# Trends in International Assessments and Outcomes in Adulthood*

Samuel Stemper†

November 29, 2023

**Abstract**

International assessments such as PISA and TIMSS are widely used to compare the academic proficiency of adolescents across countries and over time. Do scores on these assessments predict outcomes in adulthood? Combining data from PISA, TIMSS, PIAAC, and 18 representative global surveys, I find evidence that cohorts with higher scores on the PISA assessment—which tests the application of knowledge to "real world" scenarios—score higher on assessments of adulthood skills, obtain higher levels of education, and have higher incomes as adults. Conversely, I find little evidence for similar patterns with respect to the TIMSS assessment, which tests curricular knowledge.

*First version: November 29, 2023. This version: November 29, 2023. 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 often use international assessments such as PISA (Programme for International Student Assessment) and TIMSS (Trends in International Mathematics and Science Study) to monitor the academic proficiency of their students. Trends in these assessments are often used as a means to evaluate the trajectory of education systems within a country; differences in test scores across nations are often used to identify promising strategies for education policy.

Do these assessments measure the skills that ultimately equip students for success in adulthood? This paper aims to address this question by investigating whether cohorts that score higher on international assessments during adolescence ultimately exhibit differential educational and scent 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. 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 predict 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.

Across measures of skills, education, and income, I find that both PISA and TIMSS scores are positively associated with outcomes in adulthood. However, I find that cohort-level PISA scores in math exhibit larger and more statistically robust relationships across all three sets of outcomes. A 1 standard deviation cohort-level increase in PISA scores is associated with a 0.3 standard deviation increase in adulthood numeracy test scores, a 1-year increase in years of education, and a 14 percentage-point increase in household income percentile. For TIMSS scores, these effect sizes are much smaller and generally statistically insignificant.

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 is Doty et al. (2022), who study the relationship between changes in National Assessment of Educational Progress (NAEP) scores and outcomes in adulthood across U.S. states, and find that cohorts with higher NAEP scores generally have higher incomes, higher levels of education, and lower rates of teen motherhood, incarceration, and arrest. I employ a comparable approach to theirs, diverging in my use of multiple testing regimes to evaluate different skill measures and my context, which involves changes in skills across countries rather than U.S. states.

A much larger set of literature examines the relationship between measures of skills and adulthood outcomes at the individual, rather than aggregate (country or state), level. Goldhaber and Özek (2019), Hanushek and Woessmann (2008), 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." At the aggregate level, a separate literature studies how country-level differences in test scores and education levels are related to economic growth more broadly (Barro, 1991; Mankiw et al., 1992).

Finally, 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.

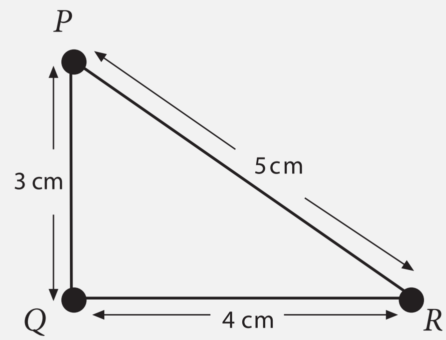
# 1 Background on PISA and TIMSS

## 1.1 Assessment Methodologies

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.

PISA and TIMSS differ substantially in the material used to assess student skills. TIMSS is curriculum-based, and tests students' knowledge of the material that they are taught in school. Alternatively, PISA assesses students' ability to apply knowledge to "real-world" problems (Loveless, 2013). Sample questions from the 2011 8th Grade TIMSS exam and the 2012 PISA exam illustrate this difference. As an example, in 2011, the TIMSS 8th Grade 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. $5 < 3 + 4$

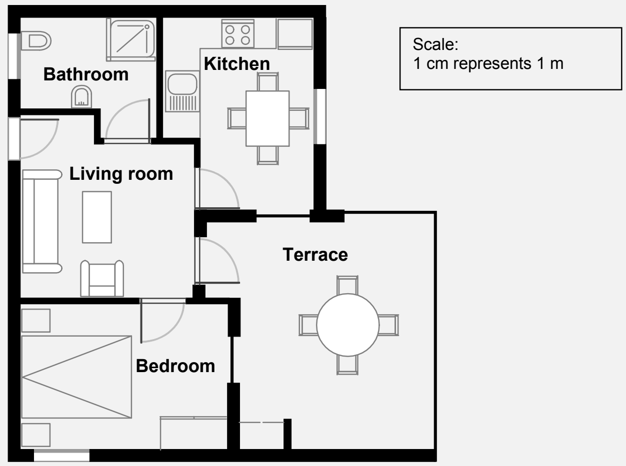
C. $3 + 4 = 12 - 5$

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. Appendix A shows a number of additional example questions from PISA and TIMSS exams.

### 1.2 Trends in PISA and TIMSS

PISA and TIMSS not only differ in the content of their assessments; they differ 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 8th Grade assessments in 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 8th Grade math assessments to date. Lines show linear time trends, separately for PISA and TIMSS; the 18-year change associated with this time trend (e.g. the estimated difference between 2000 and 2018 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 standard deviations between 2000 and 2018. Over the same period, TIMSS 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.[1] 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.

In Appendix Table B.1, I formally test for differences in the trajectories of PISA and TIMSS assessments using student-level data from all PISA and TIMSS participating countries. On average, TIMSS 4th and 8th Grade math scores grew 0.023 and 0.010 standard deviations faster per year than PISA math scores. 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. Over an 18-year period, this translates to a divergence of 0.18 standard deviations between TIMSS 8th Grade math and PISA math scores.

This phenomenon, while significant and widespread, has not unfolded uniformly across all nations. Figure 2 illustrates country-level variation in long-term test score trends. Specifically, the horizontal axis displays the change in PISA math score from the earliest two rounds of testing in 2000 and 2003 to the two most recent rounds in 2015 and 2018. The vertical axis displays the equivalent change in TIMSS 8th Grade math scores from the two rounds of testing in 1999 and 2003 to the testing rounds in 2015 and 2019. Consistent with the evidence presented above, most countries exhibit 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 and Japan, PISA scores have decreased by more than 0.15 standard deviations, while TIMSS scores have increased by a corresponding or greater amount.

The data and methods described in the section that follows use this variation in test scores—

[1] 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 from sample of test scores from low- and middle-income countries.

over time, across countries, and between assessments—to test whether these assessments are predictive of outcomes in adulthood.

## 2 Data and Methods

### 2.1 PISA and TIMSS Test Scores

I prepare country-by-cohort test score averages for PISA, 4th Grade TIMSS, and 8th Grade TIMSS using publicly available student-level test score data. To do so, I compute the average student's birth year and average test score in math for each country and testing round (e.g. 2000 PISA, 2011 4th Grade TIMSS, 2019 8th Grade TIMSS, etc.). Both PISA and TIMSS scores are reported on a scale with mean 500 and standard deviation of 100.[2] I scale these values by dividing by subtracting 500 and dividing by 100. 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.[3]

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 when possible.

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 a larger number of 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 earlier test score.

For example, the United States participated in the 8th Grade TIMSS 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 8th Grade TIMSS score to students born between 1979 and 1983 and assign the 1985 birth cohort's score to students born between 1984 and 1987.

For birth cohorts with both 4th Grade and 8th Grade TIMSS scores available, I take the av-

[2]PISA math scores are normed relative to assessed students in OECD countries in 2003. "Comparisons Over Time on the PISA Scales," OECD. TIMSS math scores are normed relative to assessed students in 1995. "Weighting, scaling, and plausible values," NCES.

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

erage. When only one TIMSS score is available, I use that score. Throughout the remainder of this paper, I refer to this aggregated cohort-level score as the "TIMSS Score" or "Combined TIMSS Score."

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 individual-level outcomes–namely, adult PIAAC test scores in numeracy and education and income data from harmonized international surveys. Because this individual 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 "PIAAC Sample" and the "SDR Sample." Appendix Figure B.1 shows the number of observations in each country-cohort combination across both samples.

### 2.2 PIAAC Scores

To assess the numeracy skills of adults, I use individual-level 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.[4]

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). Appendix A includes two sample questions that illustrate the types of questions included in the PIAAC numeracy assessment. 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 individual numeracy test scores such that they have mean 0 and

[4]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.

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.[5]

### 2.3 Harmonized International Survey Data

For education and income, I use harmonized international survey data from the Survey Data Recycling ("SDR") project.[6] 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.[7] 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." SDR raw data contains 4.4 million responses from 23 survey projects. I link these responses to their cohort's PISA and TIMSS test scores; the resulting data contains over 115,000 responses from 18 survey regimes.[8] These surveys took place between 1998 and 2017.

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.[9]

[5] 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.

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

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

[8] The reduction in sample size is accounted for mostly by birth year: 3.7 million respondents in SDR raw data were born prior to 1982, the first birth cohort for which PISA and TIMSS data are available in any country.

[9] SDR documentation describes their process as follows: "To construct the target variable T_EDU_YEARS, we use source items about respondents': (a) exact number of education years completed; (b) age when finished full-time education; (c) year when finished school; (d) years of education derived by survey providers from various source variables. We rely on the English language and Spanish language questionnaires and codebooks describing the source survey data.

Typical questions on respondent's years of education are: "How many years of formal education have you received?" (ABS), "How many years of schooling have you completed?"(AMB/2010-2016), "About how many years of education have you completed, whether full-time or part-time?" (ESS). Typical questions on the age of completing education are: "How old were you when you finished your full-time education?" (CDCEE), "At what age did you finish your education (full-time education)?" (LB). In two cases, respondents were asked to indicate the year of their education completion: "When did you obtain this degree?" (LITS/1), "When did you obtain this qualification?" (LITS/2)."

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

Finally, 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.[10]

### 2.4 Data Description

Table 1 summarizes the main variables across my two samples. Across both samples, slightly more than half of respondents are women and the average age is roughly 23 years old. Average cohort-level PISA scores are lower than TIMSS scores, reflecting the trends discussed above. In my SDR data, the average respondent had 13 years of education. At the time of the interview, 74 percent of respondents had completed secondary school and 19 percent had completed a bachelor's degree.

As noted above, these samples contain slightly different sets of country-by-cohort pairs. Appendix Figure B.1 summarizes the set of countries and birth cohorts in my samples. These samples contain a reasonably large set of cohorts within each country. There are 29 unique countries in the PIAAC sample, 14 of which contain at least 10 unique birth cohorts.[11] There are 55 unique countries in the SDR sample, 31 of which contain at least 10 unique birth cohorts. Table 1 additionally lists the set of 18 survey sources included in my sample.

