central Asia and 7 high-income countries from other regions, using information from six university rankings (see also World Bank, 2020).

[11]Test scores have also been extensively used to estimate education production functions, that is, to analyze to what extent school resources, expenditure by student, the number of students per teacher, teacher quality or parents' income and years of schooling can explain children's educational performance. Hanushek and Wößmann (2010) and Woessmann (2016) provide excellent surveys on the main results of this literature.

[12]As shown in de la Fuente and Doménech (2015), our data set of years of schooling has a higher signal to noise ratio than Barro and Lee's (2013) estimates.

[13]This data set has been compiled using OECD data on member states' national accounts, working-age populations and a set of OECD-specific purchasing power parities.

[14]Most of the empirical growth literature has focused on the effects of human capital on economic development, using GDP per capita as its usual proxy. However, the limitations of GDP as a measure of welfare are well known. In recent decades, and particularly since the Stiglitz Commission Report (see Stiglitz, Sen, and Fitoussi 2009), there have been many attempts to propose better approximations to economic well-being. In contrast to other alternatives such as ad-hoc composite indices or dashboard approaches, Jones and Klenow's (2016) measure of economic welfare provides a well-founded procedure for the aggregation of different determinants of well-being and also allows cross-country and intertemporal comparisons.

REFERENCES

Altinok, N., Angrist, N., & Patrinos, H. (2018). Global data set on education quality (1965–2015). *World Bank, Policy Research*, Working Paper no. 8314.

Altinok, N., & Murseli, H. (2007). International database on Human Capital Quality. *Economics Letters*, *96*(2), 237–244.

Altinok, N., Diebolt, C., & de Meulemeester, J.-L. (2014). A new international database on education quality: 1960–2010. *Applied Economics*, *46*(11), 1212–1247.

Angrist, N., Djankov, S., Goldberg, P., & Patrinos, H. (2019). *Measuring human capital*. Policy Research Working Paper 8742, World Bank.

Angrist, N., Djankov, S., Goldberg, P., & Patrinos, H. (2021). Measuring human capital using global learning data. *Nature*, *592*, 403–408. https://doi.org/10.1038/s41586-021-03323-7

Angrist, N., Patrinos, H. A., & Schlotter, M. (2013). An expansion of a global data set on educational quality. (No. WPS6536). World Bank Policy Research Working Paper no. 6536.

Barro, R. J., & Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, *104*, 184–198.

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Botev, J., Égert, B., Smidova, D., & Turner, Z. (2019). A new macroeconomic measure of human capital with strong empirical links to productivity. *OECD, Economics Department*, Working paper no. 1575, Paris. https://doi.org/10.1787/d12d7305-en

Breton, T. (2011). The quality vs. the quantity of schooling: What drives economic growth? *Economics of Education Review*, *30*, 765–773.

Caselli, F. (2005). Accounting for cross-country income differences. *Handbook of Economic Growth*, 1(A), 679–741. http://t.ly/q9ZO

Castelló-Climent, A. (2019). The age structure of human capital and economic growth. *Oxford Bulletin of Economics and Statistics*, *81*(2), 394–411.

Chinapah, V. (2003). Monitoring Learning Achievement (MLA) project in Africa. *ADEA Biennial Meeting 2003*, Grand Baie, Mauritius, 3-6 December. http://bit.ly/3L3I5no

Cohen, D., & Leker, L. (2014). *Health and education: Another look with the proper data*, Mimeo Paris School of Economics.

Coulombe, S., & Tremblay, J. F. (2006). Literacy and growth. *B.E. Journals in Macroeconomics: Topics in Macroeconomics*, *6*(2), 1–32.

de la Fuente, A., & Doménech, R. (2006). Human capital in growth regression: How much difference does data quality make? *Journal of the European Economic Association*, *4*(1), 1–36.

de la Fuente, A., & Doménech, R. (2015). Educational Attainment in the OECD, 1960-2010. Updated series and a comparison with other sources. *Economics of Education Review*, *48*, 56–74.

Demirgüç-Kunt, A., & Torre, I. (2020). Measuring human capital in Europe and Central Asia. *World Bank Policy Research Working Paper, 9458*. http://bit.ly/3kij5NH

Égert, B., de la Maisonneuve, C., & Turner, D. (2022). A new macroeconomic measure of human capital exploiting PISA and PIAAC: Linking education policies to productivity. *OECD Economics Department Working Papers, No. 1709*. http://bit.ly/3NiBC91

Feenstra, R., Inklaar, R., & Timmer, M. (2015). The NextGeneration of the Penn world table. *American Economic Review*, *105*(10), 3150–3182. https://bit.ly/3udI5rH

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.

Folloni, G., & Vittadini, G. (2010). Human capital measurement: A survey. *Journal of Economic surveys*, *24*(2), 248–279.

Gaëlle, P., Sanchez Puerta, M. L., Valerio, A., & Rajadel, T. (2014). STEP skills measurement surveys: Innovative tools for assessing skills. Social Protection and Labor Discussion Paper No. 1421. World Bank. http://bit.ly/3jXIBYj

Glawe, L., & Wagner, H. (2022). Is schooling the same as learning? The impact of the learning-adjusted years of schooling on growth in a dynamic panel data framework. *World Development*, *151*, 105773. http://bit.ly/3MjpLH9

Hall, R. E., & Jones, C. I. (1999). Why do some countries produce so much more output per worker than others? *The Quarterly Journal of Economics*, *114*(1), 83–116. http://t.ly/K-wC

Hanushek, E., & Kimko, D. (2000). Schooling, labor-force quality and the growth of nations. *American Economic Review*, *90*(5), 1184–208.

Hanushek, E., & Wößmann, L. (2010). The economics of international differences in educational achievement. NBER Working Paper no. 15949, Cambridge, Mass.

