Estimation of education quality: harmonization and analysis of inequalities

Claire A Henson

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Public Health

University of Washington

2023

Committee:

Emmanuela Gakidou

Terje Andreas Eikemo

Luisa Sorio Flor

Program Authorized to Offer Degree:

Global Health

University of Washington

Abstract

Estimation of education quality: harmonization and analysis of inequalities

Claire Henson

Chair of the Supervisory Committee:

Emmanuela Gakidou

Department of Health Metrics Sciences and Global Health

Background: Education is key social determinant of health, the effects of which are related to the attainment of subject-based skills and the application of these skills to improve the health and well-being of individuals and economies. Previous analyses have shown that there is stark inequality in access to quality education, but estimates are hindered by a landscape of differing assessments of student achievement making comparability of education quality across locations difficult. We therefor aim to produce globally standardized estimates of education quality and describe inequalities in access to high quality education. **Methods**: International assessments of mathematics achievement were harmonized to a single assessment scale using a newly applied network meta-analysis strategy. Inequalities in harmonized estimates of educational quality were identified across region, country, and socioeconomic status (SES) using absolute and relative measures of inequality. **Results**: There are significant differences in educational quality across locations not explained by educational attainment alone. Individual-level variables related to household SES are consistently associated with higher quality education. **Conclusion**: Inequalities in access to quality education have wide-ranging implications. Achievement levels should be interpreted in tandem with estimates of total inequality to ensure that improvements in achievement happen in tandem with reductions in total inequality.

***Land Acknowledgement**: We acknowledge the Coast Salish peoples of this land, the land which touches the shared waters of all tribes and bands within the Duwamish, Puyallup, Suquamish, Tulalip, and Muckleshoot nations.*

Introduction

Educational attainment, which refers to the years of education completed by an individual, is a well-established social determinant of health, both morbidity and mortality, with implications beyond a single generation.[1–4] It is well understood that there are a number of direct and indirect pathways in which education influences health.[3] While not all pathways between education and improved population health are related to student achievement, many have been linked directly to school curriculum, levels of learning, and practical skills such as basic mathematics and reading proficiency.[5] These skills allow increased access to well-paying occupations, which in turn provide social and psychological benefits, and also improve knowledge of health risks and behaviors.[6]

While inequalities in educational attainment itself undoubtedly contribute to inequalities in health, since the 1980s more attention has been given to the quality of this education and the ways the quality of such education impacts learning, the attainment of skills, and subsequently population health[7] and economic growth.[8] These impacts are reflected in the focus on both educational attainment and quality in Sustainable Development Goal (SDG) number 4, which seeks to guarantee a quality education for all children.[9] Quality education, as defined by the United Nations Educational, Scientific and Cultural Organization, is a multifactorial goal including issues such as availability of equipment and school infrastructure, quality of the teaching force, and other factors which have been shown to play varied roles in learning outcomes.[7,10] Regardless of the specific mechanisms driving educational quality, there are marked differences in learning and achievement levels both between and within countries.[11] Here, we utilize measures of student achievement or 'learning' to quantify differences in educational quality.

Measuring and estimating learning requires the use of assessments to ascertain the proportion of a population achieving a given proficiency by subject. There are several international and regional organizations assessing learning, in addition to many national assessments, contributing to a landscape of assessments with differing participants and scoring criteria. Due to methodological and logistical challenges in harmonizing these measures of learning there are few globally standardized estimates, although many ways exist of tracking student outcomes. One currently available global estimate of learning utilizes linkages between alternate learning assessments to construct a globally applied scale, yet the method is subject to limitations related to the presence and configuration of such linkages and the availability of data.[12]

Absolute measures of educational quality, such as the percent meeting or exceeding proficiency, may be useful in marking progress, however these one-dimensional summaries mask important inequalities in the distribution of learning between and within groups and represent a missed opportunity to encourage the equitable access to quality education. Furthermore, an analysis of inequalities in learning can provide actionable targets for intervention, for instance indicating whether the structural or individual factors are most relevant for reducing inequalities and improving equitable learning outcomes in each location.

The overarching aim of this work was to produce globally standardized estimates of educational quality which may contribute to our understanding of the work that remains to meet SDG goals for equitable quality education. These estimates may be used in conjunction with estimates of educational attainment to better predict and address patterns of mortality and health burdens. To accomplish this, our first objective was to implement novel methods to harmonize assessments of educational quality in one subject and demonstrate the applicability of these methods to produce reliable estimates of quality. Our second objective is to analyze trends and drivers of inequality in educational quality.

Methods

Data on Learning Outcomes

Learning and skills development in both mathematics and reading has been linked to improved social and economic outcomes on a macro level.[13] For this analysis only mathematics assessments were considered due to the particular geographic coverage and overlap between identified assessments, although the methodology presented here may be applied for any subject.

Mathematics assessment series were considered if they met the following criteria: geographically representative at the national level, coverage of years between 2000 to 2020, assessment of children grades 4 to 11 or ages 10 to 16, have publicly accessible microdata, and microdata contain information on household socioeconomic (SES) and demographic status or additional predictors of inequality (see inequalities in learning section below). Finally, the selected assessments must have sufficient overlapping populations of the same country-grade groups, with overlaps between all combinations of assessment, to allowing the application of the Network Meta-Analysis (NMA) crosswalk.

Trends in International Mathematics and Science Study (TIMSS), the Program for International Student Assessment (PISA) and the Multiple Indicator Cluster Study (MICS) foundational learning (FL) module assessments were selected for inclusion in this analysis. Each assessment assesses mathematics proficiency at different levels of education. TIMSS and PISA are given to students in-school, while the MICS foundational learning module is a household-based assessment. In total, the assessments include 111 countries, shown in Figure 1, although each has varied geographic and temporal coverage. While PISA and TIMSS assess predominantly European, high-income, and Asian countries, MICS is conducted in low and middle income (LMIC) locations. Additionally, while TIMSS and PISA have regular assessment coverage throughout the 21st century, the implementation of MICS foundational learning module in MICS surveys only began in 2017. Details for the coverage of each assessment are provided in appendix figure 1 and table 1.

