A female teacher with short dark hair, wearing a teal zip-up sweater and grey trousers, sits on a white chair in a classroom. She is smiling and holding an open yellow book, reading to a group of children. The children are seen from behind, sitting on the floor. The classroom has a white table, a globe, and a chalkboard in the background. The text "Countries with performance pay for teachers score higher on PISA tests" is overlaid on the right side of the image.

**Countries with  
performance pay  
for teachers  
score higher on  
PISA tests**

# MERIT PAY International

By LUDGER WOESSMANN

American 15-year-olds continue to perform no better than at the industrial-world average in reading and science, and below that in mathematics. According to the results of the 2009 Program for International Student Assessment (PISA) tests, released in December 2010 by the Organisation for Economic Co-operation and Development (OECD), the United States performed only at the international average in reading, and trailed 18 and 23 other countries in science and math, respectively. Students in China's Shanghai province outscored everyone.

Many have identified variations in teacher quality as a key factor in international differences in student performance and have urged policies that will lift the quality of the U.S. teaching force. To that end, President Barack Obama has called for a national effort to improve the quality of classroom teaching and repeatedly indicated his support for policies that would provide financial rewards for outstanding teachers.

In a March 2009 speech to the Hispanic Chamber of Commerce, he explained,

*Good teachers will be rewarded with more money for improved student achievement, and asked to accept more responsibilities for lifting up their schools. Teachers throughout a school will benefit from guidance and support to help them improve.*

In the administration's Race to the Top initiative, the U.S. Department of Education encouraged states to devise performance pay plans for teachers in the hope that such an intervention could have a significant impact on student performance.

But is there anything in the data the OECD has accumulated to give policymakers reason to believe that merit pay works? Do the countries that pay teachers based on their performance score higher on PISA tests? Based on my new analysis, the answer is yes. A little-used survey conducted by the OECD in 2005 makes it possible to identify the developed countries participating in PISA that appear to have some kind of performance pay plan. Linking that information to a country's test performance, one finds that students in countries with performance pay perform at higher levels in

math, science, and reading. Specifically, students in countries that permit teacher salaries to be adjusted for outstanding performance score approximately one-quarter of a standard deviation higher on the international math and reading tests, and about 15 percent higher on the science test, than students in countries without performance pay. These findings are obtained after adjustments for levels of economic development across countries, student background characteristics, and features of national school systems.

stimulate individual effort on the job, it is theorized, but jobs where rewards are tied to effort attract energetic, risk-taking employees who are likely to be more productive. This latter consideration, says Stanford economist Edward Lazear, “is perhaps the most important” way in which a merit pay plan can influence worker performance. But if economists expect positive results from merit pay, many educators believe that teachers are motivated primarily by the substantive mission of the teaching profession and that they do not respond to—

## **Many educators believe that teachers are motivated primarily by the substantive mission of the teaching profession and that they do not respond to monetary incentives.**

I draw these conclusions cautiously, as my study is based on information on students in just 27 countries, and the available information on the extent of performance pay in a country is far from perfect. Further, the analysis is based on what researchers refer to as observational rather than experimental data, making it more difficult to make confident statements regarding causality.

It is possible that what I have observed is the opposite of what it seems: countries with high student achievement may find it easier to persuade teachers to accept pay for performance, thereby making it appear that merit pay is lifting achievement. More generally, both performance pay and higher levels of achievement could be produced by some set of factors other than all of those taken into account in the analysis. For example, performance pay could be more widely used in places where, as in Asia, cultural expectations for student performance are high, making it appear that performance pay systems are effective, when in fact both performance pay plans and student achievement are the result of underlying cultural characteristics. But even if my findings are not indisputable, I did carry out a variety of checks to see if any observable factor, such as Asian-European differences, could account for the conclusion. Thus far, I have been unable to find any convincing evidence that the findings are incorrect. Given that, let us take a closer look at what can be learned about the impact of performance pay from PISA data.

### **Prior Research**

Standard economic theory predicts that workers will exert more effort when monetary rewards are tied to the amount of the product they produce. Not only does performance pay

indeed, they may resent and resist—monetary incentives that tie salary levels to performance indicators.