[10] 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) a respondent's predicted age (survey year minus birth year) and (b) a respondent's report age is larger than 2.

[11] Note that, due to the presence of some countries that report ages in 5-year ranges means that this figure understates the number of cohorts represented.

## 2.5 Methodology

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

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

$y_{icbt}$ denotes an outcome (e.g. years of education, household income percentile) for individual $i$ 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 "horserace" regression that includes both scores. 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: North America, Latin America and Caribbean, Middle East and North Africa, East Asia and Pacific, and Europe and Central Asia.

My coefficient of interest, $\beta_1$, measures the association between cohort-level average test scores and individual outcomes in PIAAC or SDR data. Importantly, these estimates 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 evaluating the predictive validity of test scores on later-life outcomes, rather than the causal effect of skills.

For regressions using PIAAC data, I weight each observation by $w_{ict} / \sum_{i \in ct} w_{ict}$, where $w_{ict}$ is individual $i$ in country $c$ in year $t$'s final sampling weight, and $\sum_{i \in ct} w_{ict}$ denotes the sum of sampling weights in country $c$ in year $t$. In addition, in these regressions, I account for imputation error by using Rubin's rule and the PIAAC plausible values (Rubin, 1987). For regressions using SDR data, I weight estimates using the SDR-provided weights, which rescale sampling weights from each national survey such that the sum within each survey equals the number of respondents. In my robustness checks, I show that I obtain qualitatively similar results when I rescale these weights such that the sum of each country-cohort combination's weights equals 1. Throughout, I cluster standard errors by country.

In addition, I note that income data is not available for all respondents in all surveys. The variation in the availability of income data is driven primarily by differences across, rather than within, survey waves; some survey waves do not ask respondents about their household income. Appendix Table B.4 shows that income nonresponse is not systematically related to average cohort PISA or TIMSS scores.

## 3 Results

Table 2 shows the results of regressions using PIAAC data, which assess the relationship between adolescent test scores and measures of numeracy skills in adulthood. Regression results shown in Columns 1 to 3 include fixed effects for country, age, gender, and test year. Column 1 displays the relationship between a cohort's PISA scores and PIAAC scores in adulthood, which indicates that a 1 standard deviation increase in a cohort's average PISA score is associated with a 0.3 standard deviation increase in that cohort's PIAAC scores in adulthood. 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 0.1 increase in that cohort's PIAAC scores; this effect is not statistically significant. 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 marginally significant ($p = 0.092$s). 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 PISA and PIAAC are over three times as large as associations between TIMSS and PIAAC.

Columns 4 to 6 of Table 2 employ an additional specification by incorporating region-by-age fixed effects. The inclusion of these controls does not substantially alter the estimates. These estimates demonstrate that cohort-level variation in adolescent math skills indeed translates into later-life differences in numeracy skills, and offer suggestive evidence that PISA scores may be more strongly associated with adult skills than TIMSS scores. To assess the persistence of these patterns with respect to education and income in adulthood, I next turn to harmonized international survey data.

Table 3 shows the results of 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 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 cohort-level PISA scores is associated with a 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 negative. I test for differences between these two estimates—PISA and TIMSS effects—and display corresponding p-values below coefficient estimates in Panel A; these estimates consistently 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 shows that both PISA and TIMSS test scores are associated with higher rates of secondary school completion. Effect sizes suggest that a 1 standard deviation increase in test scores increases secondary school completion by 5 to 10 percentage points. However, these estimates are imprecise and only statistically significant in some specifications. The results in Panel C suggest that PISA scores are associated with higher rates of tertiary school completion, but the effects are imprecise and not statistically significant. I find no evidence that TIMSS scores are associated with higher rates of tertiary school completion.

In Appendix Figure B.2, I attempt to summarize these results by estimating educational effects in specific ranges. Specifically, for both PISA and TIMSS scores, I estimate effects on binary indicators for ranges of years of education (e.g. 0 to 8 years of education, 9 to 11 years of education, etc.). Appendix Figure B.2 illustrates that cohorts with higher PISA scores are generally less likely to drop out of secondary school and more likely to complete tertiary school. Effect sizes for TIMSS scores are smaller and less consistent directionally.

Finally, in Panel D of Table 3 I report the effects on household income percentages. Among respondents aged 24 or older, a 1 standard deviation increase in cohort PISA scores is associated with a 12 to 15 percentile increase in household income. TIMSS scores exhibit smaller, marginally significant effects, suggesting that a 1 standard deviation increase in TIMSS scores is associated with a 4 percentile increase in household income. However, I caution that tests for differences between these coefficients are imprecise.

I subject these results to a number of robustness tests, which appear in Appendix B. First, I confirm that my estimates are not driven by any one country. To do so, I reproduce my estimates 10 times, once after excluding each of the largest 10 countries individually. The results are shown in Appendix Figure B.3, which shows that my results are stable, regardless of which country is excluded. Next, I show that my results with respect to income are qualitatively similar when I control for fixed effects for household size. These results are shown in Appendix Table B.2. I also test whether using alternative weights changes my results; Appendix Table B.3 shows that I obtain similar results when I rescale SDR-provided weights such that the sum of each country-cohort combination's weights equals 1.

Finally, I note that PIAAC data does include many adulthood outcomes that also appear in SDR data. However, because most PIAAC testing took place in 2012, sample sizes for older respondents are extremely small and contain very little variation in test scores. For example, among respondents 24 years or older, no country has more than 2 unique PISA scores. Still, I test for effects on education and wages among a slightly expanded set of respondents and find qualitatively similar (but less precise) patterns. These results are shown in Appendix Table B.5.

## 4 Conclusion

In this paper, I estimate the degree to which outcomes in adulthood are explained by cohort-level variation in test scores. Comparing results based on two large-scale international testing regimes—PISA and TIMSS—I find that math scores on PISA exams are highly predictive of skills, education, and incomes in adulthood. In contrast, my results suggest that TIMSS scores are not systematically predictive of outcomes in adulthood. 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. My results suggest that PISA scores are more predictive of future educational and economic success, highlighting a concern for young cohorts who have generally performed worse than cohorts before them.

With respect to 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

**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.

**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.

**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.

**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.

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

**_ and Ludger Woessmann**, "The role of cognitive skills in economic development," *Journal of economic literature*, 2008, *46* (3), 607–668.

**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.

**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.

**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.

**Rubin, Donald B.**, *Multiple imputation for nonresponse in surveys*, John Wiley  Sons., 1987.

**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.

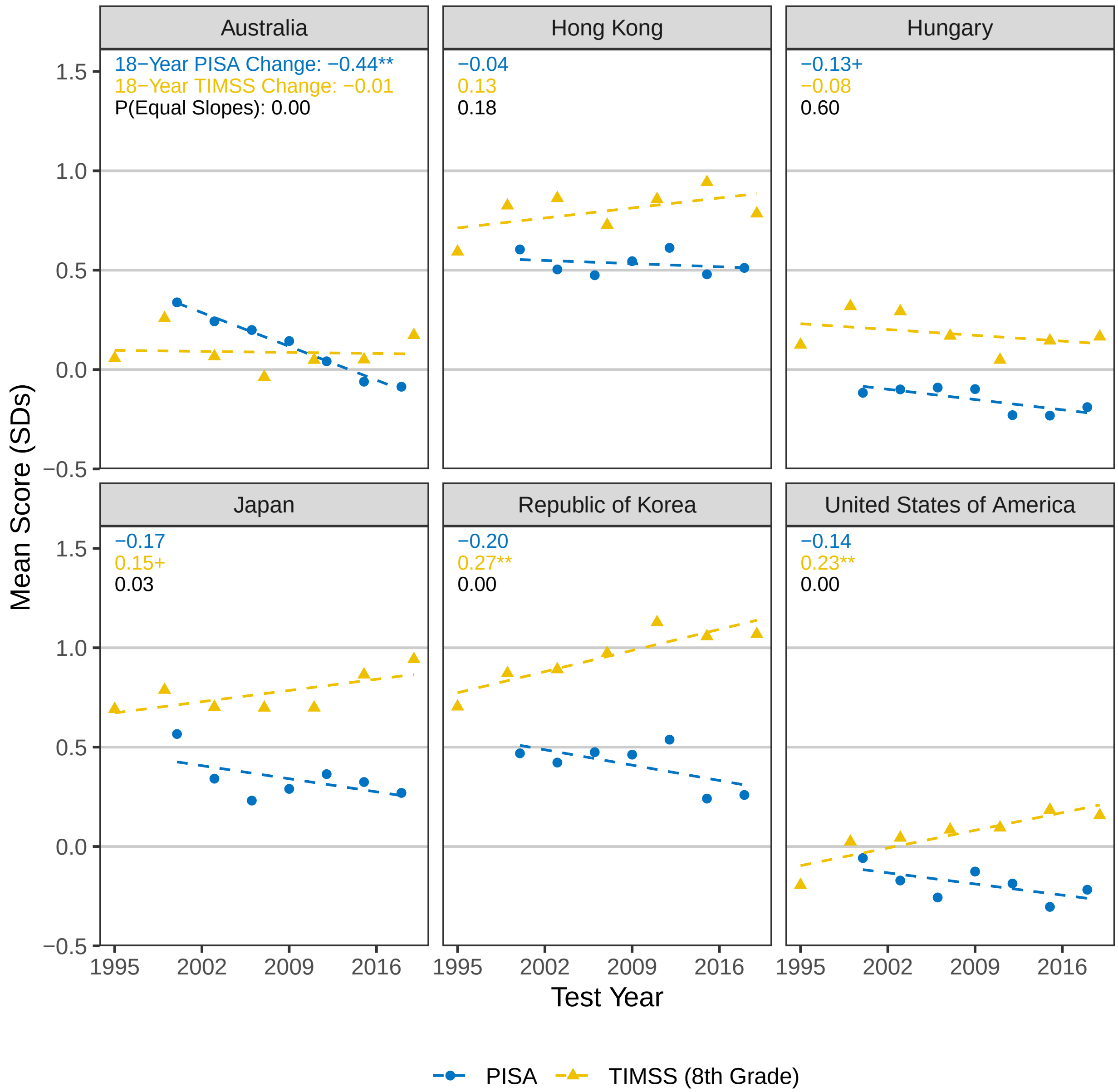



**Figure 1:** Trends in PISA and TIMSS 8th-Grade Math Scores Across Continuously Participating Countries

**Note:** Figure displays average PISA and TIMSS 8th-Grade math scores among the 6 countries that have participated in every round of PISA and TIMSS 8th-Grade testing since their inception in 1995 and 2000, respectively. Both PISA and TIMSS scores are reported on a scale with mean 500 and standard deviation of 100, based on a scale normed relative to students tested in 2003 and 1995, respectively. Displayed scores are scaled by dividing by subtracting 500 and dividing by 100. Figures in the top left of each panel summarize the results of the following regression, run separately for each country and test (PISA or TIMSS 8th-Grade):