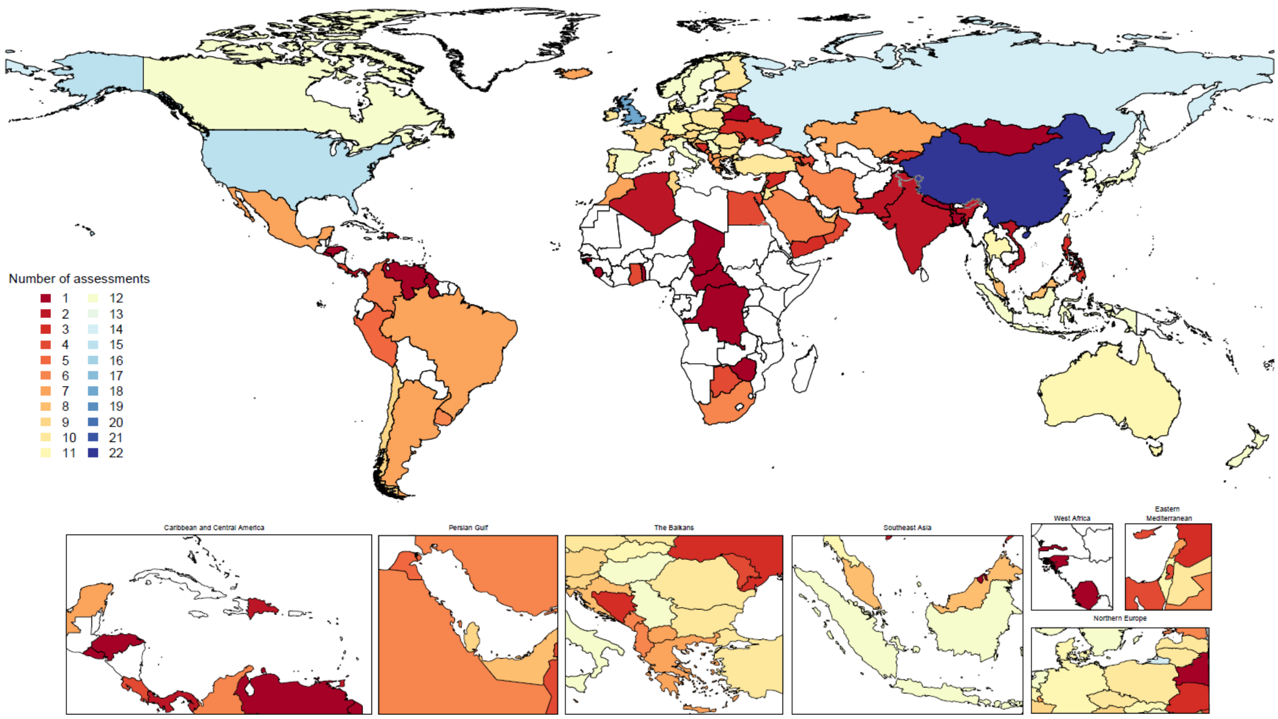


Figure 1. Coverage of assessments completed by TIMSS, PISA, and MICS. Color represents the number of assessments per country after combining datasets.

Learning Assessment Preparation and Harmonization

Microdata from each assessment were extracted to a standardized template. TIMSS and PISA utilize plausible values (PVs) to capture learning outcomes, which are designed to produce a continuous and normally distributed signal of learning and capture uncertainty in estimates of learning. Plausible values have no definitive range by nature of their construction and are intended to allow researchers to analyze learning outcomes while incorporating sometimes substantial measurement error[14], but do vary consistently between assessments. TIMSS PVs range from 60 to 800, and PISA PVs from 0 to 1000.

The MICS Foundational Learning Assessment, contrary to other continuous assessment outcomes, is constructed to assess for foundational mathematics ability on a pass/fail scale and to be employed in resource-limited settings. To achieve the aim of comparability and harmonization, MICS outcomes were transformed to a normally distributed signal of mathematics achievement utilizing a logistic regression incorporating individual characteristics, with additional details found in the appendix.

After extraction and standardization, learning assessment scores were rescaled to between 0 and 1 using the maximum and minimum values for each assessment. Samples of less than 30 students per country-grade-sex group were excluded, as were assessments of grades less than 4 and greater than eleven. MICS, as a household-based assessment, is given both to children attending school and out of school children, and only scores from children in school were kept improving the comparability with PISA and TIMSS.

Network Meta Analysis (NMA) Crosswalk and Adjustment

Matched populations, defined as a country-grade where two or more differing assessments were given regardless of year of assessment, were identified between the three assessments. Multiple overlaps were allowed; for example, a single MICS country-grade assessment could overlap with multiple different TIMSS covering that country-grade. Description of the characteristics of matched pairs can be found in table 1. After identification of overlapping groups between differing assessment types, we assigned learning scores a decile ranking for each survey-country-year-grade group within an assessment, and data were collapsed to produce mean and standard error for each survey-country-year-grade-decile group. While equipercentile linking between assessment families is an established method to harmonize learning assessments[12], here observations were collapsed into deciles to improve the stability of subsequent modeling steps.

Table 1. Matched pairs utilized in crosswalk model

Assessment-pair	Number of matched pairs (number of countries)	Number of matched pairs including multiple overlaps	Average difference between matched pairs (years)
TIMSS-PISA	40 (40)	540	6
TIMSS-MICS	8 (7)	18	10
PISA-MICS	8 (5)	29	8.7

We then utilized the CrossWalk package model in R to adjust MICS and PISA assessments to the TIMSS scale.[15] Here, 'crosswalk' references the process of adjusting PISA and MICS (alternative) assessment scales to the TIMSS (reference) scale. TIMSS was selected for the reference scale to increase the comparability of results with existing literature, although any scale could be selected. The CrossWalk package allows for the application of a network meta-analysis, which ensures that not only is the information from direct overlaps used in the adjustment process (alternative to reference), but also from the indirect comparisons (reference to reference). The model allows for the automated trimming of outliers, and 10% trimming was used.

We then took the difference between matched (location-grade-decile) pairs in logit space and used this as the dependent variable in a logistic mixed effects model. We calculated the logit-difference as

$$\delta_{ijd} = logit\left(y_{ijd}^{alt}\right) - logit\left(y_{ijd}^{ref}\right)$$

The model included two covariates; first, the difference in matched observations years, with the assumption that observations closer in time should be more reflective of true differences across scale, and second, the decile of the matched pair to account for differences in the relationship between scales by percentile. Year-difference and decile are incorporated as fixed effects, with random effects by country-grade:

$$\delta_{ijd} = \beta_0 + \beta_d + \beta_y + u_i + \epsilon_i$$

Where:

- δ_{ijd} is the modeled difference in logit space
- β_0 is the intercept
- β_d is the coefficient on score decile
- β_y is the coefficient on year difference between matched pairs
- u_i are random effects by country-grade
- ϵ_i is non-specific error

To obtain the estimated mathematics outcome for a country-grade-decile alternative, we applied the predicted adjustment factor to original microdata by subtracting it from the original observation in logit space, with in-sample year-difference set to 0. Predicted outcomes were rescaled to the TIMSS scale of 60 to 800.