To see whether the education sector is an exception to general economic theory, a number of performance pay experiments have been carried out, and in Israel and India such studies have shown positive impacts on student achievement. Experimental studies have tracked only the short-term impact of merit pay, however, and so have not identified any long-term effects that might come from changes in the kinds of people who choose to go into this line of work. Conceivably, a merit pay system could discourage entry into the profession of potentially excellent teachers reluctant to subject themselves to the requirements of a pay-for-performance scheme. Alternatively, if performance pay makes teaching more attractive to talented workers, short-term evaluations could understate its benefits.

One way to capture the long-term effects of teacher performance pay, including changes in the characteristics of those choosing to become a teacher, is to compare countries with performance pay systems to those without. This is now possible because the OECD in 2005 administered a separate survey to each of its member countries concerning the teacher compensation systems in place during the 2003–04 school year, when the 2003 PISA study was conducted. The 2003 PISA provides test score results in math, reading, and science for representative samples of 15-year-olds within each country, or nearly 200,000 students altogether. (The relative performance of countries on the PISA changed only slightly between the 2003 and 2009 tests. For example, in no subject did the scores for the United States differ significantly between 2003 and 2009.)

The PISA study is particularly useful, because it also includes information on a wide variety of family, school, and



institutional factors that are likely determinants of student achievement. My analysis adjusts, at the level of the individual student, for such characteristics as the student's gender and age, preprimary education, immigration status, household composition, parent occupation, and parent employment status. Nine measures of school resources and location are available, including class size, availability of materials, instruction time, teacher education, and size of community. Country-level variables included in the analysis were per capita GDP, teacher salary levels, average expenditure per student, external exit exams, school autonomy in budget and staffing decisions, the share of privately operated schools, and the portion of government funding for schools.

The PISA sampling procedure ensured that a representative sample of 15-year-old students was tested in each country. The student sample sizes in the OECD countries range from 3,350 students in 129 schools in Iceland to 29,983 students in 1,124 schools in Mexico. I therefore use weights when conducting my analysis so that each country contributes equally to the estimated effect of performance pay on student achievement.

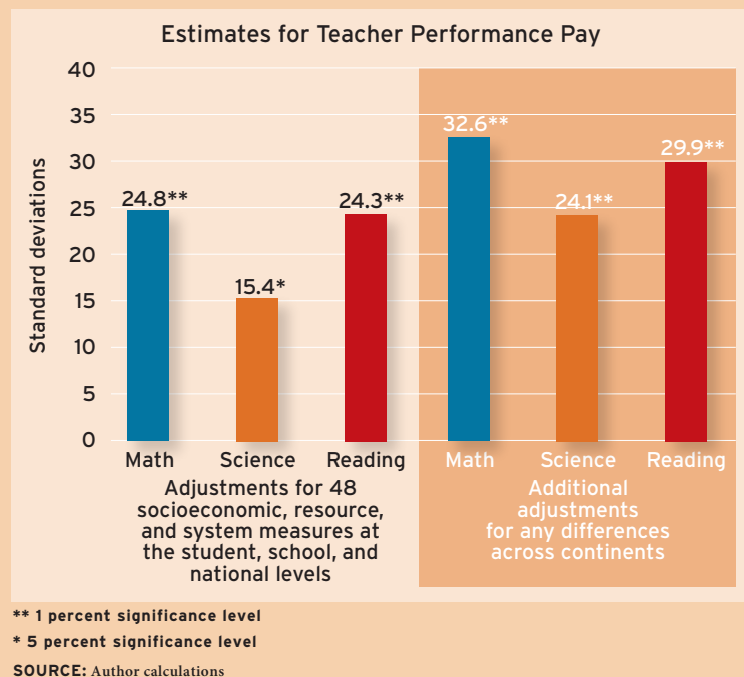
### Measuring Teacher Performance Pay

The measure of performance pay available from the OECD survey is less precise than one would prefer.

