Trends in International Assessments and Outcomes in Adulthood*

Samuel Stemper†

August 25, 2025

Abstract

International assessments such as PISA and TIMSS are widely used to compare the academic proficiency of adolescents across countries and over time. Are scores on these assessments associated with outcomes in adulthood? Combining data from mathematics scores in PISA, TIMSS, and PIAAC, and adulthood outcomes from 18 representative global surveys, I compare the relative associations of PISA and TIMSS scores with later outcomes among cohorts that took both tests during adolescence. Results suggest that cohorts with higher test scores perform better on assessments of adulthood skills, obtain higher levels of education, and have higher incomes as adults. I find suggestive evidence that PISA scores exhibit a relatively stronger relationship with education and income in adulthood compared to TIMSS scores.

*First version: November 29, 2023. This version: August 25, 2025. I am grateful to Thomas Kane and Joshua Goodman for helpful comments and suggestions. All remaining errors are mine.

†University of Auckland. Email: sam.stemper@auckland.ac.nz

Math skills play an important role in the academic and economic trajectory of individuals throughout their lives. Training to develop these skills is concentrated during childhood and adolescence. Policymakers frequently use international assessments like the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS) to monitor students' academic proficiency and their potential to succeed academically and economically.[1] This paper compares the strength of association between cohort-level performance on these two major international assessments and subsequent numeracy, educational attainment, and economic outcomes in adulthood.

This question is of particular significance for two reasons. First, PISA and TIMSS exams test distinct skills, even within the same subject. TIMSS emphasizes curriculum-based knowledge, focusing on material that students (ought to) learn in school. Alternatively, PISA measures students' ability to apply their knowledge in "real-world" scenarios, with less emphasis on curricular material.[2] Second, in many countries, PISA and TIMSS scores have moved in opposite directions since 2000. While TIMSS math scores have increased in most participating countries, PISA scores have stagnated or declined. This phenomenon is not explained by changes in the composition of participating countries or within-country time trends.

In this paper, I use variation in PISA and TIMSS test scores across cohorts and across countries to test the degree to which country-by-cohort average test scores are associated with average cohort-level outcomes in adulthood. I use data from the Programme for the International Assessment of Adult Competencies (PIAAC) to measure skills in adulthood and use harmonized international survey data to measure education and income. I focus on cohorts that took both assessments and evaluate the relative strength of correlations between these cohort-level measures and the average skills and outcomes of these cohorts later in life.

Across measures of numeracy skills, educational attainment, and income, I find that both PISA and TIMSS scores are positively associated with outcomes in adulthood. An increase in

[1]For example, following the release of the 2012 PISA results, U.S. Secretary of Education Arne Duncan characterized the U.S. performance as "a picture of educational stagnation," highlighting that "to land a job that pays a living wage, most people will need at least some college," underscoring the connection between skills, educational attainment, and access to quality employment opportunities (Duncan, 2013). Similarly, upon the release of the 2011 TIMSS results, Secretary Duncan remarked, "Given the vital role that science, technology, engineering, and math play in stimulating innovation and economic growth, it is particularly troubling that eighth-grade science achievement is stagnant and that students in Singapore and Korea are far more likely to perform at advanced levels in science than U.S. students" (Duncan, 2012).

[2]This is one of many differences between these assessments. I discuss these differences more broadly in Section 1.

average cohort-level PISA scores of 1 student-level standard deviation (i.e. a 100-point increase on the student-level PISA scale) is associated with a 0.2 standard deviation increase in adulthood numeracy test scores, a 1-year increase in years of education, and a 7 to 12 percentage-point increase in household income percentile. For TIMSS scores, effect sizes with respect to adulthood numeracy test scores are similar, but generally weaker with respect to education attainment and income. Among these measures of educational attainment and income, I find suggestive evidence for differences in these magnitudes, both in separate regressions and joint regressions that include both scores.

This work relates to a broad set of literature on the relationship between measures of human capital and education, income, and growth. Of particular relevance to my study are Doty et al. (2022) and Égert et al. (2024), who use cohort-level variation (across US states and countries, respectively) in test scores to estimate the relationship between test scores and outcomes in adulthood. Methodologically, I employ a comparable approach to Doty et al. (2022). However, I differ from both papers in my use of two measures of skills to evaluate the relationship between skills and outcomes across different testing regimes. (The differences between these assessments are discussed further in Section 1.)

More broadly, many studies examine the relationship between test scores and individual outcomes within countries or differences in economic growth across countries. Regarding the prior, Goldhaber and Özek (2019) and Hanushek (2012) summarize this literature, with the latter concluding that, in developed countries "[t]here is now considerable evidence that cognitive skills measured by test scores are directly related to individual earnings, productivity, and economic growth."[3] Regarding the latter, many studies study the role of country-level differences in test scores and education levels in driving differences in economic growth rates. This work typically uses variation in test scores across countries, rather than over time (Barro, 1991; Hanushek, 2013; Mankiw et al., 1992).[4]

Some of this work seeks to go beyond associations to establish causal links between test scores and outcomes. For instance, Hanushek et al. (2015) use variation in compulsory schooling laws to

[3]My work relates specifically to math skills, which have been studied in more depth in relation to high school curriculum (Goodman, 2019; Joensen and Nielsen, 2009) and college majors (Kirkeboen et al., 2016). Altonji et al. (2012) summarize this literature in their review.

[4]Notable exceptions include Coulombe et al. (2004) and Coulombe and Tremblay (2006), who estimate growth regressions using variation across cohorts and across countries.

instrument for PIAAC numeracy scores in individual-level regressions of income on scores. At the country level, Hanushek and Woessmann (2012) instrument for an index of cognitive skills (derived from multiple test scores) using features of national education systems (e.g., relative teacher salaries, private enrollment shares, etc.). My setting is not well-suited for such approaches, as identifying causal effects would require separate instruments for both PISA and TIMSS scores. Consequently, my estimates should be interpreted as measures of the association between these scores and adulthood outcomes at the cohort level. I discuss differences between cohort-level and individual-level associations in more detail in Section 2.

I contribute to this literature by evaluating the relative strength of association between different measures of cohort skills and adulthood outcomes, using two prominent testing regimes: PISA and TIMSS. Using across-cohort variation in test scores among cohorts that took both tests, I estimate how strongly average cohort-level scores are associated with average cohort-level outcomes.

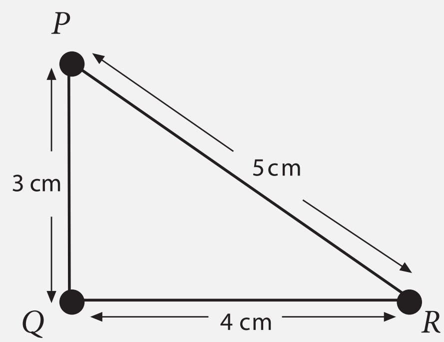
1 Background on PISA and TIMSS

PISA and TIMSS are both international assessments designed to evaluate and compare the educational performance of adolescent students across countries. The Organisation for Economic Co-operation and Development (OECD) conducts the PISA assessment every three years and tests 15-year-old students in mathematics, reading, and science. TIMSS, organized by the International Association for the Evaluation of Educational Achievement (IEA), assesses mathematics and science proficiency among 4th and 8th graders every four years. For both surveys, countries choose whether to participate in each round of assessment, and sampling and test administration are typically coordinated separately within each country.