Identifying Inequalities in Learning

Variables considered for extraction included factors that could potentially drive inequalities in health or were available across the three assessments. The selected variables were grouped into structural or individual categories. Structural variables are associated with the student's environment and included whether the school was in an urban or rural area, location (country or subnational unit), and region. Individual variables are associated with household or individual characteristics, and include three binary variables (student's sex, whether the primary language spoken at home is the language in which the test is given, and preschool attendance), and four categorical variables (number of books in home, wealth-score, and maternal and paternal education levels). Wealth scores were calculated as the proportion of material possessions that a given student has at home, with the total possible number varying by assessment and location, and subsequently grouped into wealth quantiles. Mother's and father's education was standardized to International Standard Classification of Education (ISCED)[16] categories corresponding to no education (ISCED 0), completion of primary education (ISCED 1) completion of lower secondary education (ISCED 2) completion of upper secondary education (ISCED 3), and attainment of any education higher than upper secondary (ISCED 4+). Mothers' education was available for all assessments, fathers' education was only available for TIMSS and PISA. Details for the availability of these variables across assessments may be found in Appendix table 1. Differences in mean scores between groups were assessed using Welch's two sample t-test with adjustment for multiple comparisons where relevant.

Inequality metrices vary in their ability to describe underlying inequalities, depending on the structure and distribution of the value in question. Here we focused on utilizing measures of equality of opportunity and of condition in tandem. Measures of equality of condition may be used to quantify the absolute variation within a group, or the dispersal of educational quality regardless of circumstances – these metrics include Gini and Average Absolute Interpersonal Difference (AAID), coefficient of variation (standard deviation divided by mean), the cumulative distribution function, and more.[17,18] Equality of condition was assessed by different structural dimensions (region, country) to ascertain the relative inequality within and across locations. Measures of equality of opportunity quantify the relationship between education quality and specific circumstances, such as the variables above. Selected metrics include differences (mean), interquartile range (IQR), AAID, and others. All calculations were applied to adjusted scores on the TIMSS scale, and absolute differences thus reflect differences in mathematics proficiency using the TIMSS criteria. Mathematics outcomes were compared to mean educational attainment estimates for ages 10-14 using covariate estimates from the Global Burden of Disease.[19]

Results

Learning Assessments

Learning assessment from TIMSS covered years 2003 to 2018, and from PISA years 2000 to 2018. The first MICS learning assessment was done in MICS round six and covers years 2017-2019 and 18 country years. Data are predominantly from high-income and central and eastern European regions, with 44% and 20% of observations, respectively. 2.3% of observations are from Sub-Saharan Africa, and less than one percent from South Asia [Figure 1]. Further details on data included in the analysis as well as the proportions of each characteristic in the input data by assessment are shown in appendix table 1 and appendix table 2.

Network Meta-analysis Crosswalk

We identified fifty-six unique country-grade populations to utilize in crosswalk modeling, for a total of 559 matched pairs when utilizing multiple overlaps. Table 1 describes the characteristics of these matched pairs, and Table 2 provides the resulting coefficients of the crosswalk model.

Table 2. Parameter estimates for CrossWalk

Survey	Covariate	Beta	Beta SD	p-value
UNICEF MICS	Intercept	1.157039	0.004728	p<0.001
UNICEF MICS	Year-difference	0.002438	0.000186	p<0.001
UNICEF MICS	Decile	-0.003464	0.000026	p<0.001
OECD PISA	Intercept	-0.258772	0.002473	p<0.001
OECD PISA	Year-difference	-0.001168	0.000040	p<0.001
OECD PISA	Decile	-0.003425	0.000007	p<0.001

Total Inequalities in Learning

Figure 2 displays two measures of total inequality in mathematics outcomes, AAID, and Gini. Utilizing Gini, the countries with the greatest inequality (high Gini coefficient) in mathematics outcomes are those with a substantial proportion of students scoring low on mathematics assessments and with only a small proportion scoring high and are predominantly in Sub-Saharan Africa and South Asia. Countries in High-income and Central European locations have the least inequality (low Gini coefficients), as countries these regions showed the highest mean learning scores on aggregate. The average Gini coefficient across super-regional groupings ranged from 0.16 in Sub-Saharan Africa, to 0.08 in High-income countries.

The relationship between mathematics outcomes and total inequality, and conclusions about which societies have the greatest total inequality, are different when leveraging AAID. AAID is related to the Gini coefficient, but is formulated to capture differences between individuals rather than within societies.[20] Thus, it highlights the greater total inequality in learning outcomes in several North African and High-Income locations, indicating that large proportions experience high quality education, but large proportions also experience low quality education. However, South Asia and Sub-Saharan Africa regions still displayed the greatest within-country inequality when utilizing AAID, with an average of 110 and 103.1 respectively, and the High-income region the least total inequality, with an AAID of 91 [Figure 2]. These values are in the scale of the TIMSS test and represent the within-country difference in mathematics scores between any two given individuals. The average country-level AAID across all locations was 93.1.

