PISA AND LABOR PRODUCTIVITY: IS PISA AN ACCURATE MEASURE OF THE FUTURE ECONOMIC CAPACITY OF STUDENTS?

A Thesis
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Master of Public Policy

By

Patrick C. Sims, B.A.

Washington, DC
April 16, 2013
PISA AND LABOR PRODUCTIVITY: IS PISA AN ACCURATE MEASURE OF THE FUTURE ECONOMIC CAPACITY OF STUDENTS?

Patrick C. Sims, B.A.

Thesis Advisor: Robert Bednarzik, Ph.D.

ABSTRACT

This paper is an econometric analysis of the relationship between education and labor productivity. While economists agree that human capital is an important factor in determining productivity, it has been difficult to measure. Traditionally, the number of years of schooling has been used to measure education, but this fails to account for the quality of that education. The Organization for Economic Cooperation and Development (OECD) implemented the Programme for International Student Assessment (PISA) in 2000 as a cross-country test of cognitive ability. This paper seeks to determine if PISA may be a more useful measure of skill level than the number of years of education. The regression analysis results indicate that PISA is, indeed, a useful metric for identifying human capital. It also raises important questions about academic achievement. According to its findings, the PISA scores of the highest achieving students appear to be most closely associated with higher productivity. A somewhat harder to explain finding was that wider gaps between high and low PISA groups (i.e. higher inequality) had a positive association with productivity. This presents interesting questions for how policymakers should approach education, particularly with respect to equality.
The research and writing of this thesis is dedicated to Professor Bednarzik for his invaluable insight and support, Madzia, my family, and everyone who helped along the way.

Many thanks,
Patrick C. Sims
# Table of Contents

Introduction

A Review of the Literature
   - Table 1: Correlation of GDP per worker and GDP per hour

Policy Implications

Data and Methodology
   - Figure 1: Distribution of Average PISA Scores
   - Figure 2: Map of World Labor Productivity 2009
   - Figure 3: Trends in PISA Scores and Productivity
   - Figure 4: Trends in High vs. Low PISA Scores and Productivity
   - Table 2: Distribution of Average Years of Education for Survey Years 2000-2009
   - Exhibit 1. Definition and predicted relationship of all variables

Analysis
   - Table 3: OLS Regression Results

Policy Recommendations and Conclusion

Appendix A: Database Sources

Appendix B: PISA Participants

Appendix C: Average PISA Scores Across Percentiles for Survey Years 2000-2009

Appendix D: F-Test for OLS vs. Fixed Effects

Appendix E: Chow Test for Poolability

Appendix F: Model Diagnostics

Appendix G: Fixed Effects Regression Results

References
INTRODUCTION

Numerous studies have highlighted the importance of a strong education system in developing a robust economy (Hanushek & Woessmann 2008; Appleton, Atherton, & Bleaney 2008; Altinok 2007; Hanushek & Kimko 2000). It is a key driver of productivity and research has clearly outlined the link between increases in productivity and broader economic growth (Schwedt and Turunen 2008; Lange and Topel 2006; Becker 1993). However, measuring the effectiveness of a country’s educational system can be difficult, especially when attempting to conduct a cross-country analysis. The most common measurements in education, such as grades or national standardized tests, are difficult to compare among countries. Some researchers have used number of years of school as a means of comparison, but this fails to account for quality (Sweetman 1999; Oreopoulos 2006).

In an effort to address these standardization issues, the Organization for Economic Cooperation and Development (OECD) first released the Programme for International Student Assessment (PISA) in 1997. The purpose of the test is not to compare curricula across countries, or to measure how much students have learned. Rather, it “looks at students’ ability to apply knowledge and skills in key subject areas and to analyse, reason and communicate effectively as they examine, interpret and solve problems” (About PISA).

PISA is administered on a triennial basis, with the first test having occurred in 2000. Each administration tests students’ ability in math, science, and/or reading. The test is given to 15 year-olds, which is typically the last year in which schooling is mandatory.
among participant countries. To date, over 70 countries have participated (see Appendix B for map of participants).

Since its first administration, the test has significantly gained prominence. The assessment is commonly used to compare educational outcomes among countries and at least 14 nations with federal systems use PISA to compare educational success across regions (Breakspear 2012). Some advocates have argued that the United States should adopt the test at the state-level (Jerald 2008).