1.1 Assessment Methodologies

PISA and TIMSS differ substantially in the material used to assess student skills. TIMSS is curriculum-based and tests students' knowledge of the material taught in school. Alternatively, PISA assesses students' ability to apply knowledge to "real-world" problems (Loveless, 2013). Sample questions from the 2011 Grade 8 TIMSS exam and the 2012 PISA exam illustrate this difference. As an example, in 2011, the TIMSS Grade 8 exam included the question below.

Example TIMSS Question



Which of these is the reason that triangle PQR is a right angle triangle?

A. $3^2 + 4^2 = 5^2$

B. $5 < 3 + 4$

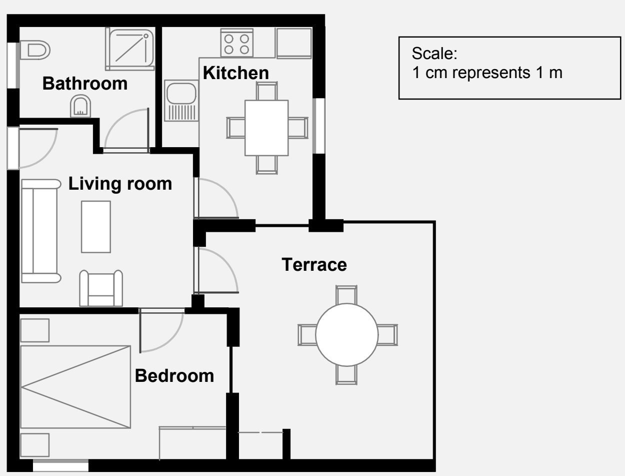
C. $3 + 4 = 12 - 5$

D. $3 > 5 - 4$

In contrast, the PISA 2012 exam included the question below.

Example PISA Question

This is the plan of the apartment that George's parents want to purchase from a real estate agency.



To estimate the total floor area of the apartment (including the terrace and the walls), you can measure the size of each room, calculate the area of each one and add all the areas together. However, there is a more efficient method to estimate the total floor area where you only need to measure 4 lengths. Mark on the plan above the four lengths that are needed to estimate the total floor area of the apartment.

Both questions above involve reasoning with geometric shapes. While the TIMSS question is a standalone mathematical problem involving knowledge of the Pythagorean theorem, the PISA question requires students to apply geometric reasoning to a real-world scenario. Additional example questionnaires from PISA and TIMSS exams are publicly available online.[5]

1.2 Differences in Tested Population

Beyond the content of the questions used in assessments, PISA and TIMSS also differ in the test-taking population. Most importantly, PISA has an age-based sample, testing 15-year-olds (regardless of grade), whereas TIMSS has a grade-based sample, testing students in grades 4 and 8 (Loveless, 2013). As a result, students tested during PISA assessments are typically one year older than students tested during TIMSS Grade 8 assessments. More substantively, TIMSS data contains students who are either ahead of or behind the typical grade progression in their country. In Section 2, I describe my process to make these test-taking populations more comparable by restricting my focus to students who follow their country's typical grade progression.

1.3 Trends in PISA and TIMSS

PISA and TIMSS differ not only in the content and student population used in their assessments but also in their conclusions about the trajectory of global math skills. While TIMSS math scores have steadily increased over the past 25 years, PISA math scores have been flat or decreasing.

Figure 1 illustrates this phenomenon among the most well-represented countries in PISA and TIMSS assessments: the six countries that have participated in PISA and TIMSS Grade 8 assessments every year since their inception. These countries are Australia, Hong Kong, Hungary, Japan, Korea, and the United States. For each country, Figure 1 displays the country's average scores in the 7 PISA math assessments and the 7 TIMSS Grade 8 math assessments between 1995 and 2019.[6] Lines show linear time trends, separately for PISA and TIMSS; the slope coefficient, representing

[5] NCES, TIMSS Released Assessment Questions; OECD, PISA Test Questions.

[6] These scores incorporate the sample restrictions and rescaling procedure described in Section 2 below.

the average annual change in scores, is shown in the corner of each panel.

Figure 1 demonstrates that, across many countries, PISA scores have fallen relative to TIMSS scores. These are large discrepancies: PISA math scores in Australia fell by over 0.4 student-level standard deviations between 2000 and 2018. Over the same period, TIMSS Grade 8 scores were flat. For reference, Bloom et al. (2008) and Evans and Yuan (2019) find that a year of schooling typically increases test score performance by 0.3 and 0.2 standard deviations, respectively.[7] All 6 of the countries in Figure 1 exhibit relatively better trends on TIMSS tests than PISA tests; 4 of these differences are statistically significant.[8]

This phenomenon has not unfolded uniformly across all nations. Figure 2 illustrates the country-level variation in long-term test score trends. Specifically, the horizontal axis displays estimated time trends in PISA math scores and the vertical axis displays the equivalent estimates for TIMSS Grade 8 math scores; these time trends are estimated identically to those in Figure 1. In this analysis, I limit the sample to countries with at least one PISA test and one TIMSS Grade 8 test in 2006 or earlier and at least one PISA test and one TIMSS Grade 8 test in 2012 or later. Consistent with the evidence presented above, most countries exhibit relatively larger growth in TIMSS scores relative to PISA scores. This can be seen visually in Figure 2; most points fall above the 45-degree line. However, some countries deviate from this pattern, with greater growth in PISA scores relative to TIMSS scores; Italy is a notable example. Furthermore, there are countries that exemplify an extreme manifestation of this trend; for instance, in Korea, time trends in PISA scores suggest an annual decrease in PISA scores of more than 0.01 standard deviations, while TIMSS scores have increased by over 0.01 standard deviations each year.

The data and methods described in the section that follows use this variation in test scores—over time, across countries, and between assessments—to test whether these assessments are associated with outcomes in adulthood.

[7] Bloom et al. (2008) use nationally-representative data from the U.S. The 0.3 estimate refers to the effect of a year of schooling for 8th to 9th graders, who are typically 13 to 15 years old. Evans and Yuan (2019) use a sample of test scores from low- and middle-income countries.

[8] In the Supplemental Appendix, I formally test for differences in the trajectories of PISA and TIMSS assessments using student-level data from all PISA and TIMSS participating countries and show that these patterns are not driven by changes in the composition of test-taking countries; regressions that include country fixed effects and country-specific time trends yield nearly identical results.

2 Data and Methods

2.1 PISA and TIMSS Test Scores

I prepare country-by-cohort test score averages for PISA math and Grade 8 TIMSS math using publicly available student-level test score data.

As discussed above, PISA has an age-based sample whereas TIMSS has a grade-based sample. To make these test-taking populations more comparable, I restrict my PISA sample to students who are in the modal grade among test-takers in their country. To generate the equivalent sample in TIMSS data, I restrict my focus to students who are within the 1 year range centered at the modal age in TIMSS data. Altogether, these restrictions mean that both samples include students who are in the typical grade for their age.