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

The displayed number corresponds to $18 \times \beta_1$. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

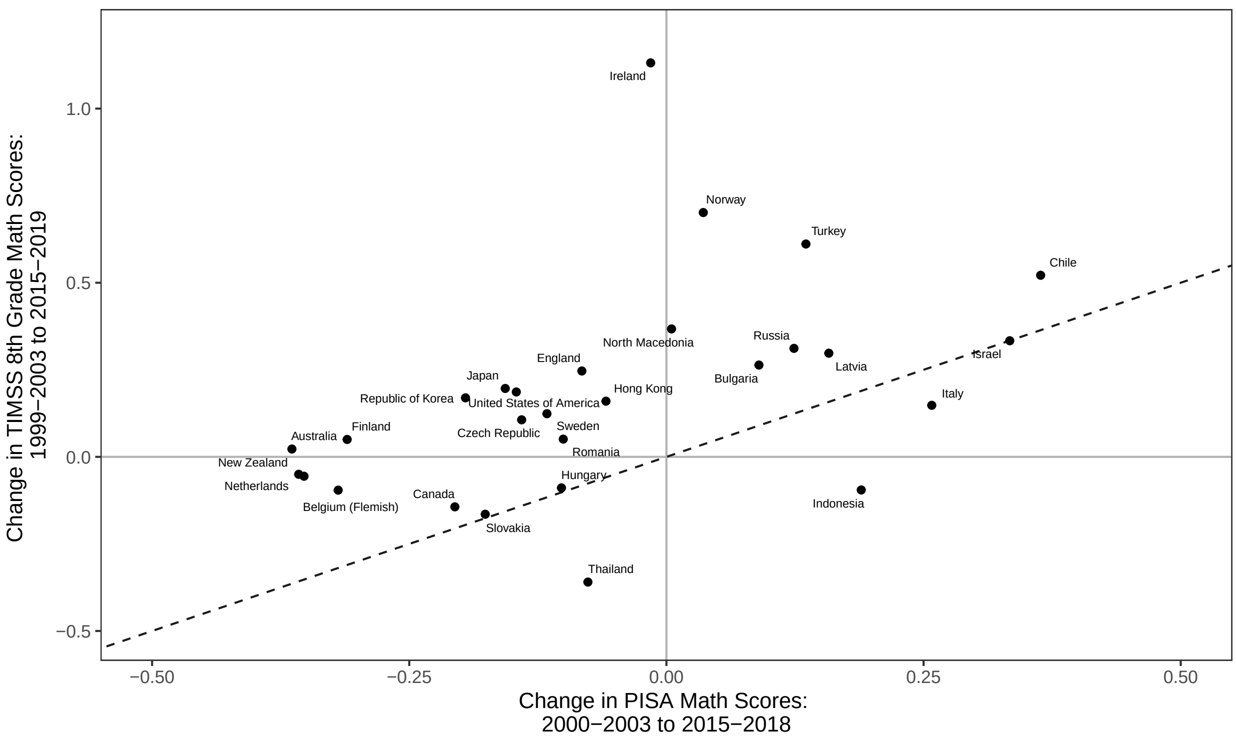



**Figure 2:** Growth in PISA and TIMSS 8th-Grade Math Scores

**Note:** Figure displays changes in average PISA and TIMSS scores across countries. The horizontal axis shows the change in average PISA math scores between 2000-2003 and 2015-2018. The vertical axis shows the change in average TIMSS 8th-Grade math scores between 1999-2003 and 2015-2019. I construct these changes by first taking average scores by country and year. Then I calculate the change between the specified ranges for each test (PISA or TIMSS 8th-Grade). In cases in the specified year range includes 2 tests, I take the average of the 2 scores. Both PISA and TIMSS scores are reported on a scale with mean 500 and standard deviation of 100, based on a scale normed relative to students tested in 2003 and 1995, respectively. Displayed scores are scaled by dividing by subtracting 500 and dividing by 100. Dashed line is the 45-degree line.

**Table 1:** Summary Statistics

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| Panel A: PIAAC Sample | | | | | |
| Female | 35,554 | 0.523 | 0.499 | 0 | 1 |
| Age | 35,555 | 23.246 | 4.513 | 16 | 32 |
| Survey Year | 35,555 | 2,013.074 | 1.428 | 2,012 | 2,017 |
| Year of Birth | 35,555 | 1,989.829 | 4.690 | 1,982 | 2,001 |
| PISA Math Score | 35,555 | −0.054 | 0.384 | −1.163 | 0.735 |
| TIMSS Combined Math Score | 35,555 | 0.014 | 0.386 | −1.130 | 1.052 |
| TIMSS 8th Grade Math Score | 31,406 | 0.029 | 0.405 | −1.130 | 1.110 |
| TIMSS 4th Grade Math Score | 21,839 | 0.068 | 0.379 | −0.710 | 0.994 |
| PIAAC Numeracy Score | 35,555 | 0.000 | 1.000 | −4.687 | 3.651 |
| Panel B: SDR Sample | | | | | |
| Female | 115,823 | 0.520 | 0.500 | 0 | 1 |
| Age | 115,848 | 22.847 | 4.222 | 16 | 35 |
| Survey Year | 115,848 | 2,010.261 | 4.123 | 1,998 | 2,017 |
| Year of Birth | 115,848 | 1,987.393 | 4.313 | 1,982 | 2,001 |
| PISA Math Score | 115,848 | −0.199 | 0.492 | −1.820 | 0.641 |
| TIMSS Combined Math Score | 115,848 | −0.091 | 0.483 | −1.925 | 1.128 |
| TIMSS 8th Grade Math Score | 105,184 | −0.061 | 0.488 | −1.925 | 1.128 |
| TIMSS 4th Grade Math Score | 45,723 | −0.059 | 0.481 | −1.726 | 1.068 |
| Years of Education | 100,493 | 13.058 | 3.046 | 0 | 22 |
| Complete Secondary School | 115,848 | 0.740 | 0.438 | 0 | 1 |
| Complete Bachelors | 115,848 | 0.190 | 0.392 | 0 | 1 |
| Percentile of HH Income | 77,996 | 52.304 | 29.076 | 0 | 100 |
| Source: Afrobarometer | 115,848 | 0.005 | 0.070 | 0 | 1 |
| Source: Americas Barometer | 115,848 | 0.007 | 0.085 | 0 | 1 |
| Source: Arab Barometer | 115,848 | 0.016 | 0.126 | 0 | 1 |
| Source: Asia Europe Survey | 115,848 | 0.002 | 0.048 | 0 | 1 |
| Source: Asian Barometer | 115,848 | 0.026 | 0.158 | 0 | 1 |
| Source: Caucasus Barometer | 115,848 | 0.006 | 0.075 | 0 | 1 |
| Source: Comparative National Elections Project | 115,848 | 0.010 | 0.100 | 0 | 1 |
| Source: Consolidation of Democracy | 115,848 | 0.001 | 0.023 | 0 | 1 |
| Source: Eurobarometer | 115,848 | 0.001 | 0.034 | 0 | 1 |
| Source: European Quality of Life Survey | 115,848 | 0.046 | 0.209 | 0 | 1 |
| Source: European Social Survey | 115,848 | 0.262 | 0.440 | 0 | 1 |
| Source: European Values Study | 115,848 | 0.029 | 0.168 | 0 | 1 |
| Source: International Social Survey Programme | 115,848 | 0.422 | 0.494 | 0 | 1 |
| Source: Latinobarometro | 115,848 | 0.036 | 0.186 | 0 | 1 |
| Source: Life in Transition Survey | 115,848 | 0.053 | 0.224 | 0 | 1 |
| Source: New Baltic Barometer | 115,848 | 0.001 | 0.035 | 0 | 1 |
| Source: New Europe Barometer | 115,848 | 0.005 | 0.071 | 0 | 1 |
| Source: World Values Survey | 115,848 | 0.072 | 0.259 | 0 | 1 |

**Note:** Table displays summary statistics for PIAAC and SDR data in Panels A and B, respectively. Both PISA and TIMSS scores are reported on a scale with mean 500 and standard deviation of 100, based on a scale normed relative to students tested in 2003 and 1995, respectively. Displayed scores are scaled by dividing by subtracting 500 and dividing by 100.

**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 | 0.330** | | 0.321** | 0.335* | | 0.334* |
| | (0.105) | | (0.102) | (0.126) | | (0.123) |
| TIMSS | | 0.099 | 0.077 | | 0.054 | 0.050 |
| | | (0.098) | (0.089) | | (0.107) | (0.101) |
| Num.Obs. | 32087 | 32087 | 32087 | 32087 | 32087 | 32087 |
| R2 | 0.123 | 0.123 | 0.123 | 0.125 | 0.125 | 0.125 |
| p-value: PISA = TIMSS | - | 0.092 | 0.105 | - | 0.078 | 0.087 |
| Country FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Age and Gender FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Test Year FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Region-by-Age FEs | | | | ✓ | ✓ | ✓ |

