

Supplementary Information for

Skills-adjusted human capital shows rising global gap

Wolfgang Lutz, Claudia Reiter, Caner Özdemir, Dilek Yildiz, Raquel Guimaraes, Anne Goujon

Corresponding author: Claudia Reiter Email: claudia.reiter@univie.ac.at

This PDF file includes:

Supplementary text Figures S1 to S8 Tables S1 to S4 SI References

Supplementary Information Text

DATA AND METHODS

The SLAMYS dataset was developed in three steps: first, for 44 countries SLAMYS were computed using comprehensive adult literacy skills data (PIAAC (1) and STEP (2)). Second, to increase coverage among developing countries, we used DHS (3) tested literacy data to provide skills adjustments for 59 additional countries. Finally, to expand the dataset to a global scale, we used a prediction regression model for countries where no empirical adult skills data are available. A detailed description of all data sources used can be found elsewhere (4).

Literacy skills represent only one domain of a variety of skills considered essential for the formation of human capital. However, the limited availability of assessment data of other skill domains (e.g. numeracy, problem-solving skills, etc.) for many countries and over time constrains us in using a more comprehensive definition of skills for the estimation of skills-adjusted human capital. Thus, analyses throughout this paper exclusively focus on literacy skills. Despite this limitation, sensitivity analyses revealed that literacy is highly correlated with other type of skills. Figure S1 displays the correlation between mean literacy scores and mean numeracy scores (left-hand plot) and between mean literacy scores and mean scores in problem-solving in technology-rich environments (right-hand plot) by age, sex, and country for all countries participating in PIAAC. The high Pearson correlation coefficients, particularly for numeracy (0.96) but also for problem-solving (0.87), and the high statistical significance (p-value < 0.001 in both correlation tests) point out that the level of literacy skills is a good proxy for the overall skill level in the population - particularly when considering them at the aggregate level.

In addition, it is important to note that even for countries with comprehensive adult literacy skills data available, issues of measurement errors cannot be completely ruled out. Despite advanced and complex survey designs which specifically aim at reducing potential sources of error and maximizing the quality of the data produced, tests as in PIAAC or STEP may still suffer from a variety of problems. Amongst others, these may include the sampling of knowledge in the particular domain, the reliability of question, or the impact of test taking conditions on scores (5). A discussion on potential measurement errors resulting from the prediction regression model and from measuring the quantitative dimension, i.e. Mean Years of Schooling, can be found at the end of the SI.

Calculation of SLAMYS based on empirical PIAAC and STEP data

Computations for the base year (2015)

When adding a skills dimension to educational attainment, a standard of comparison needs to be established, whether it is a perfect (unattainable) score (e.g. 500 in PIAAC; 1,000 in PISA, etc.), a benchmark result of the top-performer, or the performance of any group of individuals. Since our estimates are based on the average performance of populations, we decided to use the mean proficiency of the OECD population, disaggregated by age-sex-education groups, as a standard of comparison. Specifically, our standard equals the 2015 population-weighted* OECD[†] mean PIAAC literacy test score, calculated separately for males, females, 5-year age groups (between 15-19 and 60-64), and four educational attainment categories (primary or less, lower secondary, upper secondary, and post-secondary education).

The skills adjustment was designed in such a way that, for our standard of comparison, the Mean Years of Schooling (MYS) is set to be equal to the Skills in Literacy Adjusted Mean Years of

^{*} Population estimates by age, sex, and educational attainment come from the Wittgenstein Centre Data Explorer.

[†] Since PIAAC literacy test results are freely available for only 30 of the 36 OECD countries, six OECD countries had to be excluded in the calculation of the benchmark: Australia, Iceland, Latvia, Luxembourg, Portugal, and Switzerland.

Schooling (SLAMYS) for OECD. As a consequence, if a country's age-sex-education subpopulation group performs worse than the population-weighted OECD mean, its SLAMYS will be lower than its MYS; accordingly, for any country-specific age-sex-education sub-population group which scores better than the OECD mean, the opposite holds.