Both PISA and TIMSS scores were scaled at their introduction to have mean 500 and standard deviation of 100. However, because the two assessments are scaled relative to different reference populations, a given score does not reflect the same relative performance across tests. To place PISA and TIMSS scores on the same scale, I apply the procedure used in Gust et al. (2024). This rescaling helps ensure that observed differences in test scores across assessments reflect meaningful differences in performance rather than artifacts of differing scale definitions. To do so, I first restrict my focus to countries that participated in both the TIMSS 2019 and PISA 2018 tests.[9] Among these countries, I then standardize the TIMSS Grade 8 math scores to have a mean of zero and a standard deviation of one and estimate conversion parameters to rescale these scores to match the mean and standard deviation of the PISA math scale. With these conversion parameters, I apply this rescaling to all data for Grade 8 TIMSS scores. More specifically, for a given TIMSS Grade 8 math score t_i, the corresponding PISA-scale score p_i is calculated as

$$p_i = \frac{(t_i - m^C_{TIMSS})}{s^C_{TIMSS}} s^C_{PISA} + m^C_{PISA},$$

where m^C and s^C represent the mean and standard deviation of scores in the common (linking) countries. I refer to these rescaled TIMSS scores as “PISA Scale" scores throughout this paper.

After this transformation, I compute the average birth year and average test score in math

[9] These years are the same as those used in Gust et al. (2024).

for each country and testing round (e.g. 2000 PISA, 2015 Grade 8 TIMSS, etc.). In some years, student-level data does not report students' year of birth; in these cases, I compute birth years based on each student's age and the test date.[10] Throughout the analysis, PISA and TIMSS scores are reported after subtracting 500 and dividing by 100, so that a one-unit change corresponds to one student-level standard deviation.

The set of countries identified by PISA and TIMSS administrators are generally comparable, with a few exceptions such as Belgium, which is administered separately in Flemish and French regions, and the UK, which is administered separately in England, Northern Ireland, and Scotland. Where possible, I separately identify these regions in PISA and other data.

Tests are typically administered every three years for PISA and every four years for TIMSS. Due to this periodicity, there is rarely direct overlap in individual birth years across the two tests. To generate larger overlapping test scores for both tests, I impute observations by taking the closest non-missing test score that is no more than two birth years away. In the case of ties, I take the later test score.[11]

For example, the United States participated in the Grade 8 TIMSS math exam in 1995 and 1999 (among other years). Students taking these exams were, on average, born in 1981 and 1985, respectively. Following the procedure described above, I assign the 1981 birth cohort's TIMSS Grade 8 score to students born between 1979 and 1982 and assign the 1985 birth cohort's score to students born between 1983 and 1986.[12]

Finally, I keep only country-by-cohort observations with both PISA and TIMSS scores. This restriction limits the set of cohorts that appear in my data but ensures that statistical tests using PISA scores have the same sample as tests using TIMSS scores, and vice versa.

I link this country-by-cohort data with average cohort-level outcomes—namely, adult PIAAC test scores in numeracy and education and income data from harmonized international surveys. Because this data is not available for all birth years and all countries, this generates two slightly different sets of birth cohort-country combinations. I refer to these different samples as the "PI-

[10]I assume PISA tests are administered in May and TIMSS are administered in July.

[11]In the Supplemental Appendix, I show that I obtain similar estimates using differing levels of cohort-level imputation.

[12]Students born in 1987 are assigned the score of the next cohort: students who took the Grade 8 TIMSS math exam in 2003.

AAC Sample" and the “SDR Sample."[13]

2.2 PIAAC Scores

To assess the numeracy skills of adults, I use test results from PIAAC. Similar to PISA and TIMSS, PIAAC is an international survey of skills. Unlike PISA and TIMSS, PIAAC assesses the skills and competencies of adults aged 16 to 65. PIAAC rounds took place in 2012, 2014, and 2017. Each country only participated in one round of PIAAC testing, with one exception: the United States administered PIAAC tests in all three years.[14]

I focus on PIAAC numeracy scores. The OECD defines numeracy as “the ability to access, use, interpret, and communicate mathematical information and ideas, in order to engage in and manage the mathematical demands of a range of situations in adult life" (OECD et al., 2009). Example PIAAC questions are publicly available online from the OECD.[15] The OECD reports PIAAC scores on a scale ranging from 0 to 500. This scale is based on “information-processing tasks of increasing complexity," and is calculated such that “[a]t each point on the scale, an individual with a proficiency score of that particular value has a 67% chance of successfully completing test items located at that point" (OECD, 2013). To simplify interpretation, I scale numeracy test scores such that they have mean 0 and standard deviation 1.

I link PIAAC participants to their cohort's PISA and TIMSS scores based on country and birth year, which I calculate based on the year of the test and the respondent's age. For a small number of countries, respondent ages are reported in 5-year ranges (e.g. 20-24); in these cases, I take the midpoint and round up to the nearest integer.

To make PIAAC data consistent with the cohort-level PISA and TIMSS data, I collapse the PIAAC data to the cohort-by-country-by-survey year level, taking the average outcome (e.g., numeracy score) for each country-by-cohort-by-survey year cell. I link these average country-by-cohort-by-survey outcomes to their cohort's PISA and TIMSS test scores.

[13] In the Supplemental Appendix, I show the number of observations in each country-cohort combination across both samples.

[14] The United States' 2014 PIAAC round is known as the National Supplement to the Main Study. This round was not administered to a nationally representative sample and was instead meant to “enhance and expand" the Round 1 data. See NCES, PIAAC Participating Countries.

[15] OECD, PIAAC Numeracy – sample items.

2.3 Harmonized International Survey Data

For education and income, I use harmonized international survey data from the Survey Data Recycling ("SDR") project.[16] The SDR project collects publicly available data from numerous large-scale international survey projects, such as the International Social Survey Programme and the European Social Survey, and harmonizes responses to common demographic questions, among others.[17] The Survey Data Recycling project selects surveys based on the following criteria: surveys must be "designed as cross-national [...]; samples intended to be representative of the adult population of given country or territory; projects contain questions about political attitudes and behaviors; projects are freely available in the public domain; and their documentation [...] is provided in English."[18]

I consider three main educational outcomes: years of education as well as whether the respondent completed secondary or tertiary school. Many surveys do not ask for years of education directly but ask for the age a respondent finished education, the year in which a respondent finished education, or the highest level of education they received. The SDR project harmonizes across all of these response types to generate a value that is comparable across survey rounds.

Additionally, I consider household income percentiles. In reporting their household income, many surveys ask respondents to select from a set of pre-coded income brackets. To account for different reporting schemes across surveys and countries, SDR data estimates each respondent's position in the national income distribution by converting these brackets into percentiles within each national survey. More technically, income brackets are sorted in ascending order (from the lowest to the highest household income) and are assigned values of the mid-point from the cumulative distribution.

One concern with using household income, rather than personal income or wages, is that household income may be affected by household composition. While I do not have data on particular aspects of household composition (e.g. whether respondents are living with their parents

[16] Specifically, I use the Survey Data Recycling (SDR) v.2.0 database. This data and supporting documentation is available online: SDR2 Database.

[17] Slomczynski and Tomescu-Dubrow (2018) contains more information on this harmonization process.

[18] Individually, these surveys have been used extensively in economics and political science research. For example, Aghion et al. (2010) and Mayda (2006) use data from the International Social Survey Programme, Hainmueller and Hiscox (2007) and Roth and Wohlfart (2018) use data from the European Social Survey, and Besley and Persson (2019) and Guriev and Zhuravskaya (2009) use data from the World Values Survey.

and/or their spouse), I later show that my main results are robust to flexible controls for household size.