**Note:** Table displays regression results estimating the relationship between average PISA and TIMSS math scores and PIAAC skills. In the table, PISA and TIMSS represent average PISA and TIMSS scores at the country-by-birth cohort level, transformed to reflect student-level standard deviations. Region refers to World Bank regions: North America, Latin America and Caribbean, Middle East and North Africa, East Asia and Pacific, and Europe and Central Asia. Robust standard errors clustered at the country level and adjusted for multiple imputation using Rubin's rule in parentheses. Observations are weighted by $w_{ict}/\sum_{i \in ct} w_{ict}$, where $w_{ict}$ is individual $i$ in country $c$ in year $t$'s final sampling weight, and $\sum_{i \in ct}$ denotes the sum of sampling weights in country $c$ in year $t$. 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 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 | 1.046** | | 1.081** | 0.735+ | | 0.786+ | 1.111* | | 1.286* |
| | (0.341) | | (0.346) | (0.426) | | (0.438) | (0.540) | | (0.533) |
| TIMSS | | -0.132 | -0.208 | | -0.140 | -0.212 | | -0.325+ | -0.469* |
| | | (0.184) | (0.194) | | (0.187) | (0.189) | | (0.176) | (0.176) |
| Num.Obs. | 100475 | 100475 | 100475 | 100475 | 100475 | 100475 | 100475 | 100475 | 100475 |
| R2 | 0.279 | 0.278 | 0.279 | 0.282 | 0.282 | 0.282 | 0.303 | 0.302 | 0.303 |
| p-value: PISA = TIMSS | - | 0.002 | 0.003 | - | 0.047 | 0.052 | - | 0.006 | 0.003 |
| Panel B: Complete Secondary School; Sample: 19+ Years Old | | | | | | | | | |
| PISA | 0.099+ | | 0.093 | 0.080 | | 0.068 | 0.072 | | 0.061 |
| | (0.059) | | (0.058) | (0.061) | | (0.057) | (0.060) | | (0.058) |
| TIMSS | | 0.057 | 0.053 | | 0.066* | 0.061* | | 0.041 | 0.035 |
| | | (0.034) | (0.033) | | (0.030) | (0.029) | | (0.027) | (0.025) |
| Num.Obs. | 97434 | 97434 | 97434 | 97434 | 97434 | 97434 | 97434 | 97434 | 97434 |
| R2 | 0.103 | 0.102 | 0.103 | 0.107 | 0.107 | 0.107 | 0.135 | 0.135 | 0.135 |
| p-value: PISA = TIMSS | - | 0.431 | 0.476 | - | 0.783 | 0.899 | - | 0.558 | 0.670 |
| Panel C: Complete Bachelors; Sample: 24+ Years Old | | | | | | | | | |
| PISA | 0.047 | | 0.047 | 0.060 | | 0.060 | 0.048 | | 0.049 |
| | (0.093) | | (0.093) | (0.081) | | (0.081) | (0.080) | | (0.080) |
| TIMSS | | 0.018 | 0.018 | | 0.005 | 0.003 | | 0.000 | -0.002 |
| | | (0.037) | (0.037) | | (0.041) | (0.041) | | (0.041) | (0.042) |
| Num.Obs. | 45808 | 45808 | 45808 | 45808 | 45808 | 45808 | 45808 | 45808 | 45808 |
| R2 | 0.081 | 0.081 | 0.081 | 0.083 | 0.083 | 0.083 | 0.109 | 0.109 | 0.109 |
| p-value: PISA = TIMSS | - | 0.778 | 0.781 | - | 0.548 | 0.560 | - | 0.572 | 0.588 |
| Panel D: Percentile of Household Income; Sample: 24+ Years Old | | | | | | | | | |
| PISA | 12.537* | | 12.522* | 13.029* | | 12.671* | 14.653* | | 13.866* |
| | (5.360) | | (5.418) | (5.004) | | (5.172) | (6.462) | | (6.697) |
| TIMSS | | 4.058+ | 4.048+ | | 3.935+ | 3.673 | | 4.072+ | 3.483 |
| | | (2.189) | (2.192) | | (2.286) | (2.295) | | (2.315) | (2.227) |
| Num.Obs. | 33707 | 33707 | 33707 | 33707 | 33707 | 33707 | 33707 | 33707 | 33707 |
| R2 | 0.084 | 0.084 | 0.084 | 0.086 | 0.086 | 0.086 | 0.103 | 0.103 | 0.103 |
| p-value: PISA = TIMSS | - | 0.182 | 0.184 | - | 0.131 | 0.153 | - | 0.117 | 0.159 |
| Age FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Survey Year FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Survey Wave 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 adulthood outcomes in SDR data. In the table, PISA and TIMSS represent average PISA and TIMSS scores at the country-by-birth cohort level, transformed to reflect student-level standard deviations. Region refers to World Bank regions: North America, Latin America and Caribbean, Middle East and North Africa, East Asia and Pacific, and Europe and Central Asia. Robust standard errors clustered at the country 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. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

# A Example PISA and TIMSS Test Questions

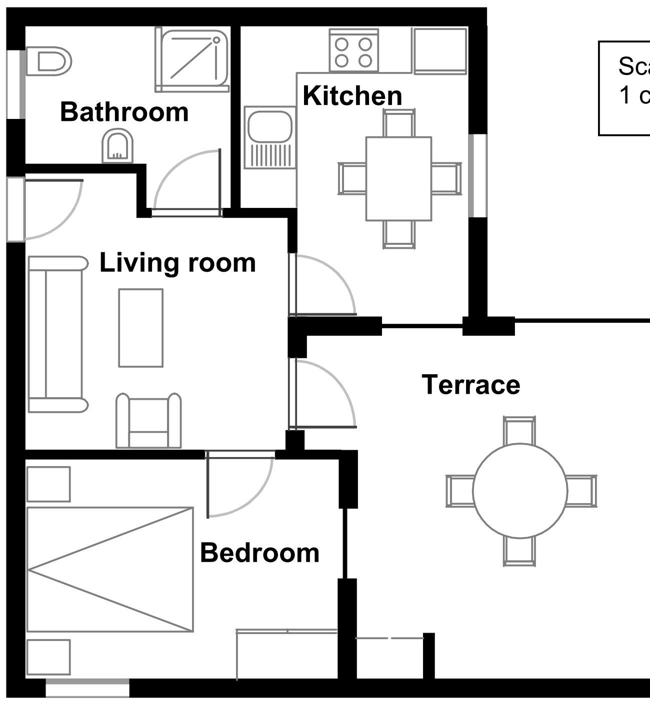
The extracts below are taken from publicly available PISA, TIMSS, and PIAAC documentation. The full documents are available at the links below:

- OECD: Sample PIAAC questions and questionnaire
- NCES: TIMSS Released Assessment Questions
- OECD: PISA Test Questions

# SECTION 1: PISA 2012 MAIN SURVEY ITEMS

## APARTMENT PURCHASE

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




Scale:
1 cm represents 1 m

**Translation Note: In this unit please retain metric units throughout.**

**Translation Note: Translate the term "real estate agency" into local terminology for businesses that sell houses.**

### Question 1: APARTMENT PURCHASE

*PM00FQ01 – 0 1 9*

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.

**Translation Note: In some languages the term used for "area" varies according to the context. As this unit focuses on the areas of rooms, you may choose to use in the first instance here both terms with one between parentheses as in the FRE source version: "La superficie (l'aire) totale de l'appartement".**

**APARTMENT PURCHASE SCORING 1**

QUESTION INTENT:

Description: Use spatial reasoning to show on a plan (or by some other method) the minimum number of side lengths needed to determine floor area

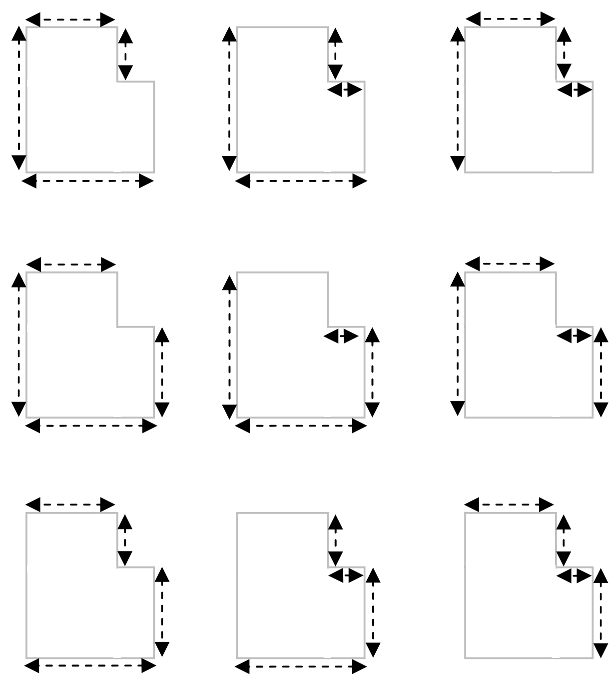
Mathematical content area: Space and shape

Context: Personal

Process: Formulate

***Full Credit***

Code 1: Has indicated the four dimensions needed to estimate the floor area of the apartment on the plan. There are 9 possible solutions as shown in the diagrams below.



- A = (9.7m x 8.8m) – (2m x 4.4m), A = 76.56m$^2$ *[Clearly used only 4 lengths to measure and calculate required area.]*

***No Credit***

Code 0: Other responses.

Code 9: Missing.

# DRIP RATE

Infusions (or intravenous drips) are used to deliver fluids and drugs to patients.

Nurses need to calculate the drip rate, *D*, in drops per minute for infusions.

They use the formula $D = \frac{dv}{60n}$ where

> *d* is the drop factor measured in drops per millilitre (mL)
>
> *v* is the volume in mL of the infusion
>
> *n* is the number of hours the infusion is required to run.

**Translation Note: Use relevant and appropriate words or expressions that are used for infusions (or intravenous drips) – there does not need to be two equivalent terms used if one is well known. Please also avoid using names of infusion brands (such as Baxter in French) even if they are well known.**

**Translation Note: Please use the appropriate scientific/medical term for "drop factor".**

**Translation Note: Please use consistently the national convention for writing the abbreviation of millilitre (mL or ml).**

**Translation Note: Initial letters of key words used in the formula may be adapted to suit a different language, but be careful that the changed letter does not conflict with letters used elsewhere in a unit or the formula.**

**Translation Note: Change expressions and symbols into the standard conventions for writing formulas in your country. For example, you may need to insert a sign like ˙ or * or × between the pronumerals to specify multiplication.**

## Question 1: DRIP RATE

*PM903Q01 – 0 1 2 9*

A nurse wants to double the time an infusion runs for.

Describe precisely how *D* changes if *n* is **doubled** but *d* and *v* do not change.

...................................................................................................................................

...................................................................................................................................

...................................................................................................................................

### DRIP RATE SCORING 1

QUESTION INTENT:

Description: Explain the effect that doubling one variable in a formula has on the resulting value if other variables are held constant

Mathematical content area: Change and relationships

Context: Occupational

Process: Employ

***Full Credit***

Code 2: Explanation describes <u>both</u> the <u>direction</u> of the effect and its <u>size</u>.

- It halves
- It is half
- D will be 50% smaller
- D will be half as big

***Partial Credit***

Code 1: A response which correctly states EITHER the direction OR the size of the effect, but not BOTH.