Given its stated goals, high PISA scores should be an estimate of how effectively students are prepared to become a part of the labor market. As PISA continues to grow in popularity and relevance, it is important to determine if it is successfully measuring what it is intended to be evaluating. This paper will use average PISA scores from the years 2000, 2003, 2006, and 2009 in 65 economies to determine if they are positively associated with labor productivity. By identifying a positive correlation, we can feel confident that PISA is, in fact, a useful tool for identifying which country’s educational system best prepares its students to be effective members of the workforce. After all, well-educated students become skilled workers who, when successful in the labor market, enhance output.
A Review of the Literature

*Human Capital and Labor Productivity*

Modern growth theory places particular importance on human capital as a means of generating economic growth through increases in labor productivity (Barro & Sala-i-Martin 2004). Many researchers have argued that human capital is at least one of the most important sources for long-term increases in economic welfare (Schwedt and Turunen 2008; Lange and Topel 2006; Becker 1993). Jorgenson, Ho, and Stiroh (2005) identified a positive association between the growth of U.S. workers with college degrees and productivity. Dal Bianco (2010) found similar gains from tertiary education. His analysis suggested that rapid technology diffusion leads to a convergence of productivity levels among advanced economies, but that the absorptive power of a highly educated workforce can create differentiation in productivity levels.

Dal Bianco is not alone in identifying the need to develop education to keep up with the demands of an increasingly high-tech economy. Goldin and Katz (2008) argued that without high educational achievement, technological progression would create greater inequality between those with a good education and those without one. In their estimation, students do not necessarily need more years of schooling, but need to acquire the skills that allow them to be more analytical and flexible. Coulombe et al (2004) also concluded that raising human capital through cognitive ability was effective for improving macroeconomic outcomes, with an emphasis on closing achievement gaps and decreasing the proportion of the least educated. Other research has found that the primary source for human capital accumulation is through formal education (Becker 1993).
Cumulatively, these studies provide a solid grounding to assume that human
capital can have a positive effect on labor productivity. Given that these relationships are
identified as being dependent on improvements in the quality of human capital, it is
necessary to use a measure of labor productivity that would be most representative of the
quality of labor inputs, not quantity.

Baldassarini and Di Veroli (2008) described various measures of productivity and
concluded that GDP per total hours worked (output per hour) is the best measure. From a
qualitative perspective, this is a logical conclusion. We are concerned with how
effectively workers use each hour rather than how much they produce on aggregate,
regardless of the number of hours worked. Assuming well-educated workers can more
efficiently use one hour of work time, we can expect that GDP per hour worked will
disaggregate the effect of a less educated workforce working more hours to compensate
for their lower human capital.

While GDP per hour worked is available for OECD countries, this measurement
is not available for all economies that have participated in PISA. One alternative measure
is GDP per person employed, which includes both full-time and part-time workers. Haine
and Kanutin (2008) found that in the Eurozone, GDP per person employed and GDP per
hour worked are closely correlated when compared between the years 1995 and 2005.
Occasional variation does occur between quarters, but the authors ascribe this to the need
for the employment indicator to become a monthly indicator. Regardless, they conclude
that these measures are valid substitutes. Table 1 illustrates the strong correlation that
exists between these two measures within this dataset.
Researchers have also found that selecting a measurement for human capital can be quite challenging, particularly for cross-country comparisons (Sianesi 2000). PISA has become an especially useful tool for this purpose due to its standardized nature. However, limited research has been conducted with it, as the data are relatively new. Other international standardized educational assessments have existed since the 1960s, but none of them was administered consistently enough that they could be accurately compared over time (see Table 2.1, Chapter 2 Hanushek, Machin, & Woessmann). Since PISA was instituted in 2000, it has used a standardized scoring system, which makes it possible to analyze as panel data.

<table>
<thead>
<tr>
<th></th>
<th>GDP per worker</th>
<th>GDP per hour worked</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP per worker</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>GDP per hour worked</td>
<td>0.9445</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
POLICY IMPLICATIONS

Why do policymakers care whether there is a positive relationship between PISA scores and labor productivity? The first, and most obvious, implication is that the OECD can use these findings to determine if PISA is successfully testing what it was designed to examine – future success in the job market. If there is a positive relationship, this would provide justification to continue allocating resources to the administration of this assessment. If no positive relationship is found, then the OECD may want to examine PISA and recalibrate it to better fulfill its stated goal.

Secondly, there is a growing movement in the United States to invest resources to administer PISA on the state level. The National Governors Association, the Council of Chief State School Officers, and Achieve, Inc. have all voiced their support for this effort (Jerald 2008). Over 100 schools in 20 U.S. states are currently piloting a program that will allow schools to compare school-level PISA scores locally and internationally (Robelen 2012). Proponents claim that this will help pinpoint underachieving schools. However, some critics argue that PISA is flawed due its inability to identify at which point the education system is failing or succeeding. Further, they contend that using PISA as an assessment of local education would result in the OECD having undue influence in local decisions (Schneider 2009). Determining PISA’s ability, or inability, to identify successful education systems would help to reconcile these conflicting opinions.