Formally, consider SLAMYSc,a,s,e as the skills-adjusted mean years of schooling for country c, age group a, sex s, and education level e in the base year 2015. Also, let MYS_{case} represent the respective mean years of schooling, and $MP_{c,a,s,e}$ the mean literacy score in PIAAC/STEP. Finally, consider MP^{*}a,s,e the mean performance of the benchmark (population-weighted OECD PIAAC mean literacy score) age-sex-education group. The skills-adjusted measure is defined by

SLAMYS_{c,a,s,e} = MYS_{c,a,s,e} x (MP_{c,a,s,e}/ MP^{*}_{a,s,e})
$$
(Eq. 1)
$$

In this way, we estimated SLAMYS for 44 countries for the base year 2015 ‡ , disaggregated by 5year age groups, sex, and four levels of educational attainment $\r\digamma$. SLAMYS for 36 countries are based on PIAAC data; for 8 countries on STEP literacy test results**. Data for MYS by country, age, and sex were retrieved from the Wittgenstein Centre (WIC) Human Capital Data Explorer (6).

Reconstruction of SLAMYS along cohort lines (1970-2015)

The estimation of SLAMYS for quinquennial years between 1970 and 2015 is based on the same rationale as provided by Eq. 1, but now including the time dimension t. The 2015 populationweighted OECD mean proficiency is held constant over time as the standard of comparison.

$$
SLAMYS_{c,a,s,e,t} = MYS_{c,a,s,e,t} \times (MP_{c,a,s,e,t}/MP^*_{a,s,e,2015})
$$
 (Eq. 2)

Since large-scale assessment tests of adult literacy were only introduced in the 1990s for a handful of countries, we had to follow a different approach to estimate SLAMYS for several decades. Therefore, time-series estimates for SLAMYS rest on the reconstruction of $MP_{c,a,s,e,t}$ along cohort lines, based on observed age effects from countries where MP_{caset} exists for more than one point in time.

Our cohort analyses are based on a pooled dataset of IALS (1994-1998) (7) and PIAAC (2011- 2017) (1) from which we built age cohorts^{††} to investigate the skill development of different age groups over a period of roughly 20 years. Ideally and when available, we used single age, which were then aggregated to 5-year age groups, depending on the year the surveys took place and the time lag between different surveys in each country. For example, in the United States surveys took place in 1996 (IALS) and 2014 (PIAAC); hence, our analysis follows a cohort, which was e.g.

[‡] Base year estimates within this paper refer to the year 2015. However, skills adjustments originate from any round of data collection of PIAAC cycle 1 (2011-2017) or STEP data collection between 2012 and 2016. As interpolation of skills data in single-year intervals to obtain 2015 values is not possible due to the nonavailability of more than one data points over time for most countries, PIAAC and STEP literacy test results provide the unmodified basis for 2015 SLAMYS despite small variations in time.

[§] We refrained from a more detailed disaggregation of education categories as test sample sizes would otherwise become too small.

^{**} As the target population of the STEP Skills Measurement Program is limited to urban adults, STEP results were adjusted to represent the entire country. Urban-rural corrections in literacy skills were derived from DHS, with the ratio between DHS literacy results of the total population and DHS literacy results of the urban population serving as the correction factor. For three countries (Bolivia, Ghana, and Kenya) country-specific DHS information was used; for five countries where no tested literacy data from DHS are available (Armenia, Colombia, Georgia, Ukraine, and Vietnam) corrections are based on regional averages.

^{††} Ideally, we would be able to follow the same individuals over their life course. However, as no true panel data are available, we take advantage of the fact that although we cannot observe the same people at different points in time, we are able to observe representative samples of the population at different points in time.

25-29 years old in IALS and 43-47 years old in PIAAC. A direct comparison between the two surveys is possible as they are based on the same scoring scale range and trend items from IALS were included in PIAAC, allowing literacy data to be linked for countries participating in both surveys.

At present, not enough data are available to expand these empirical cohort analyses to a global scale and a longer period. Therefore, we assumed a standard skill-age decay pattern by pooling all countries that participated in both IALS and PIAAC. Next, we adjusted for the mean score difference observable for the same age group in different years to separate the pure age effect – which is assumed to be more stable across countries and time – from the more context-sensitive cohort effect. These calculations were done for two broad education categories ('lower secondary or less' and 'upper secondary or higher') separately to account for potential differences in skill loss/gain due to attainment of formal education. Results imply that the skill loss due to age significantly differs by educational attainment levels and age (see Figure 2 in the Main Manuscript): while those with lower education tend to lose the highest share of their skills rather soon after leaving school, higher-educated people are still able to moderately gain skills up to the age of 35. After that, skills remain largely constant until the age of approximately 45 when cognitive skills eventually start decreasing. Sensitivity analyses of conducting the same kind of analysis for different countries separately confirmed that the patterns tend to be largely constant for different populations.