I restrict the sample to respondents who have non-missing PISA and TIMSS cohort scores and are age 16 or older at the time of the interview.[19]

As with PIAAC data, I collapse the SDR data to the cohort-by-country-by-survey year level, taking the average outcome (e.g., years of education, income percentile, etc.) for each country-by-cohort-by-survey year cell. I link these average country-by-cohort-by-survey outcomes to their cohort's PISA and TIMSS test scores.[20]

2.4 Data Description

Table 1 summarizes the main variables for the country-by-cohort-by-survey year cells in my two samples. On average, these cells are approximately 50 percent female, with a mean age of roughly 23 years.[21] In the SDR data, the average educational attainment within cells is 13 years. Among respondents aged 19 or older, 83 percent had completed secondary school, and among those aged 24 or older, 36 percent had completed a bachelor's degree at the time of the survey.

As noted above, these samples contain slightly different sets of country-by-cohort pairs. There are 28 unique countries in the PIAAC sample, 15 containing at least 3 unique PISA scores and 3 unique Grade 8 TIMSS scores. There are 51 unique countries in the SDR sample, 27 containing at least 3 unique PISA scores and 3 unique Grade 8 TIMSS scores. Table 1 additionally lists the set of 18 survey sources included in my sample.

2.5 Methodology

My method estimates the linear relationship between a cohort's average test scores in adolescence and their average outcomes in adulthood. My baseline regression takes the form below.

$$y_{cbt} = \beta_0 + \beta_1 TestScore_{cb} + \gamma_c + \delta_{age} + \zeta_t + \varepsilon_{icbt} \tag{1}$$

[19] A small number of respondents report ages that don't align with their reported year of birth and survey year. I drop all observations for which the difference between a respondent's predicted age (survey year minus birth year) and a respondent's report age is larger than 2.

[20] The SDR data does not include information on country of birth, so immigrants—who may not have been in the country during PISA or TIMSS testing—cannot be excluded from the linking process.

[21] I also provide age statistics for individuals aged 19 and above and 24 and above, which correspond to the samples used in analyses of secondary school completion, bachelor's completion, and household income percentiles.

y_{cbt} denotes an average outcome (e.g. years of education, household income percentile) for the cohort in country c born in year b and surveyed in year t. For all outcomes, I estimate separate regressions using either PISA or TIMSS scores individually, as well as a joint regression that includes both scores.

In Equation 1, γ_c captures factors that vary across countries but are constant over time (e.g., time-invariant country characteristics such as geography or persistent cultural factors), δ_{age} captures factors that vary with the age of the survey respondent (e.g., life-cycle effects, such as increasing earnings potential with age), and ζ_t captures factors that vary over time but are constant across countries (e.g. global economic shocks, technological changes).

Additionally, when possible, I test whether my results are sensitive to including region-by-age fixed effects—which allow age fixed effects to vary across regions of the world, as specified by the World Bank—and country-by-survey year fixed effects—which control for time-specific country factors (e.g. macroeconomic fluctuations).

My coefficient of interest, β_1, measures the association between cohort-level average test scores and average cohort-by-country-by-survey year level outcomes in PIAAC or SDR data. In regressions that include both test scores, I report the p-value from a Wald test of coefficient equality. In parallel regressions that include only one score, I compare the coefficient estimates using the stacked regression approach described in Oberfichtner and Tauchmann (2021).

For regressions using PIAAC data, I normalize individual sampling weights as $w_{ict} / \sum_{i \in ct} w_{ict}$, where w_{ict} is the final sampling weight for individual i in country c in year t, and $\sum_{i \in ct} w_{ict}$ is the total sampling weight for individuals in country c in year t. I then calculate the cell weight as the sum of these normalized individual weights. For regressions using SDR data, I weight estimates in each cell using the sum of SDR-provided weights, which rescale sampling weights from each national survey such that the sum within each survey equals the number of respondents. Following Doty et al. (2022), I cluster standard errors by country and birth year.

Prior to presenting results, I note two important caveats to this approach.

Firstly, β_1 may not necessarily reflect the *causal* effect of skills on these outcomes. There are many reasons for variation in national test scores: namely, changes in the composition of students, changes in the quality of the education they received, and changes in their environment, among others. These factors may have independent effects on outcomes in adulthood that are entirely unrelated to test scores. As such, the estimates in this paper should be viewed as assessing the association between test scores and later-life outcomes, rather than the causal effect of skills.

Second, the measures of cohort-level associations in this paper are distinct from associations using individual-level test scores and individual-level outcomes (Murnane et al., 2000; Lazear, 2003; Mulligan, 1999), and there are good reasons to believe they may be quantitatively different.

For instance, there may be within-cohort capacity constraints within a country (e.g. finite opportunities in universities or at high-paying firms). In such a scenario, increases in average skill levels may have muted effects on educational outcomes or income, because the supply of high-skill opportunities is fixed or growing slowly relative to the number of high-skill individuals in the cohort. The presence of such allocation issues would imply a weaker relationship between cohort-level scores relative to individual-level scores.

Conversely, cohort-level changes in test scores may influence the individual-level relationship between skills and outcomes. For example, positive social spillovers could arise, where broad increases in a cohort's average skill level enhance the productivity of all individuals, or specifically, the productivity of higher-skilled individuals.[22] In the simplest case of homogeneous peer effects or homogeneous social spillovers (e.g., Sacerdote (2011) Formula 4.1 or Acemoglu and Angrist (2000) Formula 6), the impact of peer or aggregate skills is entirely distinct from the effect of individual skills, and this effect may be either larger or smaller.

3 Results

As noted above, throughout the analysis, PISA and TIMSS scores are reported after subtracting 500 and dividing by 100, so that a one-unit change corresponds to one student-level standard deviation (100 points on the PISA scale). Regression coefficients can therefore be interpreted as the effect of a one-standard-deviation increase in the underlying student-level test score distribution on the outcome variable.

Table 2 shows the results of regressions using PIAAC data, which assess the relationship between adolescent test scores and numeracy skills in adulthood. Regression results shown in

[22] This phenomenon is commonly seen in models of peer effects. Sacerdote (2011) discusses various models of peer effects, distinguishing between those with homogeneous peer effects (where increases in a cohort's average skill level affect all individuals equally) and those with heterogeneous peer effects (where the effect varies based on individual skill levels). Similarly, in models of human capital externalities, wages depend not only on individual human capital but also on aggregate human capital (Acemoglu and Angrist, 2000).

Columns 1 to 3 include age and age-squared terms and fixed effects for country, gender, and test year.[23] Column 1 displays the relationship between a cohort's PISA scores and that cohort's average PIAAC scores in adulthood, which indicates that a 1 standard deviation increase in a cohort's average PISA score is associated with a 0.2 standard deviation increase in that cohort's average PIAAC scores in adulthood ($p < 0.05$). In Column 2, I repeat this specification, replacing PISA scores with TIMSS scores. Here, I find that a 1 standard deviation increase in a cohort's TIMSS average math score is associated with a 0.22 increase in that cohort's average PIAAC scores ($p < 0.05$). Below the estimate in Column 2, I display the results of a test of equality of coefficients for the estimates in Columns 1 and 2, which indicates that the difference between the PISA estimate and TIMSS estimate is not significant ($p = 0.69$). In Column 3, I estimate the effects of PISA and TIMSS scores in the same regression and find that estimates are reasonably similar to those estimated independently: associations between TIMSS and PIAAC are larger than associations between PISA and PIAAC, though differences between these effect sizes are statistically insignificant.