It simply asks officials in participating countries whether the base salary for public-school teachers could be adjusted to reward teachers who had an "outstanding performance in teaching." While the survey asked about many other forms of salary adjustments, the study protocol reports that this was the only one that "could be classified as a performance incentive." Among the 27 OECD countries for which the necessary PISA data are also available, 12 countries reported having adjustments of teacher salaries based on outstanding performance in teaching. The form of the monetary incentive and the method for identifying outstanding performance varies across countries. For example, in Finland, according to the national labor agreement for teachers, local authorities and education providers have an opportunity to encourage individual teachers in their work by personal cash bonuses on the basis of professional proficiency and performance at work. Outstanding performance may also be measured based on the assessment of the head teacher (Portugal), assessments performed by education administrators (Turkey), or the measured learning achievements of students (Mexico). Unfortunately, the coding of the measure does not allow my analysis to consider variation in the scope, structure, and incentives of performance-related pay schemes.

### Advantage for Countries with Performance Pay (Figure 1)

*Fifteen-year-old students in countries that can pay teachers based on their performance achieve at higher levels in math, reading, and science, even when compared only to students from the same continent.*



As an example of the limitation of this measure, note that the United States is coded as a country where teacher salaries can be adjusted for outstanding performance in teaching on the grounds that salary adjustments are possible for achieving the National Board for Professional Teaching Standards certification or for increases in student achievement test scores. That policy, however, affects only a few teachers in selected parts of the country. Given such weaknesses in the survey measure, it is all the more remarkable that I was able to detect impacts on student achievement.

### Main Results

As noted above, my main analysis indicates that student achievement is significantly higher in countries that make use of teacher performance pay than in countries that do not use it. On average, students in countries with performance-related pay score 24.8 percent of a standard deviation higher on the PISA math test; in reading the effect is 24.3 percent of a standard deviation; and in science it is 15.4 percent (see Figure 1). These effects are similar to the impact identified in the experimental study conducted in India and about twice as large as the one found in a similarly designed Israeli study.

Figure 2 depicts the math result graphically. The figure's vertical axis displays the average math test scores of students in each country after adjusting for all of the control variables in the model, with the exception of the variable measuring the use of performance pay. The horizontal axis in turn shows the performance pay variable, also after adjusting for those same control variables. The solid line on the figure shows the estimated relationship between these two variables across the 27 countries included in the analysis. It shows a clear positive association between the variation in country-average test scores and the variation in teacher performance pay that cannot be attributed to the other factors included in the analysis.

A lingering concern, however, is that the analysis may be contaminated by the fact that the very cultures that introduce merit pay are those that set high expectations for student achievement. The countries represent widely different cultures, including Asian ones, where expectations for students are often much higher than in Europe and North America. The best way to account for cultural differences among the continents of the world is to control in the analysis for the average effect of living on a particular continent, a strategy known to statisticians as continental fixed effects. Figure 1 thus also shows results based on models that include a fixed effect for each of the four continents with OECD countries: Europe, North America, Oceania, and East Asia. In these models, the effects of pay for performance are shown to be even larger than the results based on comparisons across continents. In other words, the findings cannot be attributed to cultural differences among the major regions of the world, because they are even larger when one looks only at patterns within these regions.

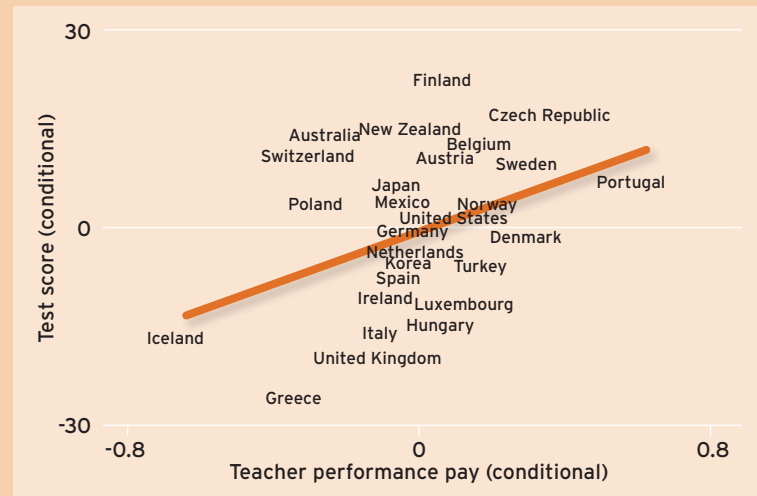
As a further test, I estimated the impact of performance pay for only the 21 participating European countries. Once again, the results showed even larger positive effects than those obtained for the full sample.