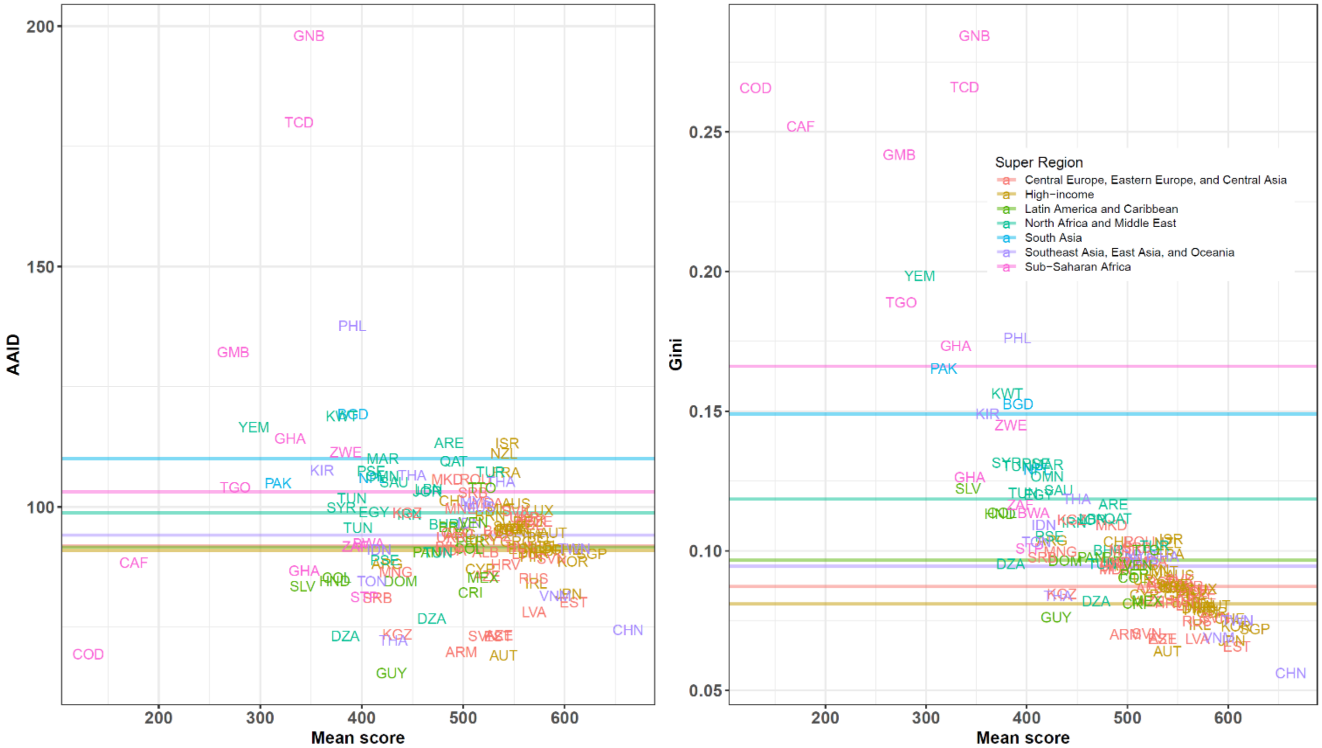


Figure 2. Relationship between learning scores (x axis) and inequality metrics of AAID (left) and Gini (right). Color indicates super-regional membership. Horizontal lines show average inequality by metric and super-region. Points are labeled using country-codes, and only most recent year of data is shown. The choice of inequality metrics influences conclusions of country-level inequalities.

To better understand factors driving the differences in educational quality we compared the mean educational attainment for 10–14-year-olds and average mathematics outcomes by location [Figure 3]. These variables had a correlation coefficient of 0.54. However, levels of learning vary widely even within countries achieving the goal level of educational attainment for this age group (5-8 years of schooling), indicating that factors related to educational access and attainment alone do not explain differences in mathematics outcomes.

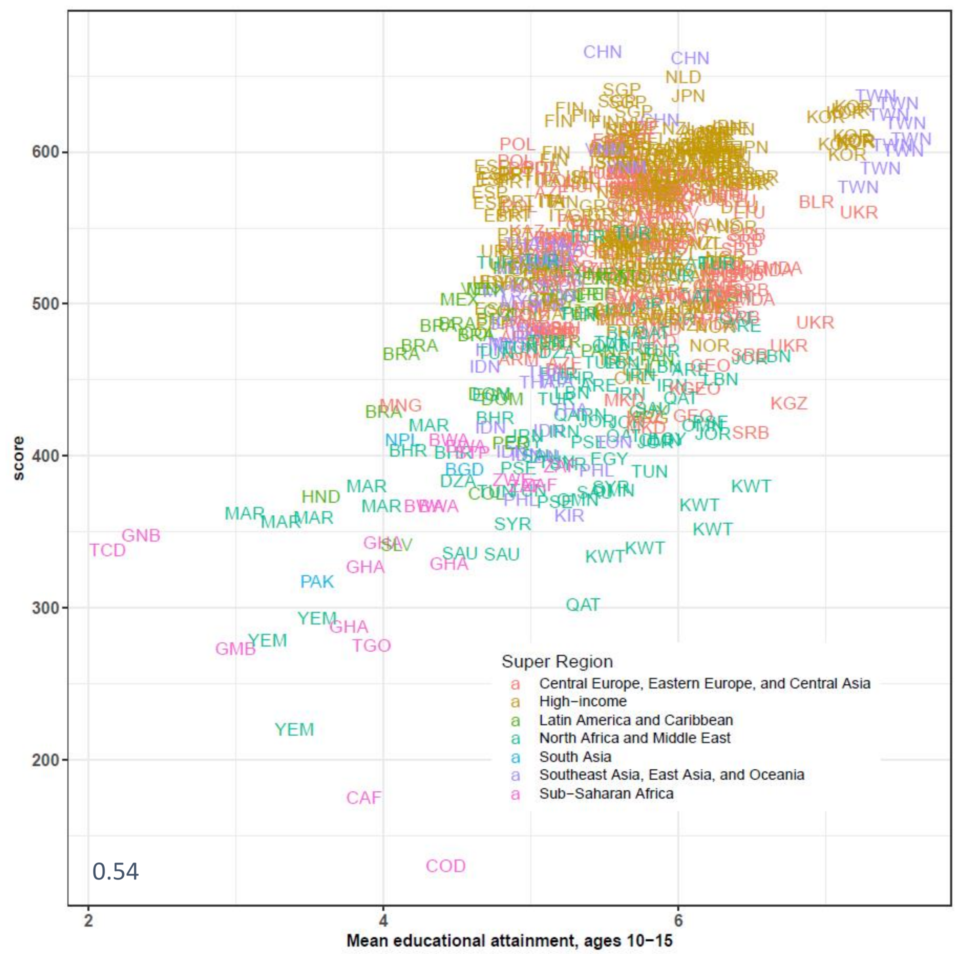


Figure 3. Relationship between country mean-educational attainment for ages 10-15 (x axis) and average learning scores per country-year (y axis). Color indicates super-region membership. Correlation coefficient = 0.54.

Group-level differences in Learning

There are stark differences in mathematics outcomes across the selected individual-level variables. Across locations, significant differences exist in mathematics outcomes relative to sex[$p<0.001$], public school attendance [$p<0.001$], preschool attendance [$p<0.001$] rural/urban status of the school [$p<0.001$], testing language (relative to language spoken at home) [$p<0.001$], number of books in students home [$p<0.001$], wealth quantiles [$p<0.001$], and maternal [$p<0.001$] and paternal [$p<0.001$] education levels. Males scored significantly higher than females in aggregate, and factors associated with higher socioeconomic status are correlated with higher mean levels of learning across all dimensions [Figure 4]. For categorical outcomes, the comparisons above reflect the difference in mean scores between 1st and 4th wealth quantile, between parental education levels of ISCED level 0 to level 4+, and number of books in home of 0-10 to >100.

Across all dimensions within-group inequality decreased as average learning score increased, in a pattern following country-level mathematics achievement and inequality. This relationship is consistent with the skewness of each of these distributions; most learning distributions are left skewed, so it follows that if the country or dimension (e.g., no preschool attendance) is associated with a lower mean score, it will demonstrate greater inequality. There are a few exceptions to these left-skewed trends, with countries from Latin-American and Caribbean standing out where the mathematics outcomes trend center or right skewed, and there is less variability in mathematics outcomes by group.