Based on these period-adjusted trends of cohorts over time, we derived an age- and educationlevel specific skill growth function over the life course, which is assumed to be constant for all countries and over time. This estimated percentage growth of skills over the life course is essential for the reconstruction of literacy test scores over time along cohort lines. More specifically, we take the scores of 60-64-year-olds tested in 2015 as the basis for the estimated score of 55-59-year-olds in 2010, adjusted by the percentage change due to the assumed reverse age pattern^{‡‡}. In this way and based on the country-, age-, sex-, and education-specific literacy scores from PIAAC and STEP, we estimated mean scores by 5-year age groups, sex, and four education categories from 1970-2015 for all 44 countries with empirical data available \S . SLAMYS were then calculated based on Equation 2, with the 2015 OECD average used as standard of comparison in all years. To aggregate SLAMYS by country and year, we weighted the scores based on the population size by age, sex, and educational attainment for each country and year, as retrieved from the WIC Data Explorer.

Adjustments for calculating SLAMYS using DHS data

As noted above, comprehensive adult literacy skills assessments are available only for 44 countries – most of them highly developed OECD members, thus unrepresentative of the world population. To include more empirical observations in our dataset, we decided to use tested literacy data from DHS for 63 countries that are more diverse in social and economic development.

As literacy assessments in DHS and PIAAC/STEP differ substantially in complexity and scope of the tests, further models had to be applied. We took advantage of the fact that four countries, i.e. Bolivia, Ghana, Kenya, and Peru, have participated in both PIAAC/STEP and DHS. Based on their concordance between PIAAC/STEP literacy proficiency and DHS literacy, adjustment scores were calculated for each population considering the proportion of the population that have tested

^{‡‡} For age groups for which we were not able to build cohorts for the whole or part of the reconstruction period (e.g. 60-64-year-olds in 2010 who were too old to be tested in 2015), we assumed age-, sex- and education-specific scores to be constant over time.

^{§§} While the empirical scores of the base year are disaggregated by age, sex, and four levels of educational attainment, the estimated standard age effect as well as the estimated skill growth function over the life course are only defined for two education categories. This crude disaggregation was found to be most consistent between countries. Given the different scores in the base year, reconstruction results, however, still differ between sexes and are available for four education categories.

literacy in DHS. As a result, PIAAC/STEP literacy proficiencies and consequently SLAMYS could be estimated for 59 additional countries. To validate our results, the ratios between the SLAMYS calculated using PIAAC/STEP results and SLAMYS estimated by DHS tested literacy data were checked for Bolivia, Ghana, Kenya, and Peru, i.e. the countries that have both sources of information. Results showed that due to conducting an easier literacy test DHS-SLAMYS estimates were consistently 25% higher than the estimates calculated by empirical PIAAC/STEP scores. For this reason, we made a further adjustment by multiplying the SLAMYS estimates derived from DHS tested literacy data by a factor of 0.8.

Calculation of SLAMYS based on prediction regression models

As explained above, SLAMYS based on PIAAC/STEP and DHS data were calculated for 103 countries. For the remaining countries SLAMYS were estimated using ordinary least squares regression models. In these models, skills-adjustment factors (SAF_{ct}) are predicted as dependent variables for every country c and time t (5-year intervals between 1970 and 2015). Based on these predictions, SLAMYS were calculated as the product of $SAF_{c,t}$ and MYS_{c,t} (as demonstrated in Eq. 1) $***$.

Several educational and demographic variables from various data sources were used as estimators in these models. First, to capture the basic literacy skills in the population, we included adult illiteracy rates ($AIR_{c,t}$) from the UIS dataset (8). Because of almost 100 percent literacy rates in most of the developed countries, this indicator is more useful to distinguish the differences among developing countries. Second, to capture the effect of schooling beyond basic education, the percentage of adult population having at least upper secondary educational attainment (aboveLS_{c,t}) from WIC Data Explorer (6) is included in the model. Third, country- and timespecific old-age dependency ratios (ODR_ct) (also from WIC Data Explorer) are included as a proxy for the state of the demographic transition in a country.

Additionally, to capture the effect of the level of quality of education, we included a 'Quality of Education Indicator'(QEL_ct) as an additional independent variable in the model. International data on quality of education are, however, limited and relatively new. Recently, the World Bank introduced a Global Data Set on Education Quality (9) (GDSEQ), covering harmonized data from international student assessments for 163 countries between 1965 and 2015. However, since these tests were only recently extended to more countries, the dataset contains many missing values – especially for earlier time periods. For this reason, we constructed a separate model that predicts QEI by using GDSEQ scores for available countries and time periods as dependent variables††† .