Finally, many critics worry about the rising influence of standardized testing. Zhao has argued that American education is in danger of losing its unique ability to educate creative thinkers who possess the necessary flexibility to adapt to a fluid labor market (2009). Along with Goldin and Katz, he notes that while countries such as China
shift their education systems to resemble that creativity of the American system, the
United States is moving to a system too heavily focused on test scores (2008). Theory
suggests that creativity and flexibility in education bolster productivity levels. PISA is
designed to test for the ability of students to apply what they have learned, not test how
much they have learned. A positive relationship between productivity and PISA scores
would indicate that it is a useful tool, despite the reservations of some critics.
DATA AND METHODOLOGY

Data

Data have been collected from several sources to create a set of panel data. The key independent variable, PISA scores, is taken as the average of all scores (math, science, and reading) for each participating country in the years 2000, 2003, 2006, and 2009. Most studies using international standardized test scores have adopted this method and used the average of all test scores (Hanushek & Woessmann 2009; Appleton, et. al. 2008; Altinok 2007).

In all, 66 countries are represented in the data set, but not all countries participated in PISA in all four years. The data were taken from the National Center for Education Statistics (NCES) and there are a total of 202 observations. Figure 1 shows the distribution of average PISA scores by country-year. The highest concentration of observations is around 500, which is the average score for OECD countries. The asymmetrical distribution of more countries below average is due to the additional non-OECD countries in the sample, which tended to have lower scores on average.
The dependent variable, productivity, is measured as GDP per person employed in the economy and the data are from the International Labour Organization’s Key Indicators of the Labour Market (ILO KILM). As previously noted, GDP per hour worked would be the preferred measure since it better assesses the quality of labor output, not just quantity. However, these data are missing for some countries and previous research has found that the two productivity measures are closely correlated, which supports the use of GDP per person employed (Haine and Kanutin 2008). Figure 2 presents a map, which illustrates GDP per person employed in 2009 by country. Clearly, OECD countries tend to have higher productivity levels than non-OECD countries.
Figure 3 suggests that a positive relationship could exist between average PISA scores and GDP per person employed. Figure 4 represents the same relationship for two discrete groups of countries: one with above-average PISA scores and the other with below-average scores. Indeed, the chart does seem to indicate that a positive relationship could exist for both low- and high-performers. However, more sophisticated analysis is necessary to confirm or refute the apparent positive relationship. An econometric model using regression analysis can help sort out the actual relationship between PISA and productivity.
Figure 3: Trends in PISA Scores and Productivity

Trends in PISA Scores and Productivity for Survey Years 2000-2009

Data Sources: NCES International Data Explorer and ILO KILM n: 202

Figure 4: Trends in High vs. Low PISA Scores and Productivity

Trends in High vs. Low PISA Scores and Productivity for Survey Years 2000-2009

Data Sources: NCES International Data Explorer and ILO KILM n: 202
The regression model will include, or control for, other influences on productivity, such as average years of education, total investment, trade, and economic freedom. Until recently, most research conducted on this subject used quantity to identify education’s contribution to human capital. Measures of cognitive ability, such as PISA, have advanced the study of the link between education and economic outcomes, but it is likely that quantity is still a factor (Hanushek and Kimko 2000). Average number of years of education in each country for each year is included to control for this relationship. As seen in Table 2, students in the majority of observations have gone to school, on average, between 9 and 12 years. This implies that students in most of the countries receive at least some secondary education and the average student does not receive tertiary education.

**Table 2: Distribution of Average Years of Education for Survey Years 2000-2009**

<table>
<thead>
<tr>
<th>Distribution of Average Years of Education (in years)</th>
<th>Percent of countries</th>
</tr>
</thead>
<tbody>
<tr>
<td># years</td>
<td></td>
</tr>
<tr>
<td>&lt;5</td>
<td>0.52</td>
</tr>
<tr>
<td>5-6</td>
<td>5.20</td>
</tr>
<tr>
<td>6-7</td>
<td>4.68</td>
</tr>
<tr>
<td>7-8</td>
<td>7.80</td>
</tr>
<tr>
<td>8-9</td>
<td>9.88</td>
</tr>
<tr>
<td>9-10</td>
<td>22.88</td>
</tr>
<tr>
<td>10-11</td>
<td>18.20</td>
</tr>
<tr>
<td>11-12</td>
<td>19.76</td>
</tr>
<tr>
<td>12-13</td>
<td>7.80</td>
</tr>
<tr>
<td>13-14</td>
<td>2.60</td>
</tr>
<tr>
<td>Total</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Source: Barro and Lee n: 192

Besides education, other important factors that contribute to economic outcomes have been included. Total investment as a percentage of GDP is included for each country in each testing year. This variable helps to account for variation in productivity
levels that may be explained by fluctuations in capital expenditures (Mankiw, Romer, and Weil 1992). In a paper for the OECD, Englander and Gurney found that trade is an important determinant of productivity for at least two reasons (1994). Trade creates competition, which spurs innovation, and it also leads to a faster diffusion of technology among trading partners. Total trade as a percentage of GDP from the ILO KILM has been included to control for these effects.