Columns 4 to 6 of Table 2 employ an additional specification by incorporating region-by-age interaction terms. The inclusion of these controls does not substantially alter the estimates; coefficients on TIMSS scores are nearly identical to corresponding coefficients on PISA scores. These estimates demonstrate that cohort-level variation in adolescent math skills is associated with later-life differences in numeracy skills.[24] To assess the persistence of these correlations with respect to education and income in adulthood, I next turn to harmonized international survey data.

Table 3 shows these estimates. Regression results shown in Columns 1 to 3 include fixed effects for country, age, gender, survey year, and survey wave, and results shown in Columns 4 to 6 add region-by-age fixed effects. The presence of multiple surveys within the same country over time allows for the inclusion of country-by-survey year fixed effects, which I add in Columns 7 to 9.

Panels A, B, and C of Table 3 display effects on measures of educational attainment. In Panel

[23] I use continuous terms for age, rather than fixed effects, due to the coarse rounding of ages for some countries in PIAAC data, described above.

[24] In similar specifications using TIMSS data and combined student test score data from PISA and the World Bank, Égert et al. (2024) find similar-sized estimates. Specifically, Égert et al. (2024) estimate log-log regressions of PIAAC scores on PISA scores extended backward with two vintages of World Bank data. Their estimated coefficients are between 0.2 and 0.7.

A, I estimate associations between adolescent test scores and years of education. Across numerous specifications, these results suggest that a 1 standard deviation increase in average cohort-level PISA scores is associated with a 0.5 to 1-year increase in years of education. TIMSS scores do not exhibit any systematic relationship with years of education; these estimates are consistently small and not statistically significant. I test for differences between these two estimates—PISA and TIMSS effects—and display corresponding p-values below coefficient estimates in Panel A; the estimates in my most stringent specification (in columns 7 to 9) suggest that PISA scores exhibit a statistically larger effect on years of education than TIMSS scores.

Panels B and C of Table 3 show effects on discrete levels of education: completion of secondary school and tertiary school, respectively. Panel B focuses on average cohort rates of secondary school completion. Effect sizes suggest that a 1 standard deviation increase in PISA scores increases average secondary school completion rates by roughly 10 to 12 percentage points. Increases in TIMSS scores are not systematically associated with an effect on secondary school completion rates. The results in Panel C do not suggest a clear relationship between test scores and tertiary school completion, as the effects are imprecise and not statistically significant for both PISA and TIMSS scores.

Finally, in Panel D of Table 3 I report the effects on household income percentiles. Among respondents aged 24 or older, a 1 standard deviation increase in cohort PISA scores is associated with a 7 to 12 percentile increase in average cohort household income. TIMSS scores exhibit slightly smaller but significant and positive effects, suggesting that a 1 standard deviation increase in TIMSS scores is associated with a 3 to 4 percentile increase in average cohort household income. The difference between the estimated effects of PISA and TIMSS scores varies in statistical significance across specifications. In my most stringent specification (in Columns 7 to 9), the effect of PISA scores is statistically significantly different from the effect of TIMSS scores.

I subject these results to several robustness tests, which appear in the Supplemental Appendix.

First, as noted above, periodicity in the timing of PISA and TIMSS tests means that there is rarely direct overlap in individual birth years across the two tests. In my main estimates, I impute observations by taking the closest non-missing test score that is no more than two birth years away. This choice balances a tradeoff between sample size (greater imputation allows for larger overlapping samples) and cohort-level precision (cohorts further away from the tested cohort are more dissimilar).

In the Supplementary Appendix, I show how my estimates change as I vary the number of adjacent cohorts imputed in my sample. For analyses using PIAAC as well as SDR data, as the number of imputed cohorts increases, the standard errors of my coefficients generally decrease, with the most substantial reduction occurring when moving from 1 to 2 years of adjacency, and my coefficient estimates remain stable across estimates using various degrees of imputation. Altogether, these results suggest that it is likely the long-run changes in test scores (e.g. decade-to-decade variation) that drive my results, rather than year-to-year variation in skill measures.

In addition, I show that my results concerning income are qualitatively similar when I control for household size.

Next, I confirm that my estimates concerning education and income are not driven by any one country or any one survey. To do so, I reproduce my estimates 10 times, once after excluding each of the largest 10 countries individually, and the results remain stable regardless of which country is excluded. I repeat this process by dropping individual surveys, and the results are similarly robust.

Finally, I note that the PIAAC data includes many adulthood outcomes related to education and income. However, since most of the PIAAC testing was conducted in 2012 and the dataset includes fewer countries, the sample sizes—particularly for older respondents—are much smaller than those using SDR data. Despite these limitations, I test for effects on education and wages and find qualitatively similar patterns, although with less precision.

4 Conclusion

In this paper, I estimate the degree to which outcomes in adulthood are associated with cohort-level variation in test scores. Comparing results from two major international testing regimes—PISA and TIMSS—I find that math scores from both tests exhibit strong correlations with adult skills, education levels, and incomes. Additionally, the evidence suggests that PISA scores have a stronger connection with education and income levels in adulthood than TIMSS scores. These results have several implications for policymakers and researchers alike.

Most substantially, these results highlight a concern for numerous countries that have witnessed stagnant or falling PISA scores alongside comparatively stronger growth in TIMSS. To the degree that PISA scores are more strongly associated with future educational and economic success, this highlights a concern for young cohorts who have generally performed worse than cohorts before them.

Concerning research, this work stands in contrast to efforts to harmonize results from numerous international assessments into a standard measure of education quality (e.g. Angrist et al. (2021)). My results here suggest that future research should be cautious about a single definition of human capital which is measured uniformly across diverse testing regimes. Instead, researchers should embrace a richer model of skills to study both the potential drivers of skills as well as the impact of skills on economic and non-economic outcomes. Recent work by Hermo et al. (2022), which distinguishes between "reasoning" skill and "knowledge" skill, is one such example.

References

Acemoglu, Daron and Joshua Angrist, "How large are human-capital externalities? Evidence from compulsory schooling laws," *NBER macroeconomics annual*, 2000, *15*, 9–59.

Aghion, Philippe, Yann Algan, Pierre Cahuc, and Andrei Shleifer, "Regulation and distrust," *The Quarterly journal of economics*, 2010, *125* (3), 1015–1049.

Altonji, Joseph G, Erica Blom, and Costas Meghir, "Heterogeneity in human capital investments: High school curriculum, college major, and careers," *Annu. Rev. Econ.*, 2012, *4* (1), 185–223.

Angrist, Noam, Simeon Djankov, Pinelopi K Goldberg, and Harry A Patrinos, "Measuring human capital using global learning data," *Nature*, 2021, *592* (7854), 403–408.

Barro, Robert J, "Economic growth in a cross section of countries," *The quarterly journal of economics*, 1991, *106* (2), 407–443.

Besley, Timothy and Torsten Persson, "Democratic values and institutions," *American Economic Review: Insights*, 2019, *1* (1), 59–76.

Bloom, Howard S, Carolyn J Hill, Alison Rebeck Black, and Mark W Lipsey, "Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions," *Journal of Research on Educational Effectiveness*, 2008, *1* (4), 289–328.

Coulombe, Serge and Jean-François Tremblay, "Literacy and Growth," *B.E. Journals in Macroeconomics*, 8 2006, *6* (2), 4.

_, Jean-Francois Tremblay, and Sylvie Marchand, "Literacy Scores, Human Capital and Growth Across Fourteen OECD Countries," *Statistics Canada Catalogue*, 2004, (89-552-mpe, no. 11).