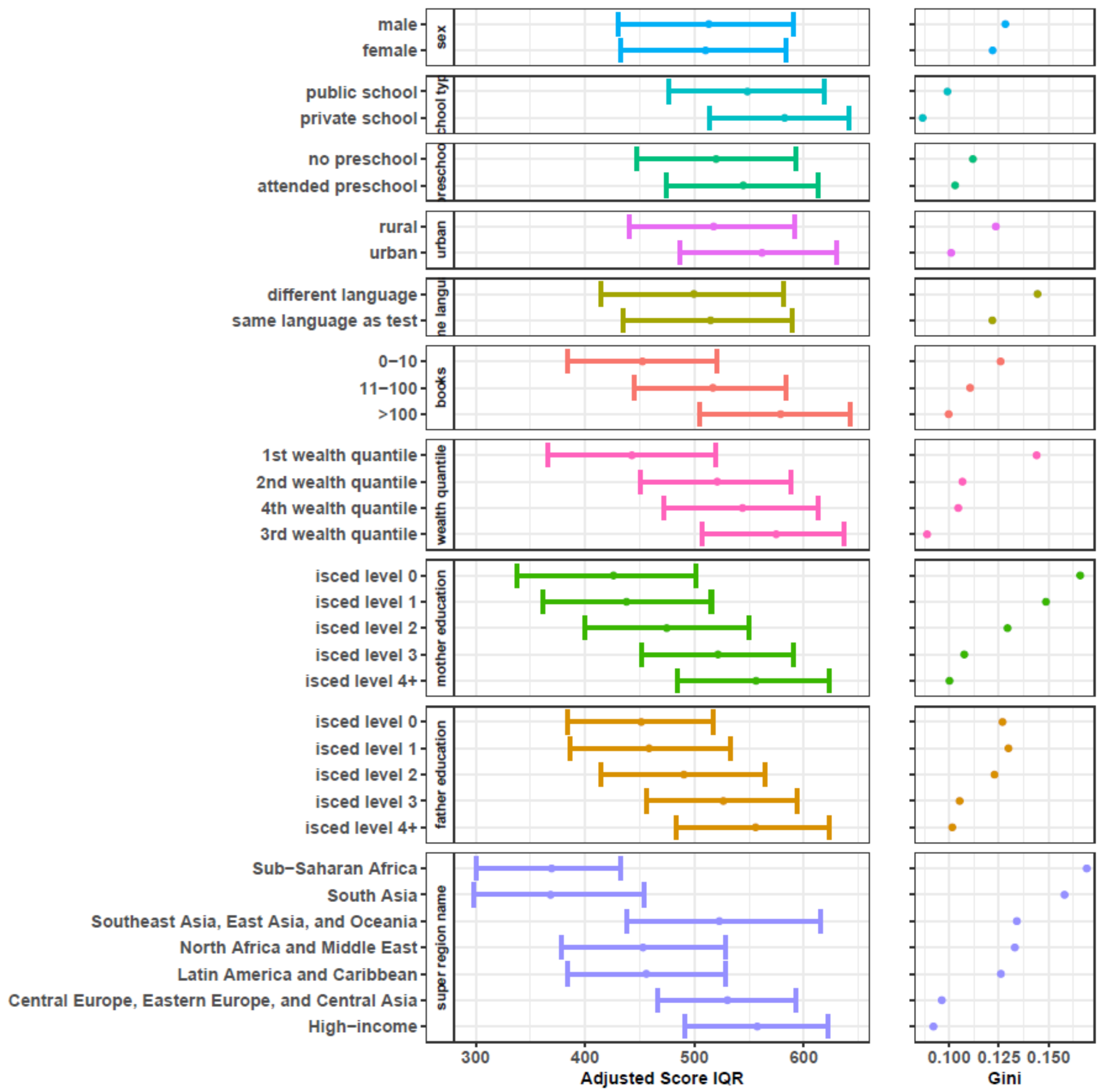


Figure 4. Relationship between inequality dimensions and mathematics outcome inter-quartile range (left) and Gini coefficient (right). Y axis is ordered according to the Gini coefficient of each outcome.

Discussion

We produced harmonized estimates of educational quality in mathematics using a previously untried methodology, leveraging direct and indirect linkages between assessments. This method and the resulting harmonized dataset allow policy makers to use both absolute and relative inequality metrics to describe education quality and better guide policies improve equitable access to quality education (SDG 4.1).

While differences in mathematics proficiency are wide-ranging across regions, reflecting broad differences in educational expenditure, national education policy and structural differences, equally important are the stark differences in educational quality across individual countries. By placing these inequalities in context of the TIMSS assessment scale we can understand the actual differences in skills which these values reflect. For instance, the locations with the least within-country inequality had levels of AAID that reflected an average difference in achievement that would not necessarily change the outcome of a student's assessment of proficiency (less than 60 points on the TIMSS scale). In contrast, locations with the greatest within-country inequality had levels of AAID that would reflect the difference between 'advanced' levels of mathematics, and not meeting proficiency. It is essential that these within-country inequalities are accounted for when benchmarking progress towards SDG 4.1, and that improvements in overall performance, where seen, are not masking increases in total inequality.

Many of the factors influencing access to quality education appear to be associated with educational attainment, for example in Yemen (with low attainment and quality) versus Taiwan (high attainment and quality). However, there is much to explore in where these two values depart from each other. For instance, there are countries with low average educational attainment with relatively high mathematics scores, such as Mexico and Brazil, while scores in Kuwait are lesser than countries with similar educational attainment levels. These departures indicate that variation in education quality is influenced by different factors than educational access and attainment. Additionally, while access to and attendance in school is a prerequisite to higher learning outcomes, there is also evidence that quality education systems (measured by student-teacher ratios) influence attainment in turn.[21] These large variations in within-country access to quality education should be considered by researchers estimating the relationship between education on population health and economic growth.

Differences in mathematics outcomes were significant across all dimensions of inequality. It is well documented that higher socioeconomic status is correlated with higher learning outcomes, and so the strength of the effect of these SES variables, such as parental education and family wealth, is not surprising. These results confirm the findings of other analyses demonstrating household SES to be one of the top predictors of mathematics achievement[22], although the effects of particular SES characteristics have also been shown vary across environments.[23] Efforts to reduce inequities require a context specific understanding, especially when country-level patterns in inequality depart from the norm. For example, despite the wide-ranging acceptance of the relationship between parental education and child health and achievement,[24,25] we found no significant differences in mathematics outcomes between parental education below primary and completion of secondary in five locations (Cyprus, Honduras, Kuwait, Russia, Ukraine).

These estimates of inequality may be used to target polices to improve access to equitable high-quality education to the locations falling behind SDG goals. Measures of student achievement alone may mask important trends in total inequality, as increasing achievement may not indicate decreasing inequality, and thus policy makers should be encouraged to interpret such measures in tandem with measures of total variation, such as AAID. While it may be a challenge to identify the drivers of educational quality that are most impactful in each setting (student-teacher ratio, curriculum support, school resources and materials), the particular landscape of inequalities – whether there is variation by location, socioeconomic status, or otherwise – can provide a starting point for which to begin.