Additionally, an index of economic freedom has been included to control for the extent of capitalism (Altinok 2007). Hall, Russell, and Crowley (2010) found that economic freedom is necessary to allow increases in human and physical capital to have an impact on economic outcomes. For this analysis, the annual Economic Freedom of the World report is used (Gwartney, et. al. 2009). The index has a variety of components: size of government; legal system and property rights; monetary policy; free trade policies and labor and market regulations. As a result, these are all implicitly controlled for here. The index is scaled from zero to ten, with higher numbers indicating a more free economy. The average for the sample is 7.10. Over the four years in this analysis, Colombia had the lowest average (5.69) and Hong Kong had the highest (8.87).
Methodology

The Ordinary Least Squares (OLS) Model is specified as follows:

\[ Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_4 X_{4it} + \beta_5 X_{5it} + \beta_6 X_{6it} + u_{it} \]

Where:

- \( Y_{it} = \text{gdpperworker} \)
  GDP per person engaged (constant 1990 USD at PPP).
- \( X_{1it} = \text{pisaaverage} \)
  Average PISA score by year and country.
- \( X_{2it} = \text{lnaveed} \)
  Average number of years of education received in a country by year (logged).
- \( X_{3it} = \text{econfree} \)
  Index of economic freedom where 0 = closed and 10 = open, by year and country.
- \( X_{4it} = \text{highincome} \)
  Dummy variable indicating whether a country is categorized as “high income” in a given year.
- \( X_{5it} = \text{trade} \)
  The total amount of goods and services traded by a country in a given year.
- \( X_{6it} = \text{Intotalinvestment} \)
  The total amount invested in the economy as a percentage of GDP by year and country (logged).

An F-test was conducted to determine if the data should be analyzed using a fixed effects or an OLS without dummies model (see Appendix D for more detail). In order to perform this test, OLS regressions are run with and without time and country dummies. A calculation is then made, which indicated that a fixed effects model would be best. However, in order to construct a fixed effects model, the data must be poolable.

To test for poolability, a Chow test was conducted, which tests to see if the slopes are the same across countries and over time. In order for a fixed effects model to be accurate, this must be the case. Appendix E shows the formula for conducting a Chow
test, which requires the analyst to sum the Residual Sum of Squares (SSE) for each country. However, given the limited number of observations per country (i.e. only four time periods), an individual regression for each country is unable to provide an SSE. Therefore, the Chow test cannot be performed on this dataset and poolability cannot be determined.

Given that a Chow test could not be performed, it would be a weak assumption to assume that fixed effects should be used when there is not even sufficient data to perform the necessary tests. Clark and Linzer (2013) also warn against the limitations of fixed effects when the number of observations per unit are low, which is precisely what is preventing the Chow test from being performed. Though poolability could not be determined, Appendix G provides the regression results when using fixed effects. None of the results is statistically significant, except for economic freedom.

If the data are not poolable, then the countries in the dataset do not all have the same slopes, which is likely caused by the different sensitivities of what drives productivity across countries. This implies that a fixed effects model is not appropriate for the data. However, it is likely that the sensitivity of important influences on productivity is similar among high-income countries. So, a dummy indicating if a country is a high-income or low-income country has been included in the regression.

Exhibit 1 provides a list of the variables included in this analysis. The expected sign, or predicted relationship with the dependent variable, is given for each independent variable. All variables are predicted to have a positive relationship with labor productivity, except for total investment. If PISA scores are predictive of students’ ability to join the workforce, then higher PISA scores should be associated with higher
productivity (Hanushek and Woessman 2009). Previous research has found that increases in the number of years of education, trade, and economic freedom are all positively associated with higher productivity, while total investment as a percentage of GDP is expected to be negatively associated (see justifications in Exhibit 1 below for references).

**Exhibit 1. Definition and predicted relationship of all variables**

<table>
<thead>
<tr>
<th>Definition</th>
<th>Variable Name</th>
<th>Expected Sign</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_{it}$</td>
<td>GDP per person engaged (constant 1990 USD at PPP) by country and year.</td>
<td>gdpperworker</td>
<td>N/A</td>
</tr>
<tr>
<td>Independent Variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_{1it}$</td>
<td>Average PISA score by country and year.</td>
<td>pisaaverage</td>
<td>Positive</td>
</tr>
<tr>
<td>$X_{2i}$</td>
<td>Average years of education by country and year (logged).</td>
<td>inaveed</td>
<td>Positive</td>
</tr>
<tr>
<td>$X_{4it}$</td>
<td>Economic Freedom by country and year: index of economic freedom where 0 = closed and 10 = open.</td>
<td>econfree</td>
<td>Positive</td>
</tr>
<tr>
<td>$X_{5}$</td>
<td>High Income dummy: indicates whether a country is categorized as “high income” by the World Bank, by year.</td>
<td>highincome</td>
<td>Positive</td>
</tr>
<tr>
<td>$X_{6}$</td>
<td>Total Trade: the total amount of goods and services traded by a country in a given year.</td>
<td>trade</td>
<td>Positive</td>
</tr>
<tr>
<td>$X_{7}$</td>
<td>The total amount invested in the economy as a percentage of GDP.</td>
<td>totalinvestment</td>
<td>Negative</td>
</tr>
</tbody>
</table>
**Analysis**