As QEI represents a measure of skills of young cohorts who are still in school at the time of assessment, its effect on skills of the working-age population can only occur with a certain time delay. Consequently, and to be demographically consistent, QEI is – if possible – considered in the model with a time lag of 25 years. However, since QEI data only go back to 1970, which would enable predictions only from 1995 onwards when using a 25-year time lag, we used QEI estimates for 1970 for the time periods from 1970 to 1995 and QEI estimates with a 25-year lag from 1995 onwards.

Dummy variables for the respective time periods are added as further independent variables. By using a stepwise regression procedure, we attain the final model for predicting $SAF_{c,t}$ in Eq. 3.

$$
log(SAF_{c,t}) = \beta_0 + \beta_1 aboveLS_{c,t} + \beta_2 ODR_{c,t} + \beta_3 AIR_{c,t} + \beta_4 QEL_{c,1970/t+25} + \Sigma_{t=1970-1990} \delta_t + \epsilon_{c,t}
$$
 (Eq.3)

^{***} The main reason to predict SAF instead of SLAMYS is to prevent any multicollinearity issues between MYS and other estimators.

^{†††} To predict QEI, dummies for UN detailed geographical regions for every country as well as two other quality of education indicators from UIS dataset (government expenditure on education as a percentage of GDP and pupil-teacher ratio in primary education) are used as predictors in the model. The model summary can be found in Table S1.

Using Equation 3, skills-adjustment factors were predicted for 185 countries (in 5-year time steps between 1970 and 2015). Finally, $SAF_{c,t}$ estimates were multiplied with MYS_{c,t} for the given country and time to calculate SLAMYS. The adjusted R^2 of 83 percent and a correlation coefficient of 0.984 between estimated and empirical SLAMYS point at a good model fit. A detailed regression model summary can be found in Table S2 and regression diagnostics are presented below.

Albeit SLAMYS were predicted for all 185 countries, our global SLAMYS dataset as presented in Table S3 is based on empirically sound data (estimates from PIAAC, STEP and DHS) whenever possible, and uses predicted values only if no empirical data are available. Table S4 provides a quantitative assessment of the data quality by summarizing on the continent level how many of the 1,850 data points (185 countries, 10 points of time) are based on PIAAC & STEP, how many on DHS, and how many are predicted from Eq. 3.

Comparison with student test results

To validate our results, we conducted a correlation analysis between SAF and student test scores. Figure S2 depicts the correlation between our estimated SAF and PISA reading scores (10). Given that our estimates cover the total working-age population, whereas PISA measures the skills of 15-year-olds, we used a time lag of 15 years, i.e. we compare PISA scores from the year 2000 with the 2015 SAF for all countries where both data are available. It is important to mention, however, that our prediction model already includes a Quality of Education Indicator as independent variable. As explained before, this indicator is based on the World Bank's Global Data Set on Education Quality (9), which covers harmonized data from international student assessments including PISA, TIMSS etc. Consequently, our measure indirectly already covers student tests; the high level of correlation (r=0.75) is therefore not surprising.

In addition, we ran age-specific correlation analyses (SAF of 15-19-year-olds vs. PISA scores of 15-year-olds in the closest year) for different years for those countries with empirically derived SLAMYS (see Figure S3). Given that SAF estimates for these countries are based on PIAAC/STEP data, these analyses represent in fact the very noticeable correlation between PISA and PIAAC/STEP scores. Relatively high correlation coefficients between 0.67 and 0.84 – depending on different years and combinations of countries – were found, suggesting that validity and reliability of the test results hold.

Regression diagnostics

There are several assumptions for OLS regression models. One of them is the normality of errors. The two plots at the top of Figure S4 shows the distribution of residuals for the model in Table S1. Residuals seem to be distributed normally.

In OLS regression models, predictors should not be correlated with residual terms. The two graphs at the bottom in Figure S4 are the scatter plots between the continuous numerical predictors in the model and residual terms. There seems to be no relationship between the predictors and residuals.

Furthermore, it is assumed that there is a linear relationship between predictors and the dependent variable in OLS models. The top left plot in Figure S5 does not show any pattern between fitted values and residuals. The horizontal line without a distinct pattern can be said to confirm the linearity assumption.