There are several limitations to this analysis. First, in applying the NMA crosswalk using country-grade-decile-linkages, we assume that mathematics outcomes follow a normal distribution and that the matched populations are equivalent. While the considerable number of overlapping populations between assessments TIMSS and PISA are likely to subvert the effects of any outliering pairs, the reduced number of overlaps between MICS and each assessment may contribute to variability in the crosswalk outcome. However, with automated trimming of outliers and increasing availability of MICS foundational learning this limitation will become less significant.

Additionally, because PISA and TIMSS are given to in-school children only we elected to use only MICS in-school students in the analysis. Thus, access to and completion of education is a dimension of inequality that is not captured in this analysis, and the total variation in outcomes shown here is under reported for all locations, as is the actual level of mathematics achievement by country. This underestimation will be greater in locations with high proportions of out of school children such as Pakistan and much of Sub-Saharan Africa and should be interpreted in context of this limitation.

Finally, in using learning assessment to assess educational quality, we assume standardized assessments can capture real differences in the quality of education. There are many factors influencing the outcomes of these assessments that may not be related to educational quality, including the timing of the assessment in the school year, the circumstances of the assessment, and ability to sample a representative population. This limitation applies to all estimates of educational quality relying on learning assessments and are another reason one-dimensional metrics such as percent at proficiency may mask important trends in educational quality.

Investments in equitable high-quality education are a proven way to improve population health and wellbeing. While absolute differences in learning may be stark, we encourage governments and educational organizations to target efforts to improve educational quality using an understanding of the dispersal of learning outcomes within groups, to ensure that improvements in education quality are accessible by all. The benefits of a high-quality education are astounding on both the individual and societal level and should be available to all learners.

Appendices

Supplementary Methods

MICS Standardization:

The sixth round of the Multiple Indicator Cluster Survey (MICS6) Foundational Learning (FL) module evaluates whether the test taker meets the minimum proficiency in mathematics. The MICS6 FL module, although narrower in scope than other major learning assessments, represents an important source of learning information, due to its wide coverage of low- and middle-income countries (LMICs). Foundational-learning level achievement is not provided on the individual-microdata level. Thus, we used the provided raw achievement-test scores to calculate 'pass/fail' for reading and mathematics, following MICS6 tabulation plan7. We then retained the component scores for mathematics proficiency, as well as the foundational learning indicator, for each individual test taker.

This binary pass/fail structure employed in MICS is inherently different from other learning assessments, where the resulting scores typically follow a continuous normal distribution. To incorporate MICS learning into the rest of the estimates we implemented a logistic regression to produce a normally distributed signal of math ability from the binary pass/fail results. The estimated distribution indicates the probability of any individual passing the module. The model was run separately by test subject with fixed intercepts on gender of learner, current school enrollment status, and education of the mother, random intercepts on learner age, and random intercepts on primary sampling unit, nested within strata and country, to predict a pass or fail outcome.

Trends by each sampling unit u, strata s, country c, enrollment status e, gender g, age a, and mothers education m, the probability of a pass (P,c,s,u,e,g,a,m) was estimated:

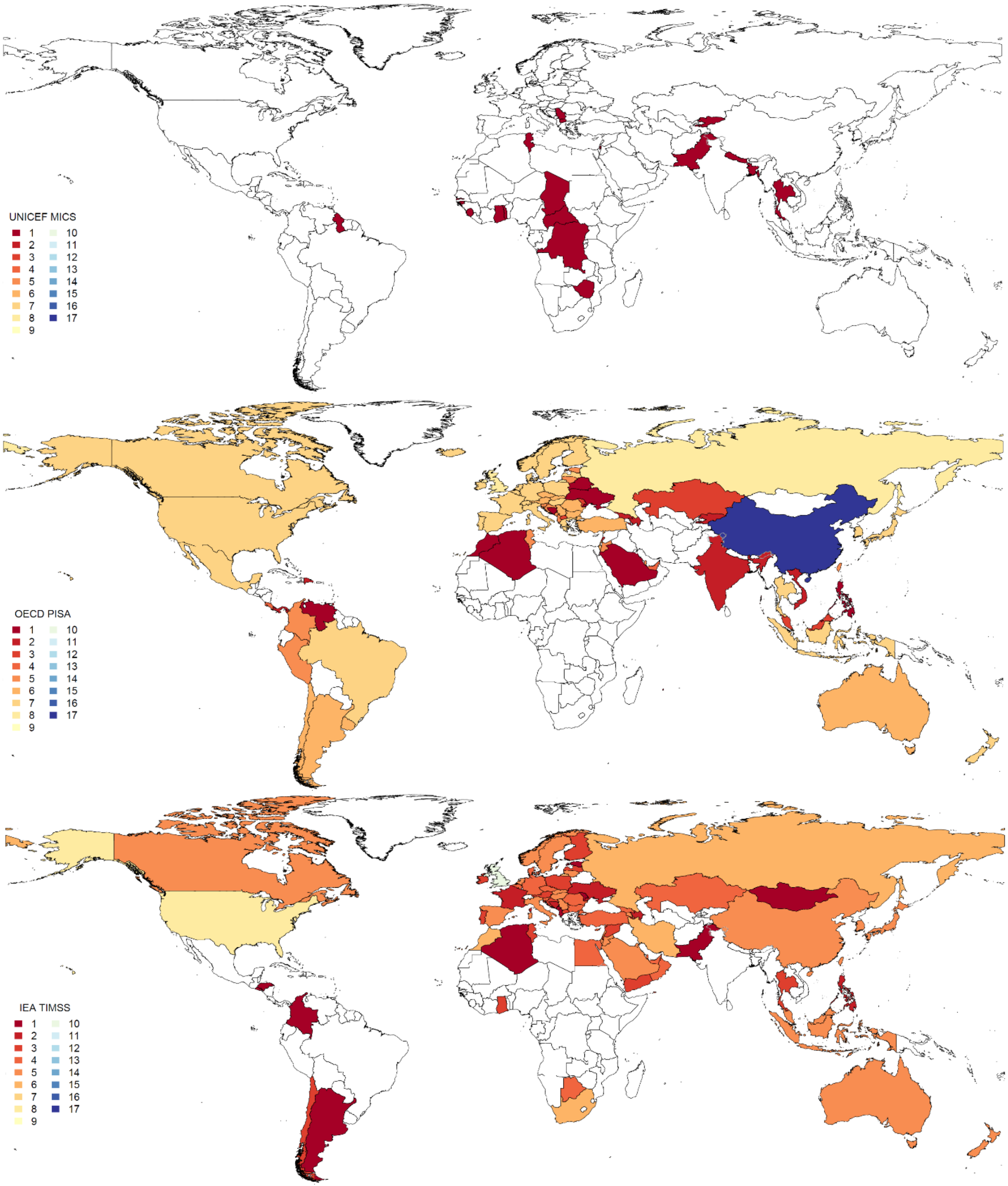
$$logit(P_{c,s,u,g,am}) = \beta0_t + \beta1_t female + \beta2_t enrollment + \beta3_t m1 + \beta4_t m2 + \beta5_t m3 + I_{csu} + I_{ta} + \epsilon_t$$

Where:

- $\beta0_t$ is a global intercept
- $\beta1_t$ is a test specific fixed intercept on female gender
- $\beta2_t$ is a test specific fixed intercept enrollment status
- $\beta3_t, \beta4_t, \beta5_t$ are test specific fixed intercepts on the three possible levels of mothers education (primary, secondary, tertiary)
- I_{csu} a location-specific random intercept
- I_{ta} is an age-specific random intercept
- ϵ_t is the global non-specific error

Resulting predictions were adjusted to ensure equivalency with initial percentage of population meeting foundational learning proficiency.