**Summary of Regressions**

In all, five regressions were run. Different education variables were included to identify their unique relationships with labor productivity. The first regression (1) included the average number of years of education in a country as the key variable of interest, which is traditionally how economists have measured the education portion of human capital. Regression (2) instead used average PISA scores as the key independent variable of interest to see if they were more closely associated with productivity than the traditional education level measure. Regression (3) used the average scores for students in the 90th percentile to see if the magnitude of the highest achievers’ scores is associated with productivity. A measure for inequality in educational achievement (4) was used to see if greater disparities in PISA scores were associated with lower labor productivity. Finally, this inequality measure was regressed with average PISA scores (5) to see their significance when regressed together.

The results indicate that PISA is more closely associated with labor productivity than average years of education. Further, the magnitude of scores matters, as it appears that students in the 90th percentile have a stronger association than the average students’ PISA scores. The results for educational inequality were inconclusive, but it does seem to have a positive relationship with productivity. These two results could follow logically. Countries with greater disparity tend to have high achievers, which creates that large gulf in achievement. If high achievers are more strongly associated with productivity, many of those countries with high achievement in the 90th percentile will also have greater disparity. Their relatively high productivity will actually be driven by those in the 90th percentile, not by the high level of disparity itself.
Table 3: OLS Regression Results

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>Mean</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average PISA Score</td>
<td>471.44</td>
<td>57.21***</td>
<td>(3.54)</td>
<td>65.91***</td>
<td>(3.99)</td>
</tr>
<tr>
<td>PISA 90\textsuperscript{th}</td>
<td>588.73</td>
<td>67.66***</td>
<td>(4.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PISA 90\textsuperscript{th}-10\textsuperscript{th} Percentiles</td>
<td>239.19</td>
<td>29.29</td>
<td>(1.12)</td>
<td>54.81**</td>
<td>(2.13)</td>
</tr>
<tr>
<td>Average Years Education (log)</td>
<td>2.26</td>
<td>7,502***</td>
<td>(5.46)</td>
<td>4,330***</td>
<td>(4.57)</td>
</tr>
<tr>
<td>Economic Freedom</td>
<td>7.15</td>
<td>4,306***</td>
<td>(4.61)</td>
<td>5,276***</td>
<td>(5.55)</td>
</tr>
<tr>
<td>High Income</td>
<td>0.59</td>
<td>16,240***</td>
<td>(8.61)</td>
<td>16,728***</td>
<td>(11.50)</td>
</tr>
<tr>
<td>Total Trade</td>
<td>4.5e+11</td>
<td>3.24e-09***</td>
<td>(4.20)</td>
<td>3.16e-09***</td>
<td>(3.93)</td>
</tr>
<tr>
<td>Total Investment (log)</td>
<td>3.09</td>
<td>-6,172**</td>
<td>(-2.33)</td>
<td>-34,184**</td>
<td>(-2.74)</td>
</tr>
<tr>
<td>Constant</td>
<td>-10,599</td>
<td>-29,569***</td>
<td>(-2.67)</td>
<td>-17,798**</td>
<td>(-1.45)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># Countries</th>
<th>Observations</th>
<th>R-squared</th>
<th>F-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>53</td>
<td>53</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>180</td>
<td>180</td>
<td>180</td>
</tr>
<tr>
<td></td>
<td>0.790</td>
<td>0.804</td>
<td>0.792</td>
</tr>
<tr>
<td></td>
<td>130.95</td>
<td>118.45</td>
<td>109.50</td>
</tr>
</tbody>
</table>

T-Statistic in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Regression Results

All five regressions had high R-squared values, indicating that they each explain about 80 percent of the variation in labor productivity. Additionally, they all had highly statistically significant F-Statistics, which provides support to the hypothesis that there is, indeed, a significant relationship in the model. While tests did indicate that heteroskedasticity was present, the inclusion of robust standard errors had no substantive effect on the regression results. Robust standard errors were, therefore, omitted. Appendix F also provides more evidence of the appropriateness of the models.

Specifically, diagnostics were conducted on all 5 regressions, which indicate none of the
models suffers from omitted variable bias or other misspecification problems. Finally, all of the independent variables were tested for multicollinearity, which was not an issue (Appendix F offers greater detail on these diagnostics).

As previous research has found, the first regression indicates that average years of education is positively associated with labor productivity. However, once average PISA scores were included in the model, its statistical significance disappeared. Average PISA scores appear to be a better indicator of the skills acquired through education than just the level of education itself and were highly statistically significant at the 1 percent level. While no other analysis has looked at this relationship using only PISA data, Hanushek and Woessmann (2009) found similar results for other international standardized assessments.