- D gets smaller *[no size]*
- There's a 50% change *[no direction]*
- D gets bigger by 50%. *[incorrect direction but correct size]*

***No Credit***

Code 0: Other responses.

- D will also double *[Both the size and direction are incorrect.]*

Code 9: Missing.

## Question 3: DRIP RATE

*PM903Q03 – 0 1 9*

Nurses also need to calculate the volume of the infusion, *v*, from the drip rate, *D*.

An infusion with a drip rate of 50 drops per minute has to be given to a patient for 3 hours. For this infusion the drop factor is 25 drops per millilitre.

What is the volume in mL of the infusion?

Volume of the infusion: .......................... mL

**DRIP RATE SCORING 3**

QUESTION INTENT:

Description: Transpose an equation and substitute two given values
Mathematical content area: Change and relationships
Context: Occupational
Process: Employ

***Full Credit***

Code 1: 360 or a correctly transposed and substituted solution.

- 360
- (60 × 3 × 50) ÷ 25 *[Correct transposition and substitution.]*

***No Credit***

Code 0: Other responses.

Code 9: Missing.

| Content Domain | Main Topic | Cognitive Domain |
|---|---|---|
| **NUMBER** | Fractions and Decimals | Applying |

## Ann and Jenny divide 560 zeds

Ann and Jenny divide 560 zeds between them. If Jenny gets $\frac{3}{8}$ of the money, how many zeds will Ann get?

Answer: ________________

*Item Number: M032064*

### SCORING

**Correct Response**

- 350

**Incorrect Response**

- 210
- 5/8
- Other incorrect (including crossed out, erased, stray marks, illegible, or off task)

### Overall Percent Correct

| Education system | Percent correct |
|---|---|
| Singapore | 76 ▲ |
| Korea, Rep. of | 67 ▲ |
| *Hong Kong-CHN* | 61 ▲ |
| *Chinese Taipei-CHN* | 60 ▲ |
| Finland | 48 ▲ |
| Russian Federation | 48 ▲ |
| Japan | 45 ▲ |
| Israel | 43 ▲ |
| Hungary | 40 ▲ |
| Sweden | 37 ▲ |
| *England-GBR* | 34 ▲ |
| Australia | 34 ▲ |
| Italy | 34 ▲ |
| Lithuania | 33 ▲ |
| Malaysia | 32 ▲ |
| Norway | 30 |
| Kazakhstan | 28 |
| Turkey | 28 |
| New Zealand | 28 |
| International average | 27 |
| **United States** | **25** |
| Slovenia | 25 |
| Ukraine | 24 |
| Armenia | 23 ▼ |
| Georgia | 23 ▼ |
| Tunisia | 21 ▼ |
| Romania | 20 ▼ |
| United Arab Emirates | 17 ▼ |
| Iran, Islamic Rep. of | 17 ▼ |
| Macedonia, Rep. of | 16 ▼ |
| Qatar | 16 ▼ |
| Chile | 14 ▼ |
| Thailand | 13 ▼ |
| *Palestinian Nat'l Auth.* | 12 ▼ |
| Lebanon | 10 ▼ |
| Bahrain | 10 ▼ |
| Indonesia | 9 ▼ |
| Saudi Arabia | 8 ▼ |
| Oman | 7 ▼ |
| Jordan | 7 ▼ |
| Morocco | 6 ▼ |
| Syrian Arab Republic | 6 ▼ |
| Ghana | 3 ▼ |

| Benchmarking education system | Percent correct |
|---|---|
| *Quebec-CAN* | 45 ▲ |
| *North Carolina-USA* | 40 ▲ |
| *Minnesota-USA* | 38 ▲ |
| *Massachusetts-USA* | 36 ▲ |
| *Ontario-CAN* | 31 |
| *Connecticut-USA* | 30 |
| *Colorado-USA* | 29 |
| *Alberta-CAN* | 29 |
| *Indiana-USA* | 28 |
| *Dubai-UAE* | 25 |
| *Florida-USA* | 23 |
| *California-USA* | 17 ▼ |
| *Abu Dhabi-UAE* | 15 ▼ |
| *Alabama-USA* | 14 ▼ |

▲ Percent higher than International average
▼ Percent lower than International average

## Ann and Jenny divide 560 zeds (continued)

*M032064*

# Student Responses

### Correct Response:

Answer: 350

### Incorrect Response:

Answer: $\frac{5}{8}$

| Content Domain | Main Topic | Cognitive Domain |
|---|---|---|
| NUMBER | Fractions and Decimals | Knowing |

## 4/100 plus 3/1000

$$\frac{4}{100} + \frac{3}{1000} =$$

A. 0.043

B. 0.1043

C. 0.403

D. 0.43

*Item Number: M032094*

| Correct Response: | A |
|---|---|

## Overall Percent Correct

| Education system | Percent correct | |
|---|---|---|
| Singapore | 92 | ▲ |
| Korea, Rep. of | 89 | ▲ |
| *Hong Kong-CHN* | 86 | ▲ |
| *Chinese Taipei-CHN* | 85 | ▲ |
| Russian Federation | 83 | ▲ |
| Lebanon | 81 | ▲ |
| Japan | 77 | ▲ |
| Israel | 77 | ▲ |
| Italy | 74 | ▲ |
| Malaysia | 73 | ▲ |
| Slovenia | 70 | ▲ |
| Kazakhstan | 69 | ▲ |
| Lithuania | 68 | ▲ |
| Finland | 68 | ▲ |
| United Arab Emirates | 68 | ▲ |
| Tunisia | 68 | ▲ |
| Australia | 68 | ▲ |
| Sweden | 67 | ▲ |
| Armenia | 66 | ▲ |
| Ukraine | 65 | |
| Hungary | 63 | |
| **United States** | **63** | |
| International average | 62 | |
| *Palestinian Nat'l Auth.* | 59 | |
| *England-GBR* | 57 | ▼ |
| Georgia | 55 | ▼ |
| Romania | 54 | ▼ |
| Thailand | 54 | ▼ |
| Macedonia, Rep. of | 52 | ▼ |
| Qatar | 52 | ▼ |
| New Zealand | 51 | ▼ |
| Iran, Islamic Rep. of | 50 | ▼ |
| Turkey | 50 | ▼ |
| Jordan | 49 | ▼ |
| Norway | 49 | ▼ |
| Morocco | 49 | ▼ |
| Saudi Arabia | 49 | ▼ |
| Bahrain | 48 | ▼ |
| Syrian Arab Republic | 48 | ▼ |
| Oman | 47 | ▼ |
| Indonesia | 46 | ▼ |
| Chile | 41 | ▼ |
| Ghana | 34 | ▼ |

| Benchmarking education system | | |
|---|---|---|
| *North Carolina-USA* | 77 | ▲ |
| *Massachusetts-USA* | 76 | ▲ |
| *Minnesota-USA* | 72 | ▲ |
| *Dubai-UAE* | 71 | ▲ |
| *Quebec-CAN* | 71 | ▲ |
| *Connecticut-USA* | 70 | ▲ |
| *Indiana-USA* | 70 | ▲ |
| *Alberta-CAN* | 69 | ▲ |
| *Abu Dhabi-UAE* | 67 | |
| *Ontario-CAN* | 65 | |
| *Florida-USA* | 64 | |
| *Colorado-USA* | 60 | |
| *California-USA* | 58 | |
| *Alabama-USA* | 45 | ▼ |

▲ Percent higher than International average
▼ Percent lower than International average

OECD Skills Surveys

OECD BETTER POLICIES FOR BETTER LIVES

# PIAAC Numeracy – sample items

In the Survey of adult skills (PIAAC). numeracy is defined as the ability to use, apply, interpret, and communicate mathematical information and ideas. It is an essential skill in an age when individuals encounter an increasing amount and wide range of quantitative and mathematical information in their daily lives. Numeracy is a skill parallel to reading literacy, and it is important to assess how these competencies interact, since they are distributed differently across subgroups of the population.

The items are presented in the form delivered by the computer-based version of the assessment. To answer the questions, respondents need to click in the appropriate box, and/or type figures in the space provided.

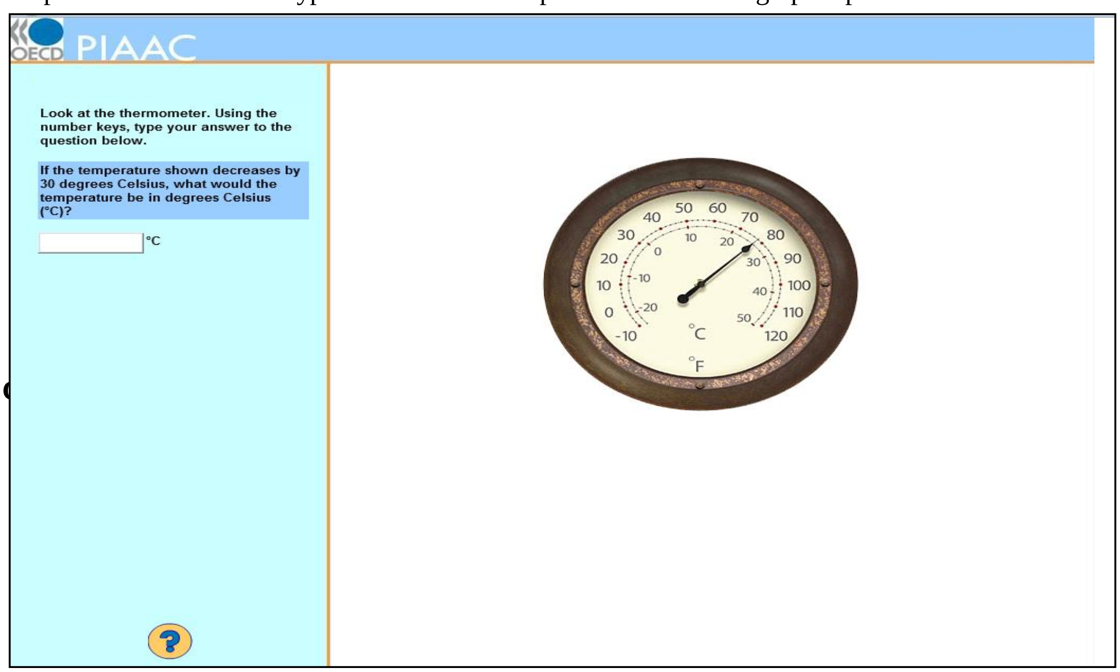
## Numeracy - Sample Items

### *Sample Item 1: Thermometer*

This item (of low difficulty) focuses on the following aspects of the numeracy construct:

| | |
|---|---|
| ***Content*** | Dimension and shape |
| ***Process*** | Act upon, use (measure) |
| ***Context*** | Every day or work |