Another assumption is the constant variance of residuals which is called homoskedasticity. The two graphs on the left-hand side of Figure S5 show that there is not a relationship between fitted values and residuals which is indicating homoskedasticity. Moreover, Breusch-Pagan test for homoskedasticity also produces a p-value of 0.32 which means there is not enough evidence to reject the null hypothesis claiming homoskedasticity.

The Residuals vs Leverage plot positioned in the bottom right in Figure S5 helps to detect outliers in the model. In the plot, there are not any cases of outliers that exceed Cook's distance measure.

Finally, in OLS regressions models, independent variables should not have a high correlation (multicollinearity) among themselves. To check this assumption, Variance Inflation Factor (VIF) scores have been calculated. None of the predictor terms have a VIF score above ten and the mean of the VIF scores is below five which means there is no multicollinearity.

Assumptions for OLS regression are also tested for the model in Equation 3. The top two plots in Figure S6 show the distribution of error terms that indicate a nearly normal distribution confirming the normality assumption. The rest of the plots in Figure S6 are scatter plots between the predictors and residuals. It can be said that there are no correlations between them.

However, the decreasing variance for residuals in some of the plots in Figure S6 indicate that there might be some level of heteroskedasticity. This pattern is also visible in the top left plot in Figure S7. Variance for the top end of the fitted values is smaller. Moreover, studentized Breusch-Pagan test produces a p-value less than 0.05 which shows evidence to reject the constant variance of residuals hypothesis.

The violation of the homoskedasticity assumption may lead to an underestimation of standard errors. Zeileis (11) offers a procedure for a robust estimation of standard errors^{##}. While the coefficients in Eq. 3 are tested with these new standard errors, the significance of the predictor terms have not changed at a 95% confidence level. Moreover, we are more interested in the estimation of the dependent variable than predicting the effects of specific independent variables. Since heteroskedasticity does not lead to a bias in the coefficient estimates (12), it should not affect our SLAMYS estimates.

The bottom right plot in Figure S7 shows that all observations are within Cook's distance which means there are not any cases of outliers that are highly influential. Lastly, VIF scores for all predictors are less than five which confirms that there is no multicollinearity.

The above tested regression diagnostics deal with the assumptions about the error related to the dependent variable. In OLS regression, it is assumed that independent variables are measured without error. However, in most cases independent variables are measured with some degree of error. Errors in variables (EIV) models offer a solution for the potential cases of measurement errors in independent variables (13). In our case, our numerical predictors may contain a degree of measurement error but it is not possible to know the exact levels. Using eivtools package in R (12) we calculated the adjustment scores with varying amounts of reliability for the numerical predictors in Eq. 3 one by one to get an understanding of the sensitivity of our model to measurement errors. For each variable we run three different models with reliability values 0.95, 0.90 and 0.85. Figure S8 plots the fitted adjustment factors of the main model (y-axis) against the fitted adjustment factors from various errors in variance models (x-axis). The figure shows that the fitted values are robust to decreases in reliability scores.

A note on measurement errors in Mean Years of Schooling

Finally, albeit the focus of this paper is on measuring skills, it is important to not completely neglect the potential of measurement errors in the quantitative dimension of our indicator, i.e. potential measurement errors in MYS. Krueger and Lindhahl (14) thoroughly discussed the problem of estimating the extent of measurement error in cross-country data on average years of schooling, concluding that measures tend to be particularly noisy when based on frequently mismeasured enrollment rates. However, also in more advanced measures of educational attainment, such as the widely used Barro and Lee dataset (15), that draws on survey- and census-based estimates reported by UIS, errors in measurement are inevitable given the doubtful quality of UIS enrolment rates in many countries.

^{‡‡‡} Robust standard errors were calculated via "sandwich" package in R.

Data on MYS used in this paper are retrieved from the WIC Data Explorer (6) which provides an advance over other existing international measures of educational attainment on several grounds: i) the WIC methodology is based on original data (as opposed to data compiled by other institutions, like UIS or EUROSTAT); ii) all data are thoroughly harmonized using available ISCED mappings in order to achieve better comparability and avoid flaws in the primary data; and iii) estimates rely on assumptions and rules, and the consistency of these over countries is important. The detailed methodology for the WIC global MYS estimates can be found elsewhere (16, 17).