Further evidence of the validity of using PISA as a measure of human capital was found when estimating the regression using the PISA scores of students in the 90th percentile. When included along with the average number of years of education, it had a positive association with productivity and was highly statistically significant at the 1 percent level. Average number of years of education, however, had no significant relationship. It is not possible to simultaneously regress average PISA scores and scores from the 90th percentile due to high collinearity (0.979).

We can, though, infer something about their respective relationships by comparing the results from regressions 2 and 3. Both are highly statistically significant, but scores from the 90th percentile seem to have a slightly stronger relationship. This could support recent research, which shows much of the benefits of recent economic growth being realized with increasing inequality (Berg and Osprey, 2011). Those who are
especially high-achieving are creating more new wealth and, therefore, fostering those types of students, may be more closely associated with higher productivity. Improved technology may also contribute to this since many low- and medium-skilled jobs are being replaced by computers and machinery, thus concentrating productivity gains among the best educated (Galor and Moav 2000).

The fourth regression sought to identify the association between high inequality in academic achievement and labor productivity. To do so, it regressed the disparity between scores from the 90th and 10th percentiles. This variable, however, had no statistically significant relationship, but the average number of years of education was significantly positively associated. However, once we included both average PISA scores and the inequality variable, both had a positive and highly statistically significant association with productivity and average number of years of education lost its significance.

While the explanation for this is not entirely clear, it may be the case that the inequality variable is not meaningful without the control of the PISA average. That is, the level of inequality may have a different effect depending on how high the average PISA score is. For example, a country with a very high average score may be less affected by high inequality. In such a case, its lowest scores would still be relatively high when compared with a country with an equal level of inequality, but a lower average score. Therefore, it is not possible to estimate the effect of inequality without average scores.

Four of the control variables were robustly statistically significant across the regressions: economic freedom; the dummy for high income; total trade; and total investment. All variables were significant at at least the 5 percent level in all five
regressions. As predicted, economic freedom, trade, and the dummy indicating if the country was high income were consistently positively and significantly associated.

Total investment (as a percentage of GDP) was consistently negatively and significantly associated. It is important to remember that this is not simply a measure of total investment, but is calculated as a share of GDP. Countries with high GDPs are likely to have higher total investments, but it will be represented as a lower portion of GDP in these data. The United States, for example, invests much less as a percentage of GDP than Thailand (e.g. 14.7 percent vs. 21.2 percent in 2009). That is, they are starting from a much larger base. The key takeaway is that countries with high labor productivity have a better-educated workforce with higher levels of cognitive ability (as measured by PISA scores in this analysis).
POLICY RECOMMENDATIONS AND CONCLUSION

The findings from the analyses provide interesting insight into how policymakers should approach education. Most obviously, it confirms once again the importance of a strong educational system in developing a robust economy. While this is not a new finding, its importance warrants repeated emphasis. Better educational outcomes have consistently been shown to have a positive association with labor productivity and governments would be wise to invest accordingly. How best to measure education, however, has been less certain.

With PISA, the OECD has sought to make a useful standardized assessment that allows for cross-country comparison. The findings from this paper support the idea that PISA is a better indicator of educational achievement than the number of years of education. This implies that the OECD has developed a useful assessment tool and continued administration of PISA exams is a worthy endeavor. This also indicates that governments should not simply add years of education to improve outcomes, but should seek to enhance the quality of the education system instead.

The regression results indicate that scores from the 90th percentile may be more strongly associated with higher productivity levels than average PISA scores. This has important implications for policymakers. While all students should be supported, it is important that the highest achieving students are given the opportunity to thrive and advance beyond their peers. The results also seem to indicate that high inequality is positively associated with productivity. However, too much should not be read into these findings since it is likely that they are actually driven by the especially high scores of the highest achievers.
Much debate has occurred in the United States about the value of PISA. Some proponents seek to employ it at the school- and state-levels, while others contend that it is a flawed measure. This analysis was conducted at the country-level and, therefore, cannot necessarily be interpreted as a useful indicator at lower levels. It has, however, found that PISA is a useful measure of educational achievement, which should lend some credibility to the claim that it could be useful at lower levels. If higher PISA scores, on average, are associated with higher productivity, it would logically follow that schools and states with higher achievement on PISA are doing something right.