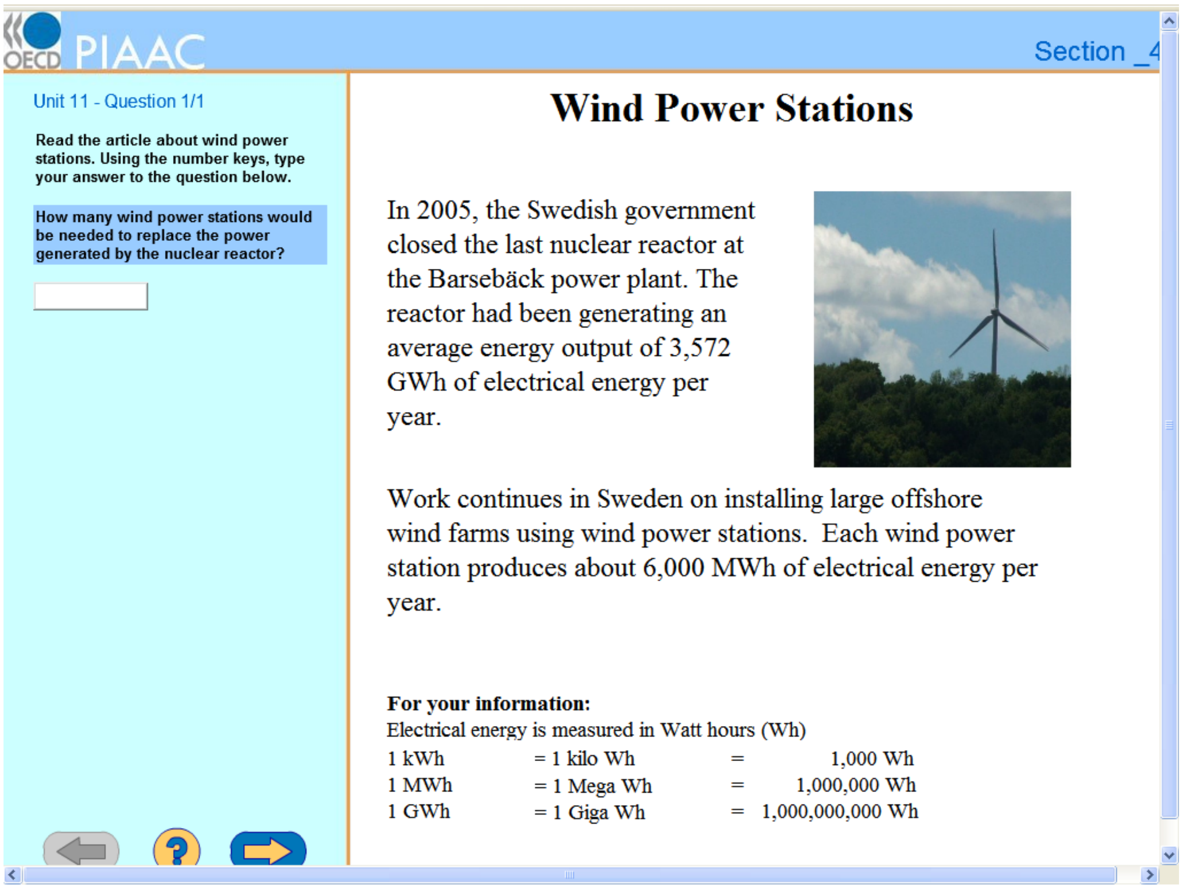
Respondents are asked to type in a numerical response based on the graphic provided.



### *Sample Item 2: Wind power stations*

This sample item (of medium difficulty) focuses on the following aspects of the numeracy construct:

| | |
|---|---|
| ***Content*** | Quantity and Number |
| ***Process*** | Act upon, use (compute) |
| ***Context*** | Community and society |



# B Additional Tables and Figures

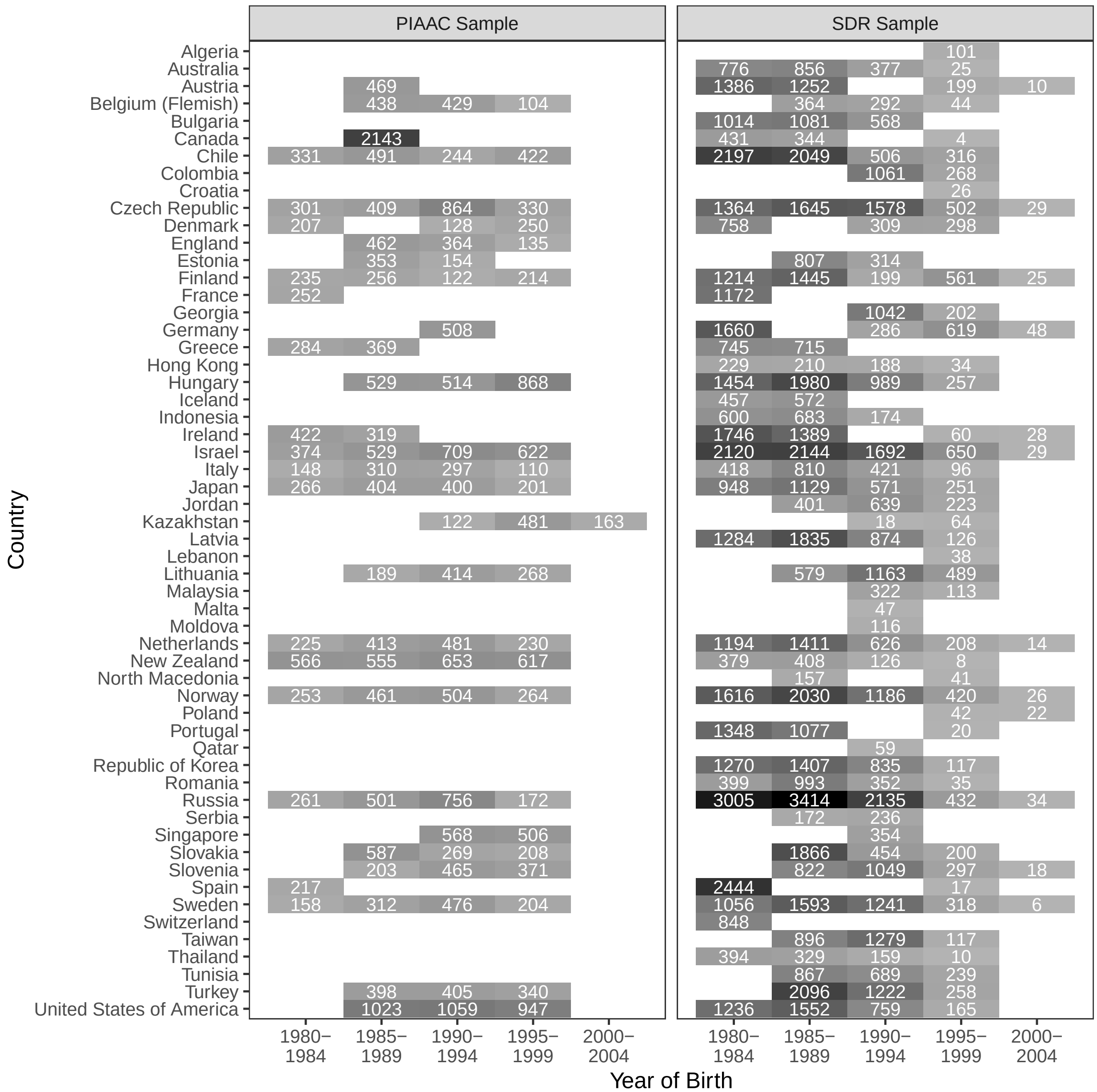



**Figure B.1:** Data Coverage by Country and Year of Birth

**Note:** Figure displays the number of observations in each country-by-year of birth cell, separately for the PIAAC and SDR samples.