Doty, Elena, Thomas J Kane, Tyler Patterson, and Douglas O Staiger, "What Do Changes in State Test Scores Imply for Later Life Outcomes?," Technical Report, National Bureau of Economic Research 2022.

Duncan, Arne, "Statement by U.S. Secretary of Education Arne Duncan on the Release of the 2011 TIMSS and PIRLS Assessments," https://web.archive.org/web/20121214014019/http://www.ed.gov/news/press-releases/statement-us-secretary-education-arne-duncan-release-2011-timss-and-pirls-assess December 2012. Accessed: 2025-01-24.

_, "The Threat of Educational Stagnation and Complacency: Remarks of U.S. Secretary of Education Arne Duncan at the release of the 2012 Program for International Student Assessment (PISA)," https://web.archive.org/web/20140207062452/http://www.ed.gov/news/speeches/threat-educational-stagnation-and-complacency December 2013. Accessed: 2025-01-24.

Égert, Balázs, Christine De la Maisonneuve, and David Turner, "A new macroeconomic measure of human capital exploiting PISA and PIAAC: Linking education policies to productivity," *Education Economics*, 2024, pp. 1–17.

Evans, David and Fei Yuan, "Equivalent years of schooling: A metric to communicate learning gains in concrete terms," *World Bank Policy Research Working Paper*, 2019, (8752).

Goldhaber, Dan and Umut Özek, "How much should we rely on student test achievement as a measure of success?," *Educational Researcher*, 2019, *48* (7), 479–483.

Goodman, Joshua, "The labor of division: Returns to compulsory high school math coursework," *Journal of Labor Economics*, 2019, *37* (4), 1141–1182.

Guriev, Sergei and Ekaterina Zhuravskaya, "(Un) happiness in transition," *Journal of economic perspectives*, 2009, *23* (2), 143–168.

Gust, Sarah, Eric A Hanushek, and Ludger Woessmann, "Global universal basic skills: Current deficits and implications for world development," *Journal of Development Economics*, 2024, *166*, 103205.

Hainmueller, Jens and Michael J Hiscox, "Educated preferences: Explaining attitudes toward immigration in Europe," *International organization*, 2007, *61* (2), 399–442.

Hanushek, Eric A, "The economic value of education and cognitive skills," *Handbook of education policy research*, 2012, pp. 39–56.

— , "Economic growth in developing countries: The role of human capital," *Economics of education review*, 2013, *37*, 204–212.

— **and Ludger Woessmann**, "Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation," *Journal of economic growth*, 2012, *17*, 267–321.

— **, Guido Schwerdt, Simon Wiederhold, and Ludger Woessmann**, "Returns to skills around the world: Evidence from PIAAC," *European Economic Review*, 2015, *73*, 103–130.

Hermo, Santiago, Miika Päällysaho, David Seim, and Jesse M Shapiro, "Labor market returns and the evolution of cognitive skills: Theory and evidence," *The Quarterly Journal of Economics*, 2022, *137* (4), 2309–2361.

Joensen, Juanna Schrøter and Helena Skyt Nielsen, "Is there a causal effect of high school math on labor market outcomes?," *Journal of Human Resources*, 2009, *44* (1), 171–198.

Kirkeboen, Lars J, Edwin Leuven, and Magne Mogstad, "Field of study, earnings, and self-selection," *The Quarterly Journal of Economics*, 2016, *131* (3), 1057–1111.

Lazear, Edward P., "Teacher incentives," *Swedish Economic Policy Review*, 2003, *10*, 179–214.

Loveless, Tom, "International tests are not all the same," *Brookings: Research*, 2013.

Mankiw, N Gregory, David Romer, and David N Weil, "A contribution to the empirics of economic growth," *The quarterly journal of economics*, 1992, *107* (2), 407–437.

Mayda, Anna Maria, "Who is against immigration? A cross-country investigation of individual attitudes toward immigrants," *The review of Economics and Statistics*, 2006, *88* (3), 510–530.

Mulligan, Casey B, "Galton versus the human capital approach to inheritance," *Journal of political Economy*, 1999, *107* (S6), S184–S224.

Murnane, Richard J, John B Willett, Yves Duhaldeborde, and John H Tyler, "How important are the cognitive skills of teenagers in predicting subsequent earnings?," *Journal of Policy Analysis and Management*, 2000, *19* (4), 547–568.

Oberfichtner, Michael and Harald Tauchmann, "Stacked linear regression analysis to facilitate testing of hypotheses across OLS regressions," *The Stata Journal*, 2021, *21* (2), 411–429.

OECD, *Reporting the results* 2013.

_ , PIAAC Numeracy Expert Group et al., "PIAAC numeracy: a conceptual framework," Technical Report, OECD education working paper, Bd. 35 2009.

Roth, Christopher and Johannes Wohlfart, "Experienced inequality and preferences for redistribution," *Journal of Public Economics*, 2018, *167*, 251–262.

Sacerdote, Bruce, "Peer effects in education: How might they work, how big are they and how much do we know thus far?," in "Handbook of the Economics of Education," Vol. 3, Elsevier, 2011, pp. 249–277.

Slomczynski, Kazimierz M and Irina Tomescu-Dubrow, "Basic principles of survey data recycling," *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)*, 2018, pp. 937–962.