This analysis, however, is not a defense of all standardized tests. Many fear that standardized testing’s influence is growing beyond its actual utility. Most certainly, not all standardized tests are the same. The emphasis of PISA is to measure critical thinking and problem solving skills. Many argue that standardized tests in the United States focus too heavily on rote memorization, which leads to “teaching to the test.” Teachers are focusing on teaching specific questions rather than providing their students with the analytical skills to solve them. There is no reason to assume that these findings can be applied to other tests; it is possible that scores from such tests would not have a similar positive association with labor productivity.
## APPENDIX A: DATABASE SOURCES

<table>
<thead>
<tr>
<th>Variable</th>
<th>Database Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>gdpperworker: GDP per person engaged.</td>
<td>International Labor Organization Key Indicators of the Labour Market</td>
</tr>
<tr>
<td>pisaaverage: average PISA score by country and year.</td>
<td>National Center for Education Statistics (U.S. Department of Education)</td>
</tr>
<tr>
<td>aveed: average years of education by country and year.</td>
<td>Barro and Lee</td>
</tr>
<tr>
<td>highincome: indicates whether a country is categorized as “high income” by the World Bank, by year.</td>
<td>World Bank</td>
</tr>
<tr>
<td>trade: the total amount of goods and services traded by a country in a given year.</td>
<td>World Trade Organization</td>
</tr>
<tr>
<td>totalinvestment: the total amount invested in the economy as a percentage of GDP.</td>
<td>International Labor Organization Key Indicators of the Labour Market</td>
</tr>
</tbody>
</table>
APPENDIX B: PISA PARTICIPANTS

OECD countries
- Australia
- Austria
- Belgium
- Canada
- Czech Republic
- Denmark
- Finland
- France
- Germany
- Greece
- Hungary
- Iceland
- Ireland
- Italy
- Japan
- Korea
- Luxembourg
- Mexico
- Netherlands
- New Zealand
- Norway
- Poland
- Portugal
- Slovak Republic
- Spain
- Sweden
- Switzerland
- Turkey
- United Kingdom
- United States

Partner countries and economies in PISA 2006
- Argentina
- Azerbaijan
- Brazil
- Bulgaria
- Chile
- Colombia
- Croatia
- Estonia
- Hong Kong-China
- Indonesia
- Israel
- Jordan
- Kyrgyzstan
- Latvia
- Liechtenstein
- Lithuania
- Macao-China
- Montenegro
- Qatar
- Romania
- Russian Federation
- Serbia
- Slovenia
- Chinese Taipei
- Thailand
- Tunisia
- Uruguay

Partner countries and economies in previous PISA surveys or in PISA 2009
- Albania
- Shanghai-China
- Dominican Republic
- Macedonia
- Moldova
- Panama
- Peru
- Singapore
- Trinidad and Tobago

Source: OECD Library
APPENDIX C: AVERAGE PISA SCORES ACROSS PERCENTILES FOR SURVEY YEARS 2000-2009
APPENDIX D: F-TEST FOR OLS VS. FIXED EFFECTS

Formula

\[
\frac{(RSS_{OLS} - RSS_{Dummies})}{N + T + 2} \cdot \frac{RSS_{Dummies}}{NT - N - T}
\]

OLS Regression

\[
\text{. reg gdpperworker pisaaverage lnaaveed econfree trade lntotalinvestment}
\]

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2.7613e+10</td>
<td>5</td>
<td>5.5227e+09</td>
</tr>
<tr>
<td>Residual</td>
<td>1.1092e+10</td>
<td>174</td>
<td>63745553.8</td>
</tr>
<tr>
<td>Total</td>
<td>3.8705e+10</td>
<td>179</td>
<td>216228925</td>
</tr>
</tbody>
</table>

OLS Regression with Time and Country Dummies

\[
\text{. reg gdpperworker pisaaverage lnaaveed econfree trade lntotalinvestment i.year i. country1}
\]

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3.8442e+10</td>
<td>62</td>
<td>620025361</td>
</tr>
<tr>
<td>Residual</td>
<td>263405223</td>
<td>117</td>
<td>2251326.69</td>
</tr>
<tr>
<td>Total</td>
<td>3.8705e+10</td>
<td>179</td>
<td>216228925</td>
</tr>
</tbody>
</table>

Calculation

\[
\frac{(1.1092e+10 - 263405223)}{(55 + 4 - 2)} \cdot \frac{(263405223)}{(220 - 66 - 4)} = 108.18427
\]

108.18427 \rightarrow \text{Indicates the need for fixed effects.}
**APPENDIX E: CHOW TEST FOR POOLABILITY**

Formula

\[
\frac{(RSS_{pooled} - RSS_{individualsum})}{RSS_{individualsum}} \times \frac{1}{(n-1)(k+1)}
\]

\[
RSS_{individualsum} = \text{the sum of all Residual Sum of Squares for each individual country.}
\]
The following model diagnostics support the model specifications for each of the five regressions. The Ramsey RESET test is designed to identify if the model suffers from omitted variable bias. If the results are statistically significant, there is likely an omitted variable bias. All five models have insignificant results and therefore, likely do not have an omitted variable bias.

The linktest is performed to analyze how well the model fits the data. A well-fitting model will have a \textit{linktest-hat} that is highly statistically significant and a \textit{linktest-hatsq} that is not. As seen below, all five models meet this criteria, indicating that the models are well-fitted.

Finally, a statistically significant result from the Breusch-Pagan test indicates the presence of heteroskedasticity, which could bias the standard error estimates and, therefore, statistical significance of the independent variables in the regression. While some of the models do have heteroskedasticity, as seen below, it does not appear to be biasing the estimates. The use of the \textit{robust} command in Stata corrects for any bias, but does not change the regression results in any substantive way. Therefore, it was omitted from this analysis.