Supplementary Results



Appendix Figure 1. Coverage of assessments, number indicates observations per country. MICS (top), TIMSS (middle) and PISA (bottom).

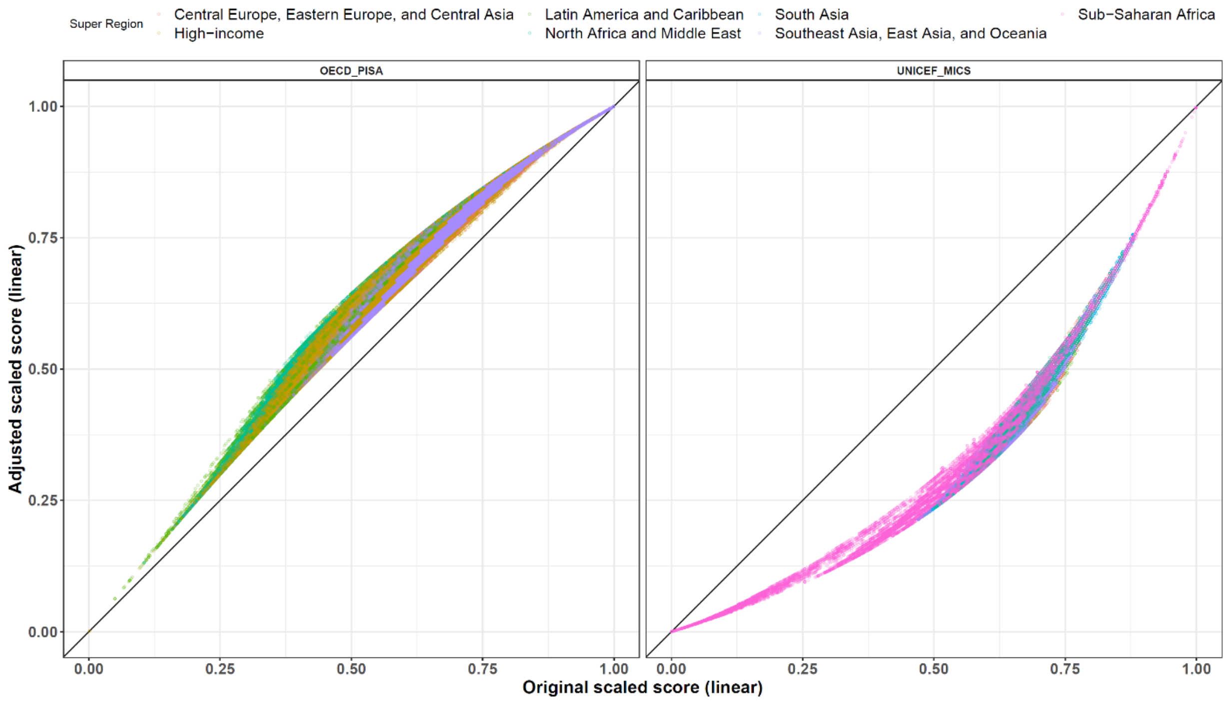
Appendix Table 1. Populations included in each assessment

Assessment	Years	Countries	Grades	Ages
OECD PISA	2002, 2010, 2012, 2015, 2018, 2006, 2009, 2000, 2003	ALB, ARE, ARG, AUS, AUT, AZE, BEL, BGR, BIH, BLR, BRA, BRN, CAN, CHE, CHL, CHN, COL, CRI, CZE, DEU, DNK, DOM, DZA, ESP, EST, FIN, FRA, GBR, GEO, GRC, HRV, HUN, IDN, IRL, ISL, ISR, ITA, JOR, JPN, KAZ, KGZ, KOR, LBN, LTU, LUX, LVA, MAR, MDA, MEX, MKD, MLT, MNE, MUS, MYS, NLD, NOR, NZL, PAN, PER, PHL, POL, PRT, QAT, ROU, RUS, SAU, SGP, SRB, SVK, SVN, SWE, THA, TTO, TUN, TUR, TWN, UKR, URY, USA, VEN, VNM	grade 10, grade 11, grade 8, grade 9, grade 7	16, 15, 14
IEA TIMSS	2018, 2007, 2010, 2014, 2003, 2015	ALB, ARE, ARG, ARM, AUS, AUT, AZE, BEL, BGR, BHR, BIH, BWA, CAN, CHL, COL, CYP, CZE, DEU, DNK, DZA, EGY, ESP, EST, FIN, FRA, GEO, GHA, HND, HRV, HUN, IDN, IRL, IRN, ISR, ITA, JOR, JPN, KAZ, KOR, KWT, LBN, LTU, LVA, MAR, MDA, MKD, MLT, MNE, MNG, MYS, NLD, NOR, NZL, OMN, PAK, PHL, POL, PRT, PSE, QAT, ROU, RUS, SAU, SGP, SLV, SRB, SVK, SVN, SWE, SYR, THA, TUN, TUR, TWN, UKR, USA, YEM, ZAF	grade 4, grade 5, grade 8, grade 9, grade 6, grade 7	10, 9, 12, 11, 13, 8, 14, 7, 6, 15, 16, 17
UNICEF MICS	2019, 2018, 2017	BGD, CAF, COD, GHA, GMB, GNB, GUY, KGZ, KIR, MKD, NPL, PSE, SRB, STP, TCD, TGO, THA, TON, TUN, ZWE	grade 10, grade 4, grade 5, grade 6, grade 7, grade 8, grade 9	14, 13, 7, 8, 9, 10, 11, 12

Appendix Table 2. Availability of variables across learning assessments and relative distribution of individual characteristics.