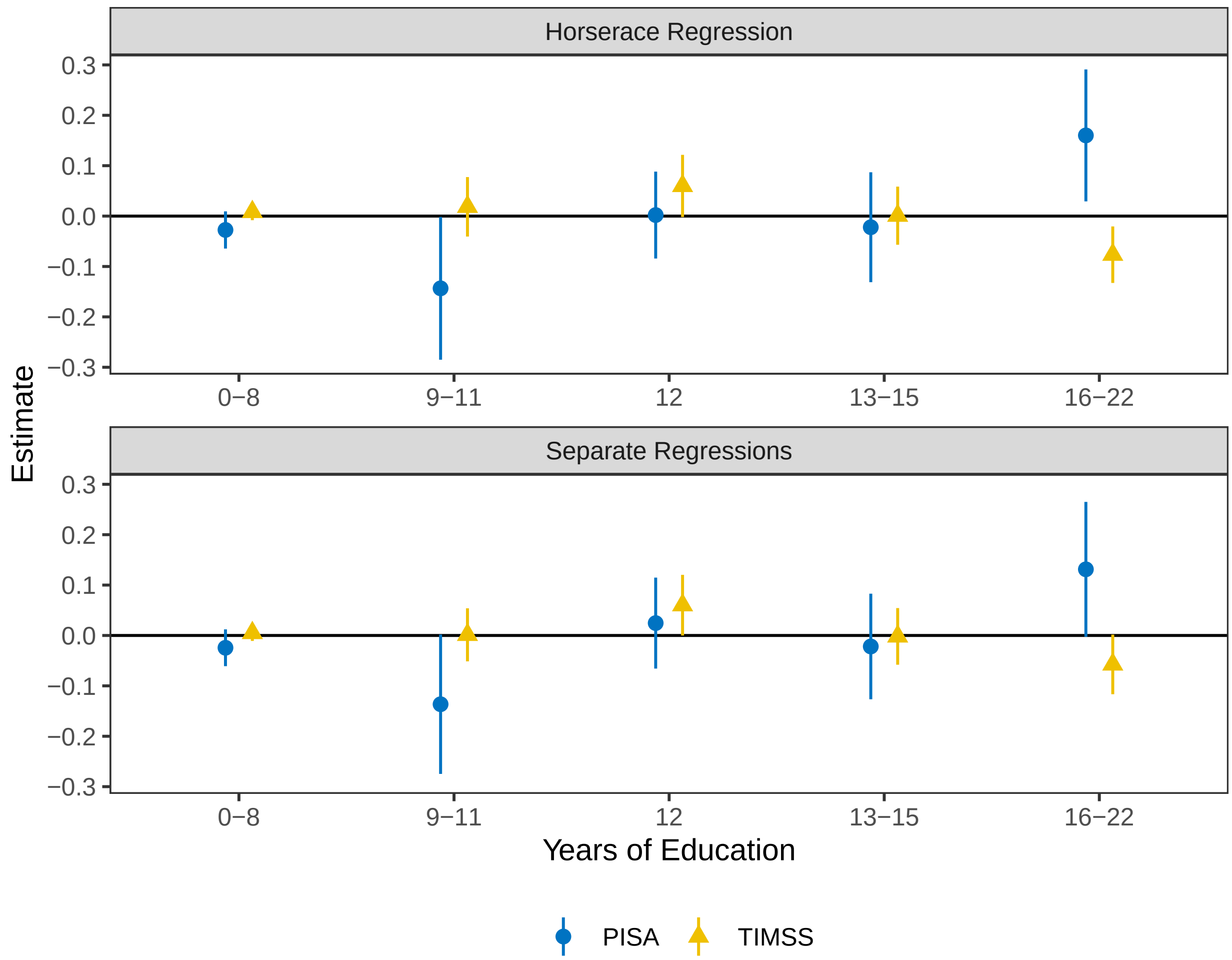



**Figure B.2:** Relationship Between Educational Levels and Adolescent Test Scores by Country Birth Cohort

**Note:** Figure displays regression results estimating the relationship between average PISA and TIMSS math scores and binary indicators for education levels in SDR data. Plotted coefficients reflect the effect of a 1 standard deviation in a cohort's test score on the probability of a respondent having the indicated level of education. PISA and TIMSS represent average PISA and TIMSS scores at the country-by-birth cohort level, transformed to reflect student-level standard deviations. The top panel summarizes the results of 5 separate regressions; each regression includes PISA and TIMSS as independent variables. The bottom panel summarizes the results of 10 separate regressions; each regression includes either PISA or TIMSS as an independent variable. All regressions include fixed effects for age, survey year, survey wave, region-by-age, and country-by-survey year. Region refers to World Bank regions: North America, Latin America and Caribbean, Middle East and North Africa, East Asia and Pacific, and Europe and Central Asia. Error bars indicate 95% confidence intervals. Observations are weighted by SDR-provided weights.

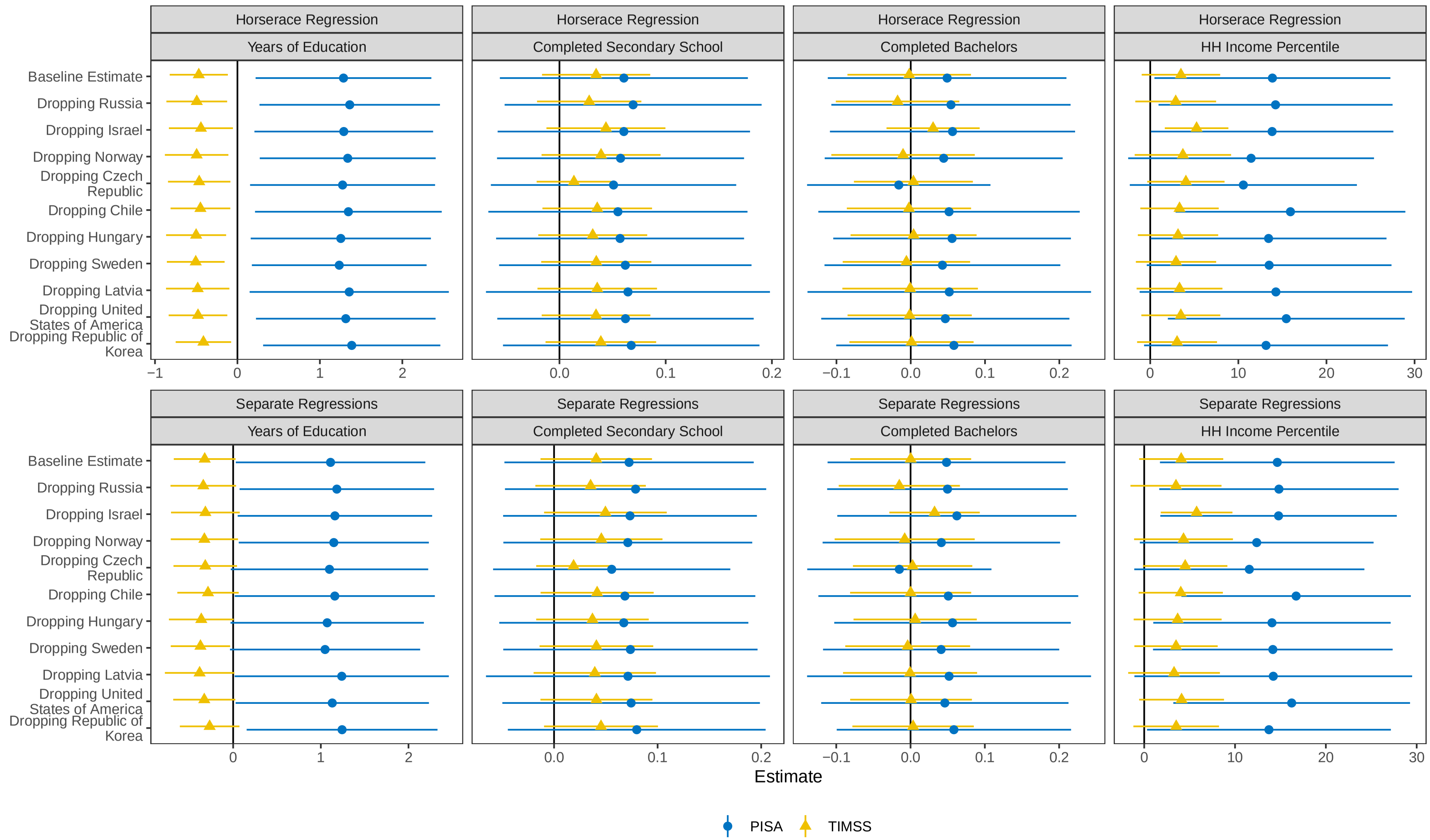



**Figure B.3:** Relationship Between Adulthood Outcomes and Adolescent Test Scores by Country Birth Cohort: Sensitivity to Country Omission

**Note:** Figure displays regression results estimating the relationship between average PISA and TIMSS math scores and adulthood outcomes in SDR data. In each panel, "Baseline Estimate" reflects the estimate shown in Table 3 Columns 7 and 8 (for separate regressions) 9 (for horserace regressions). All other rows show equivalent coefficients after dropping one country from the data. PISA and TIMSS represent average PISA and TIMSS scores at the country-by-birth cohort level, transformed to reflect student-level standard deviations. The top panel summarizes the results of 5 separate regressions; each regression includes PISA and TIMSS as independent variables. The bottom panel summarizes the results of 10 separate regressions; each regression includes either PISA or TIMSS as an independent variable. All regressions include fixed effects for age, survey year, survey wave, region-by-age, and country-by-survey year. Region refers to World Bank regions: North America, Latin America and Caribbean, Middle East and North Africa, East Asia and Pacific, and Europe and Central Asia. Error bars indicate 95% confidence intervals. Observations are weighted by SDR-provided weights.

**Table B.1:** Secular Test Score Growth in PISA versus TIMSS

| | (1) | (2) | (3) |
|---|---|---|---|
| Dependent Variable: PISA or TIMSS Test Score | | | |
| Year | -0.009* | 0.000 | |
| | (0.004) | (0.002) | |
| Year × Test = TIMSS (4th Grade) | 0.023*** | 0.024*** | 0.025*** |
| | (0.006) | (0.003) | (0.003) |
| Year × Test = TIMSS (8th Grade) | 0.010+ | 0.010*** | 0.013*** |
| | (0.005) | (0.003) | (0.002) |
| Num.Obs. | 6089931 | 6089931 | 6089931 |
| R2 | 0.011 | 0.329 | 0.341 |
| Test FEs | ✓ | ✓ | ✓ |
| Country FEs | | ✓ | ✓ |
| Country Time Trend | | | ✓ |

**Note:** Table displays regression results estimating differences in over-time growth in PISA, TIMSS 4th-Grade, and TIMSS 8th-Grade math scores. Data is stacked student-level test results from PISA, 4th-Grade TIMSS, and 8th-Grade TIMSS math assessments. Robust standard errors clustered at the country level and adjusted for multiple imputation using Rubin's rule in parentheses. Observations are weighted by $w_{ict} / \sum_{i \in ct} w_{ict}$, where $w_{ict}$ is individual $i$ in country $c$ for test $t$'s sampling weight, and $\sum_{i \in c} w_{ict}$ denotes the sum of sampling weights in country $c$ and test $t$. Here, test $t$ refers to a test (e.g. PISA, 4th-Grade TIMSS, and 8th-Grade TIMSS) and year (e.g. 1999, 2000, etc) combination. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

**Table B.2:** Sensitivity of Income Estimates to Controls for Household Composition

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Panel A: Main Income Estimates | | | | | | | | | |
| PISA | 12.537* | | 12.522* | 13.029* | | 12.671* | 14.653* | | 13.866* |
| | (5.360) | | (5.418) | (5.004) | | (5.172) | (6.462) | | (6.697) |
| TIMSS | | 4.058+ | 4.048+ | | 3.935+ | 3.673 | | 4.072+ | 3.483 |
| | | (2.189) | (2.192) | | (2.286) | (2.295) | | (2.315) | (2.227) |
| Num.Obs. | 33707 | 33707 | 33707 | 33707 | 33707 | 33707 | 33707 | 33707 | 33707 |
| R2 | 0.084 | 0.084 | 0.084 | 0.086 | 0.086 | 0.086 | 0.103 | 0.103 | 0.103 |
| p-value: PISA = TIMSS | - | 0.182 | 0.184 | - | 0.131 | 0.153 | - | 0.117 | 0.159 |
| Panel B: Main Income Estimates with Household Size Fixed Effects | | | | | | | | | |
| PISA | 10.434+ | | 10.188+ | 11.236* | | 10.738* | 13.700* | | 12.563+ |
| | (5.349) | | (5.516) | (5.043) | | (5.312) | (6.221) | | (6.532) |
| TIMSS | | 5.494* | 5.388* | | 5.247* | 5.010* | | 5.189* | 4.593* |
| | | (2.170) | (2.140) | | (2.343) | (2.301) | | (2.224) | (2.018) |
| Num.Obs. | 29112 | 29112 | 29112 | 29112 | 29112 | 29112 | 29112 | 29112 | 29112 |
| R2 | 0.178 | 0.179 | 0.179 | 0.181 | 0.181 | 0.181 | 0.197 | 0.197 | 0.197 |
| p-value: PISA = TIMSS | - | 0.437 | 0.465 | - | 0.335 | 0.385 | - | 0.192 | 0.275 |
| Age FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Survey Year FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Survey Wave 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 percentile of household income in SDR data. Panel A does not include any controls for household composition. Panel B includes fixed effects for household size. In the table, PISA and TIMSS represent average PISA and TIMSS scores at the country-by-birth cohort level, transformed to reflect student-level standard deviations. Region refers to World Bank regions: North America, Latin America and Caribbean, Middle East and North Africa, East Asia and Pacific, and Europe and Central Asia. Robust standard errors clustered at the country 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. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