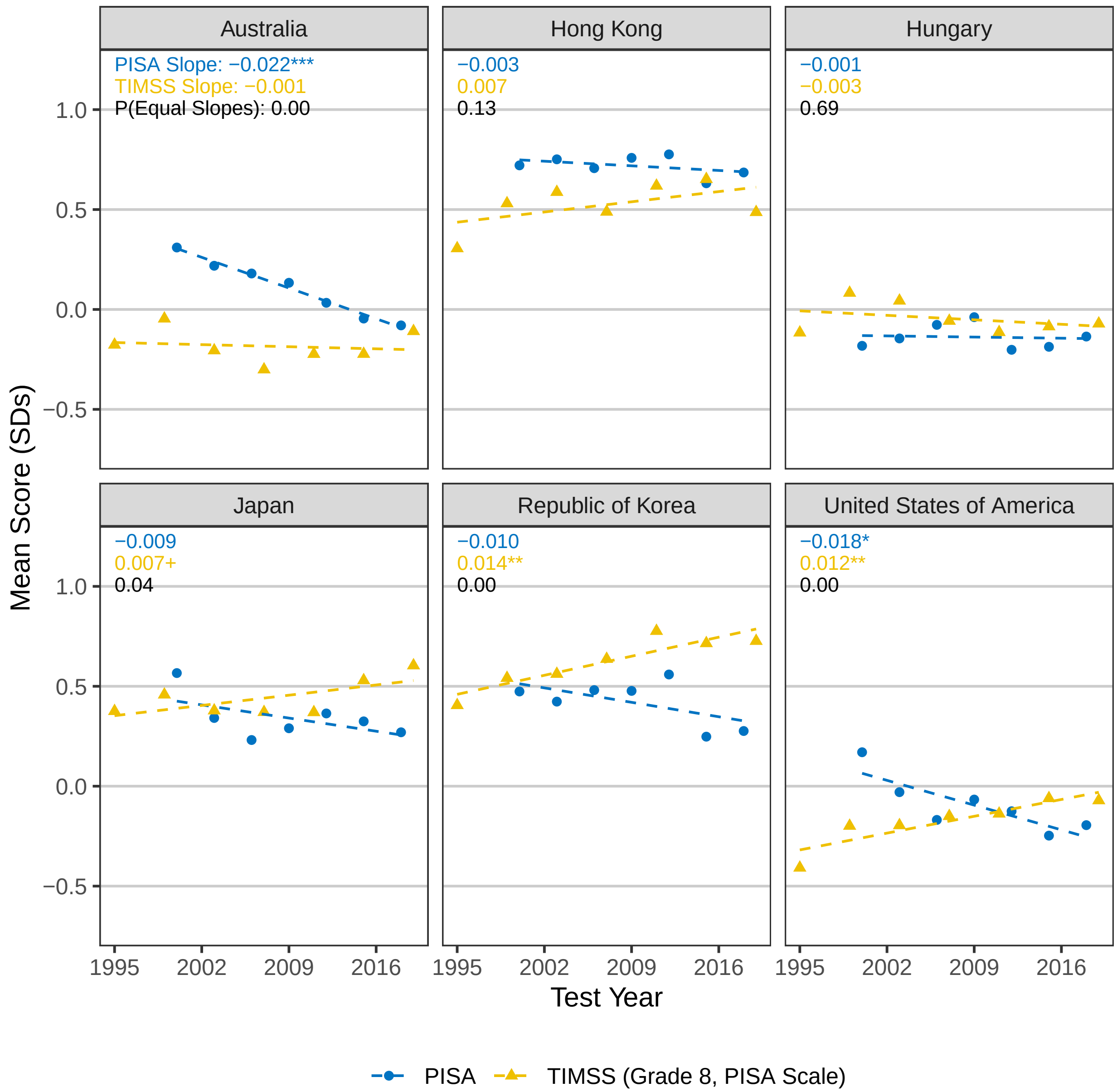


Figure 1: Trends in PISA and TIMSS Grade 8 Math Scores Across Continuously Participating Countries

Note: Figure displays average PISA and TIMSS Grade 8 math scores among the 6 countries that have participated in every round of PISA and TIMSS Grade 8 testing since their inception in 1995 and 2000, respectively. Both PISA and TIMSS scores are transformed to the PISA scale based on the procedure in Gust et al. (2024) after incorporating the sample restrictions described in Section 2. Text at the top left of each panel summarizes the results of the following regression, run separately for each country and test (PISA or TIMSS Grade 8):

$$AvgScore_{ct} = \beta_0 + \beta_1 Year_t + \varepsilon_{ct}$$

The displayed slope corresponds to my estimate of β_1. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

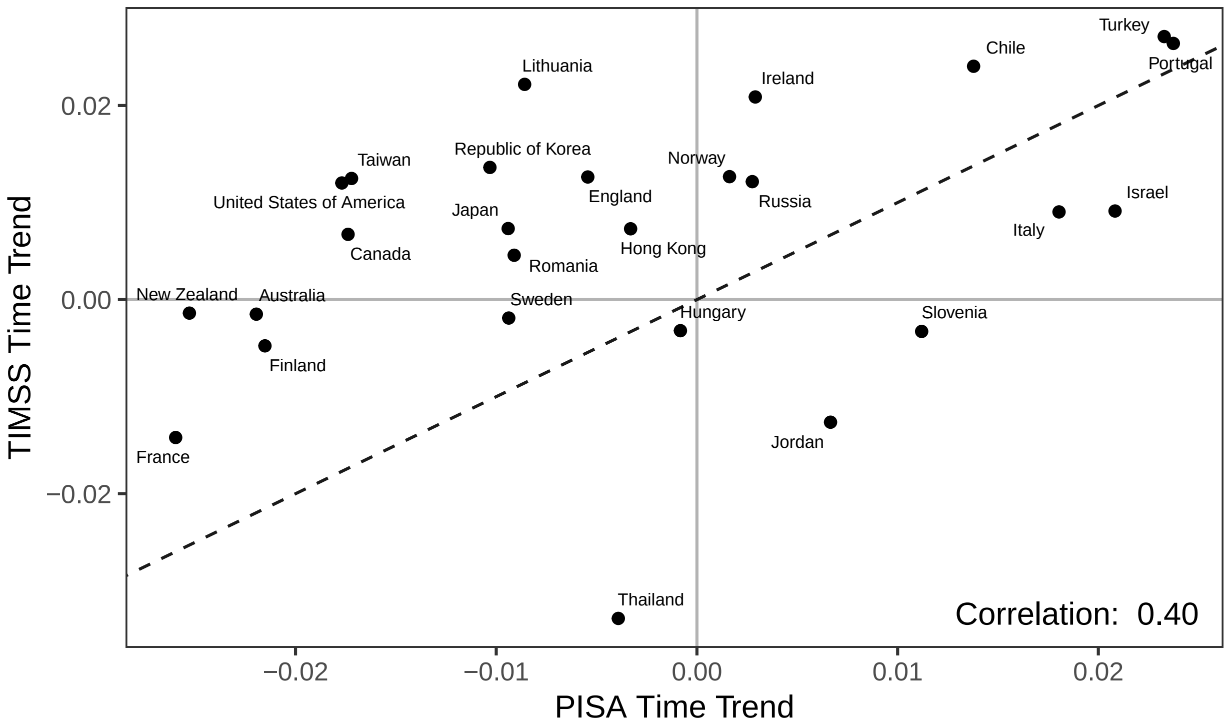


Figure 2: Growth in PISA and TIMSS Grade 8 Math Scores

Note: Figure displays estimated time trends for average PISA and TIMSS Grade 8 math scores across countries. Displayed countries are those with at least one PISA test and one TIMSS Grade 8 test in 2006 or earlier and at least one PISA test and one TIMSS Grade 8 test in 2012 or later. Both PISA and TIMSS scores are transformed to the PISA scale based on the procedure in Gust et al. (2024) after incorporating the sample restrictions described in Section 2. Time trends are estimated separately for PISA and TIMSS Grade 8 math scores using the regression equation:

$$AvgScore_{ct} = \beta_0 + \beta_1 Year_t + \varepsilon_{ct}$$

The displayed numbers correspond to estimates of β_1. Dashed line is the 45-degree line.

Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Panel A: PIAAC Sample					
Female	221	0.523	0.066	0.357	0.690
Age	221	23.199	4.464	16	32
Survey Year	221	2,012.896	1.315	2,012	2,017
Year of Birth	221	1,989.697	4.774	1,982	2,001
PISA Math Score	221	−0.041	0.422	−0.922	1.029
TIMSS Math Score (Grade 8, PISA Scale)	221	−0.277	0.383	−1.260	0.779
PIAAC Numeracy Score	221	0.217	0.388	−0.856	0.946
Panel B: SDR Sample					
Female	3,417	0.522	0.150	0.000	1.000
Age	3,417	22.961	4.464	16	35
Age $\mid$ 19 +	2,824	24.142	3.994	19	35
Age $\mid$ 24 +	1,391	27.497	2.855	24	35
Survey Year	3,417	2,010.567	4.308	1,998	2,017
Year of Birth	3,417	1,987.605	4.404	1,982	2,001
PISA Math Score	3,417	−0.059	0.504	−1.810	1.029
TIMSS Grade 8 Math Score (PISA Scale)	3,417	−0.268	0.446	−2.019	0.779
Years of Education	3,086	13.127	1.962	4.500	22.000
Complete Secondary School $\mid$ 19 +	2,824	0.828	0.165	0.000	1.000
Complete Bachelors $\mid$ 24 +	1,391	0.355	0.206	0.000	1.000
Percentile of HH Income $\mid$ 24 +	1,336	53.291	11.846	4.000	99.000
Number of Survey Responses	3,417	30.691	22.841	1	182
Source: Asian Barometer	3,417	0.034	0.171	0.000	1.000
Source: Afrobarometer	3,417	0.005	0.064	0.000	1.000
Source: Americas Barometer	3,417	0.009	0.094	0.000	1.000
Source: Arab Barometer	3,417	0.015	0.121	0.000	1.000
Source: Asia Europe Survey	3,417	0.003	0.048	0.000	1.000
Source: Caucasus Barometer	3,417	0.005	0.067	0.000	1.000
Source: Consolidation of Democracy	3,417	0.001	0.026	0.000	1.000
Source: Comparative National Elections Project	3,417	0.006	0.066	0.000	1.000
Source: Eurobarometer	3,417	0.001	0.027	0.000	1.000
Source: European Quality of Life Survey	3,417	0.048	0.179	0.000	1.000
Source: European Social Survey	3,417	0.266	0.379	0.000	1.000
Source: European Values Study	3,417	0.024	0.131	0.000	1.000
Source: International Social Survey Programme	3,417	0.410	0.414	0.000	1.000
Source: Latinobarometro	3,417	0.028	0.147	0.000	1.000
Source: Life in Transition Survey	3,417	0.077	0.244	0.000	1.000
Source: New Baltic Barometer	3,417	0.001	0.021	0.000	0.656
Source: New Europe Barometer	3,417	0.006	0.065	0.000	1.000
Source: World Values Survey	3,417	0.063	0.212	0.000	1.000

Note: Table displays summary statistics for PIAAC and SDR data in Panels A and B, respectively. Both PISA and TIMSS scores are transformed to the PISA scale based on the procedure in Gust et al. (2024) after incorporating the sample restrictions described in Section 2.

Table 2: Relationship Between Adulthood Numeracy and Adolescent Test Scores by Country Birth Cohort

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: PIAAC Numeracy Test Score						
PISA	0.198*		0.157+	0.217*		0.181*
	(0.090)		(0.075)	(0.098)		(0.080)
TIMSS (Grade 8, PISA Scale)		0.221*	0.168*		0.217+	0.176+
		(0.104)	(0.074)		(0.113)	(0.086)
Num.Obs.	221	221	221	221	221	221
R2	0.938	0.938	0.940	0.943	0.943	0.945
p-value: PISA = TIMSS	-	0.686	0.879	-	0.996	0.944
Country FEs	✓	✓	✓	✓	✓	✓
Age and Age^2	✓	✓	✓	✓	✓	✓
Test Year FEs	✓	✓	✓	✓	✓	✓
Region-by-Age and Region-by-Age^2				✓	✓	✓

Note: Table displays regression results estimating the relationship between average PISA and TIMSS math scores and average PIAAC numeracy scores. In the table, PISA and TIMSS represent average PISA and TIMSS scores at the country-by-birth cohort level. Both PISA and TIMSS scores are transformed to the PISA scale based on the procedure in Gust et al. (2024) after incorporating the sample restrictions described in Section 2. Region refers to World Bank regions: East Asia and Pacific, Europe and Central Asia, Latin America and Caribbean, Middle East and North Africa, and North America. Robust standard errors clustered at the country and birth year level. p-values shown in Columns 2 and 5 reflect the results of a test of equality of coefficients for the estimates in Columns 1 and 2, and Columns 4 and 5, respectively. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3: Relationship Between Adulthood Outcomes and Adolescent Math Test Scores by Country Birth Cohort

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Years of Education; Sample: 16+ Years Old									
PISA	0.739*		0.715*	0.540		0.532	0.938		0.956+
	(0.345)		(0.335)	(0.423)		(0.410)	(0.558)		(0.542)
TIMSS (Grade 8, PISA Scale)		0.236	0.125		0.121	0.045		-0.005	-0.106
		(0.194)	(0.122)		(0.188)	(0.134)		(0.246)	(0.178)
Num.Obs.	3086	3086	3086	3086	3086	3086	3086	3086	3086
R2	0.745	0.744	0.745	0.759	0.758	0.759	0.891	0.889	0.891
p-value: PISA = TIMSS	-	0.091	0.075	-	0.187	0.183	-	0.023	0.016
Panel B: Complete Secondary School; Sample: 19+ Years Old									
PISA	0.122*		0.118*	0.108+		0.105+	0.118*		0.118*
	(0.054)		(0.051)	(0.058)		(0.055)	(0.050)		(0.048)
TIMSS (Grade 8, PISA Scale)		0.045	0.032		0.035	0.023		0.015	0.003
		(0.050)	(0.039)		(0.048)	(0.038)		(0.037)	(0.029)
Num.Obs.	2824	2824	2824	2824	2824	2824	2824	2824	2824
R2	0.511	0.507	0.512	0.537	0.533	0.537	0.768	0.764	0.768
p-value: PISA = TIMSS	-	0.181	0.150	-	0.183	0.162	-	0.027	0.018
Panel C: Complete Bachelors; Sample: 24+ Years Old									
PISA	-0.071		-0.071	-0.062		-0.062	0.013		0.010
	(0.063)		(0.063)	(0.062)		(0.062)	(0.037)		(0.036)
TIMSS (Grade 8, PISA Scale)		0.046+	0.045+		0.046	0.046		0.025	0.025
		(0.024)	(0.023)		(0.031)	(0.031)		(0.025)	(0.024)
Num.Obs.	1391	1391	1391	1391	1391	1391	1391	1391	1391
R2	0.567	0.567	0.567	0.578	0.578	0.579	0.801	0.801	0.801
p-value: PISA = TIMSS	-	0.112	0.082	-	0.140	0.107	-	0.719	0.675
Panel D: Percentile of Household Income; Sample: 24+ Years Old									
PISA	7.470**		7.608**	7.512***		7.549***	12.127**		11.749**
	(2.357)		(2.334)	(1.665)		(1.591)	(3.684)		(3.572)
TIMSS (Grade 8, PISA Scale)		3.908*	4.031*		3.427*	3.465+		3.859*	3.303
		(1.312)	(1.494)		(1.361)	(1.578)		(1.712)	(1.844)
Num.Obs.	1336	1336	1336	1336	1336	1336	1336	1336	1336
R2	0.566	0.565	0.567	0.585	0.584	0.587	0.715	0.711	0.716
p-value: PISA = TIMSS	-	0.256	0.216	-	0.096	0.067	-	0.027	0.013
Country FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Age FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Survey Year FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region-by-Age FEs				✓	✓	✓	✓	✓	✓
Country-by-Survey Year FEs							✓	✓	✓

Note: Table displays regression results estimating the relationship between average PISA and TIMSS math scores and average adulthood outcomes in SDR data. In the table, PISA and TIMSS represent average PISA and TIMSS scores at the country-by-birth cohort level. Both PISA and TIMSS scores are transformed to the PISA scale based on the procedure in Gust et al. (2024) after incorporating the sample restrictions described in Section 2. Region refers to World Bank regions: East Asia and Pacific, Europe and Central Asia, Latin America and Caribbean, Middle East and North Africa, and North America. Robust standard errors clustered at the country and birth year level in parentheses. Observations are weighted by SDR-provided weights. p-values shown in Columns 2, 5, and 8 reflect the results of a test of equality of coefficients for the estimates in Columns 1 and 2, Columns 4 and 5, and Columns 7 and 8, respectively. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.