Hanushek, E., & Wößmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes and causation. *Journal of Economic Growth*, *17*, 267–321.

Hanushek, E., & Wößmann, L. (2015). *The knowledge capital of nations: Education and the economics of growth. CESifo book series*, MIT Press.

Jones, C. I., & Klenow, P. J. (2016). Beyond GDP? Welfare across countries and time. *American Economic Review*, *106*(9), 2426–57.

Kaarsen, N. (2014). Cross-country differences in the quality of schooling. *Journal of Development Economics*, *107*, 215–224.

Keslair, F., & Paccagnella, M. (2020), Assessing adults' skills on a global scale: A joint analysis of results from PIAAC and STEP. OECD Education Working Papers No. 230. http://bit.ly/3v1qQgS

Lee, J.-W., & Barro, R. (2001). Schooling quality in a cross-section of countries. *Economica*, *68*(272ą), 465–488.

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, *107*(2), 407–437.

Martin, M., Mullis, I., Foy, P., & Stanco, G. (2012). TIMSS 2011 international results in science. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement, Amsterdam. http://timssandpirls.bc.edu/timss2011/international-results-science.html

Michaelowa, K. (2001). Primary education quality in francophone Sub-Saharan Africa: Determinants of learning achievement and efficiency considerations. *World Development*, *29*(10), 1699–1716.

Mullis, I., Martin, M., Foy, P., & Arora, A. (2012). TIMSS 2011 international results in mathematics. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement, Amsterdam. http://timssandpirls.bc.edu/timss2011/international-results-mathematics.html

Mullis, I., Martin, M., Foy, P., & Drucker, K. (2012). PIRLS 2011 international results in reading. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement, Amsterdam. http://timssandpirls.bc.edu/pirls2011/international-results-pirls.html

Mullis, I., Martin, M., Beaton, A., Gonzalez, E., Kelly, D., & Smith, T. (1998). *Mathematics and science achievement in the final year of secondary school. IEA's Third International Mathematics and Science Study (TIMSS)*. TIMSS International Study Center. http://timssandpirls.bc.edu/timss1995i/MathScienceC.html

Murimba, S. (2005). The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ): Mission, approach and projects. Prospects. *Quarterly Review of Comparative Education*, *35*(1), 75–89. http://bit.ly/3ECbse9

OECD (2004). *Learning for tomorrow's world. First results from PISA 2003. Programme for International Student Assessment*, Paris. http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/learningfortomorrowsworld-englishversion-chapterbychapter.htm

OECD. (2009). International adult literacy and basic skills surveys in the OECD region. *Education Directorate*, Working Paper no. 26.

OECD. (2013). *OECD skills outlook 2013. First results from the survey of adult skills*. OECD Publishing. http://bit.ly/38YtwDp

OECD. (2014). *PISA 2012 results in focus. What 15-year-olds know and what they can do with what they know*. Programme for International Student Assessment, Paris. http://www.oecd.org/pisa/keyfindings/pisa-2012-results.htm

OECD and Statistics Canada. (2000). *Literacy in the information age*. Final report of the International Adult Literacy Survey, Paris. http://www.oecd.org/edu/skills-beyond-school/41529765.pdf

OECD and Statistics Canada. (2011). *Literacy for life: Further results from the Adult Literacy and Life Skills Survey. Second International ALL Report*. OECD Publishing. http://www.statcan.gc.ca/pub/89-604-x/89-604-x2011001-eng.pdf

OECD and Unesco Institute for Statistics. (2003). *Literacy skills for the world of tomorrow. Further results from PISA 2000. Programme for International Student Assessment*, Paris. http://www.keepeek.com/Digital-Asset-Management/oecd/education/literacy-skills-for-the-world-of-tomorrow_9789264102873-en#page286

Pritchett, L. (2013). The rebirth of education: Schooling ain't learning. *Center for Global Development*, http://bit.ly/3JX9qGE

Psacharopoulos, G. (1994). Returns to investment in education: A global update. *World Development*, *22*(9), 1325–1343.

Schoellman, T. (2012). Education quality and development accounting. *The Review of Economic Studies*, *79*(1), 388–417.

Schwerdt, G., & Wiederhold, S. (2018). *Literacy and growth: New evidence from PIAAC*. Mimeo, University of Konstanz and Catholic University Eichstaett-Ingolstadt.

Stiglitz, J. Sen, A., & Fitoussi, J. P. (2009), Report by the Commission on the Measurement of Economic Performance and Social Progress. http://t.ly/-kkt

UNESCO-OREALC. (2016). *Reporte Técnico. Tercer Estudio Regional Comparativo y Explicativo, TERCE*. http://bit.ly/3LfyPNi

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Walker, M. (2011). *PISA 2009 plus results: Performance of 15-year-olds in reading, mathematics and science for 10 additional participants*. ACER Press. http://t.ly/Q5G5

Wößmann, L. (2003). Specifying human capital. *Journal of Economic Surveys*, *17*(3), 239–270.

Woessmann, L. (2016). The importance of school systems: Evidence from international differences in student achievement. *The Journal of Economic Perspectives*, *30*(3), 3–32. http://bit.ly/394mHAd

World Bank (2018). World development report 2018: Learning to realize education's promise. The World Bank. http://bit.ly/3L0EhmW

World Bank. (2020). Europe and Central Asia Economic Update, Fall 2020: COVID-19 and human capital. The World Bank. http://bit.ly/3khOMqp

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Fuente, Á., & Doménech, R. (2022). Cross-country data on skills and the quality of schooling: A selective survey. *Journal of Economic Surveys*, 1–24. https://doi.org/10.1111/joes.12530

14676419, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/joes.12530 by Universitat De Valencia, Wiley Online Library on [21/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License