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|c|c|c|}
\hline
\textbf{Diagnostics Results} & (1) & (2) & (3) & (4) & (5) \\
\hline
\textbf{Ramsey RESET (significance)} & 0.5054 & 0.1700 & 0.1717 & 0.3910 & 0.1396 \\
\textbf{linktest–hat} & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
\textbf{linktest–hatsq} & 0.635 & 0.270 & 0.216 & 0.533 & 0.163 \\
\textbf{Breusch-Pagan test for Heteroskedasticity (significance)} & 0.3236 & 0.0469 & 0.0161 & 0.2387 & 0.0093 \\
\hline
\end{tabular}
\end{table}

A model cannot be specified with independent variables that are highly correlated. As seen below, only two of the variables correlated to a worrisome level (PISA average and PISA 90\textsuperscript{th} percentile scores). For this reason, no regression is run using those two variables together.
## Multiollinearity: Correlations of Independent Variables

<table>
<thead>
<tr>
<th></th>
<th>PISA Ave.</th>
<th>Average Years of Edu (log)</th>
<th>Econ. Free</th>
<th>High Income Country</th>
<th>Trade</th>
<th>Total Invest. (log)</th>
<th>PISA90</th>
<th>PISA9010</th>
</tr>
</thead>
<tbody>
<tr>
<td>PISA Average</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Years of Education (log)</td>
<td>0.5860</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economic Freedom</td>
<td>0.6045</td>
<td>0.4118</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Income Country</td>
<td>0.7163</td>
<td>0.4589</td>
<td>0.6172</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trade</td>
<td>0.2959</td>
<td>0.2865</td>
<td>0.2926</td>
<td>0.3747</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Investment (log)</td>
<td>-0.054</td>
<td>-0.0406</td>
<td>-0.046</td>
<td>-0.1783</td>
<td>-0.236</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PISA90</td>
<td>0.9794</td>
<td>0.6411</td>
<td>0.5987</td>
<td>0.7330</td>
<td>0.3177</td>
<td>-0.1036</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>PISA9010</td>
<td>0.1103</td>
<td>0.3963</td>
<td>0.1079</td>
<td>0.2621</td>
<td>0.1696</td>
<td>-0.2421</td>
<td>0.3025</td>
<td>1.000</td>
</tr>
</tbody>
</table>
### APPENDIX G: FIXED EFFECTS REGRESSION RESULTS

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>Mean</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Productivity (GDP per worker)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average PISA Score</td>
<td>471.44</td>
<td>3.94</td>
<td>(0.26)</td>
<td></td>
<td>3.37</td>
<td>(0.21)</td>
</tr>
<tr>
<td>PISA 90th Percentile</td>
<td>588.73</td>
<td>2.53</td>
<td>(0.17)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PISA 90th-10th Percentiles</td>
<td>239.19</td>
<td></td>
<td></td>
<td>-2.45</td>
<td>(-0.21)</td>
<td>-1.83</td>
</tr>
<tr>
<td>Average Years Education (log)</td>
<td>2.26</td>
<td>2,133</td>
<td>(0.46)</td>
<td>1,965</td>
<td>(0.41)</td>
<td>2,027</td>
</tr>
<tr>
<td>Economic Freedom</td>
<td>7.15</td>
<td>1,227***</td>
<td>(2.84)</td>
<td>1,189**</td>
<td>(2.60)</td>
<td>1,206***</td>
</tr>
<tr>
<td>Total Trade</td>
<td>4.5e+11</td>
<td>6.05e-10</td>
<td>(0.72)</td>
<td>6.00e-10</td>
<td>(0.71)</td>
<td>5.98e-10</td>
</tr>
<tr>
<td>Total Investment (log)</td>
<td>3.09</td>
<td>-1,175</td>
<td>(-1.10)</td>
<td>-1,121</td>
<td>(-1.03)</td>
<td>-1,133</td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td>2,2434*</td>
<td>(2.03)</td>
<td>21,042*</td>
<td>(1.71)</td>
<td>21,195</td>
</tr>
<tr>
<td># Countries</td>
<td>55</td>
<td>55</td>
<td>55</td>
<td>55</td>
<td>55</td>
<td>55</td>
</tr>
<tr>
<td>Observations</td>
<td>180</td>
<td>180</td>
<td>180</td>
<td>180</td>
<td>180</td>
<td>180</td>
</tr>
<tr>
<td>R-squared</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-Statistic</td>
<td>23.68</td>
<td>20.57</td>
<td>20.56</td>
<td>20.56</td>
<td>18.13</td>
<td></td>
</tr>
</tbody>
</table>

T-Statistic in parentheses

*** p<0.01, ** p<0.05, * p<0.1
REFERENCES