Variable	group	IEA TIMSS	OECD PISA	UNICEF MICS	Total
Number of books in home	0-10	19%	18%		18%
	11-100	52%	47%		49%
	>100	23%	32%		28%
	unknown	6%	3%	100%	5%
Father's education	ISCED level 0	2%	3%		2%
	ISCED level 1	4%	7%		6%
	ISCED level 2	8%	14%		11%
	ISCED level 3	16%	42%		30%
	ISCED level 4+	28%	27%		27%
	unknown	42%	7%	100%	24%
Mother's Education	ISCED level 0	2%	4%	21%	3%
	ISCED level 1	5%	7%	25%	6%
	ISCED level 2	8%	14%	28%	11%
	ISCED level 3	16%	43%	11%	30%
	ISCED level 4+	28%	28%	7%	28%
Grade	grade 4	48%		28%	22%
	grade 5	3%		25%	1%
	grade 6	1%		22%	0%
	grade 7	0%	1%	11%	1%
	grade 8	44%	5%	9%	23%

	grade 9	4%	33%	5%	20%
	grade 10		54%	1%	29%
	grade 11		8%		4%
Home Language	unknown	6%	4%	5%	5%
	same language as test	89%	83%	70%	85%
	different language	5%	14%	25%	10%
Attended Preschool	unknown	71%	24%	100%	46%
	no preschool	5%	10%		8%
	attended preschool	24%	66%		46%
School Type	unknown	100%	36%	100%	66%
	public school		53%		28%
	private school		11%		6%
Sex	female	50%	50%	50%	50%
	male	50%	50%	50%	50%
Super Region	Central Europe, Eastern Europe, and Central Asia	21%	20%	7%	20%
	North Africa and Middle East	31%	9%	10%	19%
	High-income	37%	51%		44%
	South Asia	0%		40%	0%
	Latin America and Caribbean	1%	13%	3%	8%
	Sub-Saharan Africa	5%		30%	2%
	Southeast Asia, East Asia, and Oceania	6%	7%	11%	7%
Urban	unknown	88%	47%	1%	65%
	rural	4%	23%	63%	15%
	urban	8%	30%	36%	20%
Wealth Quantile	unknown	4%	0%		2%
	First wealth quantile	34%	20%	81%	27%
	Second wealth quantile	24%	26%	16%	25%
	Third wealth quantile	11%	33%	3%	23%
	Fourth wealth quantile	28%	21%	0%	24%



Appendix figure 2. Scaled learning outcomes pre and post adjustment, for PISA and MICS. The x-axis shows the scaled, unadjusted scores for each survey, and the y-axis the scaled, adjusted scores for each survey, colored by super region. TIMMS data are not shown, as they are not adjusted.

References

1. Beck, K. C. *et al.* Educational inequalities in adult mortality: a systematic review and meta-analysis of the Asia Pacific region. *BMJ Open* **12**, e059042 (2022).
2. Masters, R. K., Link, B. G. & Phelan, J. C. Trends in education gradients of 'preventable' mortality: A test of fundamental cause theory. *Soc. Sci. Med.* **127**, 19–28 (2015).
3. Phelan, J. C., Link, B. G. & Tehranifar, P. Social Conditions as Fundamental Causes of Health Inequalities: Theory, Evidence, and Policy Implications. *J. Health Soc. Behav.* **51**, S28–S40 (2010).
4. Balaj, M. *et al.* Parental education and inequalities in child mortality: a global systematic review and meta-analysis. *The Lancet* **398**, 608–620 (2021).
5. Hanushek, E., Jamison, D. & Jamison, E. The Effects of Education Quality on Mortality Decline and Income Growth. *Econ. Educ. Rev.* **26**, 771–788 (2007).
6. Baker, D. P. *et al.* The Population Education Transition Curve: Education Gradients Across Population Exposure to New Health Risks. *Demography* **54**, 1873–1895 (2017).
7. Garcy, A. M. & Berliner, D. C. A critical review of the literature on the relationship between school quality and health inequalities. *Rev. Educ. Stud.* **6**, 40–66 (2018).
8. Angrist, N., Djankov, S., Goldberg, P. K. & Patrinos, H. A. Measuring human capital using global learning data. *Nature* **592**, 403–408 (2021).
9. United Nations. Transforming our World: The 2030 Agenda for Sustainable Development. Sustainable Development Knowledge Platform. https://sustainabledevelopment.un.org/post2015/transformingourworld/publication (2015).
10. UNESCO. *SDG Resources for Educators - Quality Education*. https://en.unesco.org/themes/education/sdgs/material/04 (2018).

11. Bundy, D. A. P., Silva, N. de, Horton, S., Jamison, D. T. & Patton, G. C. *Child and Adolescent Health and Development*. (The International Bank for Reconstruction and Development / The World Bank, 2017). doi:10.1596/978-1-4648-0423-6.

12. Altinok, N., Angrist, N. & Patrinos, H. A. *Global Data Set on Education Quality (1965–2015)*. (World Bank, Washington, DC, 2018). doi:10.1596/1813-9450-8314.

13. Hanushek, E. A. & Woessmann, L. Chapter 2 - The Economics of International Differences in Educational Achievement. in *Handbook of the Economics of Education* (eds. Hanushek, E. A., Machin, S. & Woessmann, L.) vol. 3 89–200 (Elsevier, 2011).

14. Wu, M. The role of plausible values in large-scale surveys. *Stud. Educ. Eval.* **31**, 114–128 (2005).

15. Sorensen, R. & Zheng, P. GitHub - ihmeuw-msca/crosswalk: Crosswalk with network analysis. (2021).

16. UNESCO Institute for Statistics. International Standard Classification of Education (ISCED). https://uis.unesco.org/en/topic/international-standard-classification-education-isced (2017).

17. Handbook on measuring equity in education; 2018. *Equity Educ.*

18. Toutkoushian, R. & Michael, R. An Alternative Approach to Measuring Horizontal and Vertical Equity in School Funding. **32**, (2007).

19. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Covariates 1980-2019. (2020) doi:https://doi.org/10.6069/CFCY-WA51.

20. Bowles, S. & Carlin, W. Inequality as experienced difference: A reformulation of the Gini coefficient. *Econ. Lett.* **186**, 108789 (2020).

21. Dearden, L., Ferri, J. & Meghir, C. The Effect of School Quality on Educational Attainment and Wages. *Rev. Econ. Stat.* **84**, 1–20 (2002).
22. Musso, M. F., Cascallar, E. C., Bostani, N. & Crawford, M. Identifying Reliable Predictors of Educational Outcomes Through Machine-Learning Predictive Modeling. *Front. Educ.* **5**, (2020).
23. Eriksson, K., Lindvall, J., Helenius, O. & Ryve, A. Socioeconomic Status as a Multidimensional Predictor of Student Achievement in 77 Societies. *Front. Educ.* **6**, (2021).
24. Christenson, S. L., Rounds, T. & Gorney, D. Family factors and student achievement: An avenue to increase students' success. *Sch. Psychol. Q.* **7**, 178–206 (1992).
25. Ma, X. & Klinger, D. A. Hierarchical Linear Modelling of Student and School Effects on Academic Achievement. *Can. J. Educ. Rev. Can. Léducation* **25**, 41–55 (2000).