**Table B.3:** Relationship Between Adulthood Outcomes and Adolescent Test Scores by Country Birth Cohort (Alternative Weights)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Panel A: Years of Education; Sample: 16+ Years Old | | | | | | | | | |
| PISA | 0.803* | | 0.846* | 0.577 | | 0.641+ | 0.882+ | | 1.036* |
| | (0.358) | | (0.355) | (0.372) | | (0.375) | (0.486) | | (0.458) |
| TIMSS | | -0.173 | -0.239 | | -0.219 | -0.278 | | -0.367+ | -0.477* |
| | | (0.192) | (0.199) | | (0.189) | (0.189) | | (0.214) | (0.203) |
| Num.Obs. | 100475 | 100475 | 100475 | 100475 | 100475 | 100475 | 100475 | 100475 | 100475 |
| R2 | 0.285 | 0.285 | 0.286 | 0.289 | 0.289 | 0.289 | 0.313 | 0.313 | 0.313 |
| p-value: PISA = TIMSS | - | 0.011 | 0.011 | - | 0.042 | 0.038 | - | 0.009 | 0.003 |
| Panel B: Complete Secondary School; Sample: 19+ Years Old | | | | | | | | | |
| PISA | 0.051 | | 0.047 | 0.039 | | 0.030 | 0.040 | | 0.035 |
| | (0.053) | | (0.053) | (0.052) | | (0.051) | (0.047) | | (0.046) |
| TIMSS | | 0.038 | 0.036 | | 0.042+ | 0.040 | | 0.021 | 0.017 |
| | | (0.027) | (0.028) | | (0.025) | (0.025) | | (0.023) | (0.022) |
| Num.Obs. | 97434 | 97434 | 97434 | 97434 | 97434 | 97434 | 97434 | 97434 | 97434 |
| R2 | 0.102 | 0.102 | 0.102 | 0.106 | 0.106 | 0.106 | 0.140 | 0.140 | 0.141 |
| p-value: PISA = TIMSS | - | 0.805 | 0.850 | - | 0.947 | 0.863 | - | 0.650 | 0.723 |
| Panel C: Complete Bachelors; Sample: 24+ Years Old | | | | | | | | | |
| PISA | 0.027 | | 0.027 | 0.027 | | 0.028 | 0.039 | | 0.038 |
| | (0.080) | | (0.080) | (0.064) | | (0.064) | (0.084) | | (0.084) |
| TIMSS | | 0.006 | 0.006 | | -0.006 | -0.007 | | 0.005 | 0.004 |
| | | (0.033) | (0.033) | | (0.037) | (0.037) | | (0.037) | (0.038) |
| Num.Obs. | 45808 | 45808 | 45808 | 45808 | 45808 | 45808 | 45808 | 45808 | 45808 |
| R2 | 0.085 | 0.085 | 0.085 | 0.087 | 0.087 | 0.087 | 0.113 | 0.113 | 0.113 |
| p-value: PISA = TIMSS | - | 0.802 | 0.803 | - | 0.644 | 0.645 | - | 0.697 | 0.711 |
| Panel D: Percentile of Household Income; Sample: 24+ Years Old | | | | | | | | | |
| PISA | 9.058 | | 9.157 | 10.126* | | 9.986* | 12.103+ | | 11.636+ |
| | (5.482) | | (5.586) | (4.721) | | (4.834) | (6.097) | | (6.145) |
| TIMSS | | 3.197 | 3.259 | | 2.845 | 2.734 | | 3.548 | 3.211 |
| | | (2.260) | (2.275) | | (2.180) | (2.169) | | (2.271) | (2.096) |
| Num.Obs. | 33707 | 33707 | 33707 | 33707 | 33707 | 33707 | 33707 | 33707 | 33707 |
| R2 | 0.085 | 0.085 | 0.086 | 0.088 | 0.087 | 0.088 | 0.107 | 0.107 | 0.107 |
| p-value: PISA = TIMSS | - | 0.368 | 0.362 | - | 0.196 | 0.207 | - | 0.169 | 0.199 |
| Age FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Survey Year FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Survey Wave 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 adulthood outcomes in SDR data. Panel A does not include any controls for household composition. Panel B includes fixed effects for household size. In the table, PISA and TIMSS represent average PISA and TIMSS scores at the country-by-birth cohort level, transformed to reflect student-level standard deviations. Region refers to World Bank regions: North America, Latin America and Caribbean, Middle East and North Africa, East Asia and Pacific, and Europe and Central Asia. Robust standard errors clustered at the country level in parentheses. Observations are weighted by SDR-provided weights, rescaled such that the sum of each country-cohort combination's weights equals 1. 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. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

**Table B.4:** Test Scores and Income Nonresponse

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Dependent Variable: Income Nonresponse | | | | | | | | | |
| PISA | -0.006 | | -0.006 | 0.013 | | 0.013 | -0.071 | | -0.071 |
| | (0.093) | | (0.093) | (0.086) | | (0.086) | (0.043) | | (0.044) |
| TIMSS | | -0.007 | -0.007 | | 0.002 | 0.002 | | -0.005 | -0.001 |
| | | (0.032) | (0.032) | | (0.031) | (0.031) | | (0.032) | (0.033) |
| Num.Obs. | 45808 | 45808 | 45808 | 45808 | 45808 | 45808 | 45808 | 45808 | 45808 |
| R2 | 0.233 | 0.233 | 0.233 | 0.235 | 0.235 | 0.235 | 0.289 | 0.289 | 0.289 |
| Age FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Survey Year FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Survey Wave 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 the likelihood of income nonresponse in SDR data. In the table, PISA and TIMSS represent average PISA and TIMSS scores at the country-by-birth cohort level, transformed to reflect student-level standard deviations. Region refers to World Bank regions: North America, Latin America and Caribbean, Middle East and North Africa, East Asia and Pacific, and Europe and Central Asia. Robust standard errors clustered at the country level in parentheses. Observations are weighted by SDR-provided weights. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

**Table B.5:** Relationship Between Adulthood Outcomes and Adolescent Test Scores by Country Birth Cohort: PIAAC Sample

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Panel A: Years of Education; Sample: 16+ Years Old | | | | | | | | | |
| PISA | 0.903 | | 0.818 | 0.955 | | 0.926 | 0.955 | | 0.926 |
| | (0.841) | | (0.821) | (1.438) | | (1.399) | (1.439) | | (1.399) |
| TIMSS | | 0.876* | 0.825+ | | 0.890 | 0.877 | | 0.890 | 0.877 |
| | | (0.421) | (0.405) | | (0.579) | (0.565) | | (0.579) | (0.565) |
| Num.Obs. | 34590 | 34590 | 34590 | 34590 | 34590 | 34590 | 34590 | 34590 | 34590 |
| R2 | 0.385 | 0.385 | 0.386 | 0.388 | 0.388 | 0.389 | 0.388 | 0.388 | 0.389 |
| p-value: PISA = TIMSS | - | 0.974 | 0.994 | - | 0.961 | 0.970 | - | 0.961 | 0.970 |
| Panel B: Complete Secondary School; Sample: 19+ Years Old | | | | | | | | | |
| PISA | 0.175* | | 0.180* | 0.177 | | 0.181 | 0.177 | | 0.181 |
| | (0.071) | | (0.071) | (0.119) | | (0.121) | (0.119) | | (0.121) |
| TIMSS | | 0.025 | 0.037 | | 0.011 | 0.025 | | 0.011 | 0.025 |
| | | (0.051) | (0.050) | | (0.062) | (0.067) | | (0.062) | (0.067) |
| Num.Obs. | 27310 | 27310 | 27310 | 27310 | 27310 | 27310 | 27310 | 27310 | 27310 |
| R2 | 0.371 | 0.370 | 0.371 | 0.372 | 0.372 | 0.372 | 0.372 | 0.372 | 0.372 |
| p-value: PISA = TIMSS | - | 0.062 | 0.059 | - | 0.209 | 0.200 | - | 0.209 | 0.200 |
| Panel C: log(Hourly Earnings, PPP Adjusted); Sample: 19+ Years Old | | | | | | | | | |
| PISA | 0.119 | | 0.164 | 0.238* | | 0.250* | 0.238* | | 0.250* |
| | (0.251) | | (0.205) | (0.109) | | (0.115) | (0.110) | | (0.115) |
| TIMSS | | 0.258 | 0.271 | | 0.015 | 0.039 | | 0.015 | 0.039 |
| | | (0.181) | (0.178) | | (0.039) | (0.041) | | (0.039) | (0.041) |
| Num.Obs. | 10205 | 10205 | 10205 | 10205 | 10205 | 10205 | 10205 | 10205 | 10205 |
| R2 | 0.381 | 0.382 | 0.383 | 0.393 | 0.393 | 0.393 | 0.393 | 0.393 | 0.393 |
| p-value: PISA = TIMSS | - | 0.697 | 0.721 | - | 0.074 | 0.051 | - | 0.074 | 0.051 |
| Age FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Survey Year FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Survey Wave 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 adulthood outcomes in PIAAC data. In the table, PISA and TIMSS represent average PISA and TIMSS scores at the country-by-birth cohort level, transformed to reflect student-level standard deviations. Region refers to World Bank regions: North America, Latin America and Caribbean, Middle East and North Africa, East Asia and Pacific, and Europe and Central Asia. Robust standard errors clustered at the country level in parentheses. Observations are weighted by SDR-provided weights, rescaled such that the sum of each country-cohort combination's weights equals 1. 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. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.