Given the before-mentioned advances, variations in the MYS between WIC and Barro-Lee as well as between WIC and UIS are not surprising and can be traced back to different types of source data, different definitions on the educational categories, flaws in the input data, different procedures employed in estimation of the educational shares, and differences in the estimation of durations of schooling for incomplete levels. Consequently, and given the consistency and comparability across countries, MYS estimates used in this paper are unlikely to distort our SLAMYS results.

DATA AND CODES AVAILABILITY

Data and codes used to generate the results are available in the GitHub repository, [https://github.com/clreiter/WIC-Skills-Adjusted-Human-Capital/tree/final-preparation-repository.](https://github.com/clreiter/WIC-Skills-Adjusted-Human-Capital/tree/final-preparation-repository)

Figure S1. Correlation between PIAAC mean literacy and numeracy scores (left) and PIAAC mean literacy and problem-solving in technology-rich environment scores (right), by age, sex and country, all PIAAC countries.

Figure S2. Correlation between estimated SAF 2015 and PISA mean reading score, by country (all countries with available data are considered).

Figure S3. Correlation between SAF for 15-19-year-olds (based on PIAAC/STEP data) and PISA mean reading score, by country (all countries with available data are considered).

Figure S4. Residual plots for the model estimating QEI scores.

Figure S5. Residual plots for the model estimating QEI scores.

Figure S6. Residual plots for the model estimating SLAMYS (Eq. 3).

Figure S7. Residual plots for the model estimating SLAMYS (Eq. 3).

Figure S8. Fitted adjustment factors with different reliability values in Errors in Variance model (xaxis) vs fitted adjustment factors from the main model (Eq. 3) (y-axis).

Table S1. Model summary and estimated coefficients for estimating QEI. Dependent variable is QEI.

Table S2. Model summary and estimated coefficients for Eq. 3. Dependent variable is log(SAF_{c,t}). Base period for year dummies is Year(2015-19).

Table S3. Mean Years of Schooling (MYS) and Skills in Literacy Adjusted Mean Years of Schooling (SLAMYS) for 185 countries (1970-2015) ranked by their 2015 SLAMYS score.

Table S4. Quantitative assessment of data quality. Numbers represent how many of the SLAMYS scores were calculated using which data source/method, by year and continent.

SI References

- 1. OECD, Programme for the International Assessment of Adult Competencies. Data available at: https://webfs.oecd.org/piaac/puf-data/.
- 2. The World Bank, The STEP Skills Measurement Program. Data available at: https://microdata.worldbank.org/index.php/catalog/step.
- 3. USAID, The Demographic and Health Surveys (DHS) Program. Data available at: https://dhsprogram.com/data/available-datasets.cfm.
- 4. C. Reiter, *et al.*, The Demography of Skills-Adjusted Human Capital. *IIASA Working Paper* **WP-20-006** (2020).
- 5. E. A. Hanushek, L. Woessmann, The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature* **46**, 607–668 (2008).
- 6. Wittgenstein Centre for Demography and Global Human Capital, Wittgenstein Centre Data Explorer (Beta) (2018). Data available at: http://dataexplorer.wittgensteincentre.org/wcde-v2/.
- 7. U.S. Department of Education. National Center for Education Statistics., International Adult Literacy Survey (IALS). Data available at: https://nces.ed.gov/surveys/ials/data.asp.
- 8. UNESCO Institute for Statistics (UIS), UIS.Stat. Data available at: http://data.uis.unesco.org/.
- 9. N. Altinok, N. Angrist, H. A. Patrinos, Global data set on education quality (1965-2015). *World Bank Policy Research Working Paper* **No. 8314**, 1–78 (2018).
- 10. OECD, Programme for International Student Assessment. Data available at: http://www.oecd.org/pisa/data/.
- 11. A. Zeileis, Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software* **11**, 1–17 (2004).
- 12. Frost, Jim, *Regression analysis: An intuitive introduction for using and interpreting linear models* (2019).
- 13. W. A. Fuller, *Measurement error models* (Wiley, 1987).
- 14. A. B. Krueger, M. Lindahl, Education for growth: Why and for whom? *Journal of Economic Literature* **39**, 1101–1136 (2001).
- 15. R. J. Barro, J. Lee, *Education matters: global schooling gains from the 19th to the 21st century* (Oxford University Press, 2015).
- 16. M. Potančoková, S. KC, A. Goujon, "Global estimates of mean years of Schooling: A new methodology" (International Institute for Applied Systems Analysis (IIASA), 2014).
- 17. M. Speringer, *et al.*, Global Reconstruction of Educational Attainment, 1950 to 2015: Methodology and Assessment (2019).