## Other Sensitivity Tests

When findings are based on small samples, it is important to ascertain whether a conclusion is sensitive to the particular analysis being conducted. Even after conducting a preferred analysis that maximizes use of the information available and best conforms to underlying economic theory, it is important to make sure that the pattern that one has identified is not a statistical accident that readily disappears if a slightly different analysis is conducted. For this reason, I performed a variety of sensitivity tests for math achievement because the reliability of the math test across countries and cultures

## Merit Pay and Student Achievement across Countries (Figure 2)

*All else being equal, countries with performance-related pay policies for teachers perform at higher levels on the PISA math exam.*



NOTE: The y-axis shows average math test scores after adjusting for all of the control variables in the model. The x-axis shows the performance-pay variable, also adjusted for the control variables. The solid line shows the estimated relationship between these two variables across the 27 countries.

SOURCE: Author calculations

is usually considered higher than it is for reading or science. Remarkably, the relationship between performance pay and math achievement remained essentially unchanged, regardless of the sensitivity test that I ran.

My first sensitivity check focused on cultural differences among countries that were not captured by the continental fixed effects analysis. In this sensitivity check, I excluded two countries, Mexico and Turkey, which have particularly low levels of GDP per capita. Since it is known that the level of GDP is strongly correlated with educational performance, it may be that the inclusion of these two countries is producing misleading results. But dropping these countries hardly affects results.

The second sensitivity test excluded the level of educational attainment of the teachers, on the grounds that teacher quality might itself be affected by a country's performance pay policies and therefore should not be used as a control variable. Excluding this variable did not materially change the results from those reported in Figure 1. In a third series of sensitivity tests, I excluded from the analysis one country at a time to make sure that the situation in no one country was driving the overall pattern of results. I found no evidence that that was happening.

The incidence of performance pay is, to some extent, clustered in two regions: Scandinavia (Denmark, Finland, Norway, and Sweden) and Eastern Europe (Czech Republic and Hungary). In a fourth set of sensitivity tests, I separately excluded the countries from these two regions from the analysis to see whether results were highly dependent on one or the other cluster. The results remained unchanged, indicating

that neither of these regional clusters is solely responsible for the main result.

A fifth set of sensitivity tests was possible because I have information on other policies that lead to differential pay among teachers. Salaries may vary depending on 1) the teaching conditions and responsibilities (such as taking on management responsibilities, teaching additional classes, and teaching in particular areas or subjects), 2) teacher qualifications and training, or 3) a teacher's family status and/or age. Since it is possible that student achievement is higher whenever pay schedules are flexible, regardless of the connection to teacher

an incentive system washes out any differences that may be caused by variations in teacher training.

## Conclusions

The analysis presented above represents the first evidence that, all other observable things equal, students in countries with teacher performance pay plans perform at a higher level in math, reading, and science. The differences in performance are large, ranging from 15 percent (in science) to 25 percent (in math and reading) of a standard deviation. Since one-quarter

# All other observable things equal, students in countries with teacher performance pay plans perform at a higher level in math, reading, and science.

classroom effectiveness, I estimated the impact of each of these three sets of factors on math achievement. None showed a significant impact on performance, and the effect of performance pay remained large and significant, even when these other possible salary adjustments were included in the analysis.

In sum, the main results shown in Figure 1 survive a wide variety of sensitivity tests. That the results are robust to multiple model specifications provides strong evidence that performance pay helps to explain the variation in student performance on the PISA tests.

## Differential Effects

With one exception (immigrants benefited less than native-born students from a performance pay regime), I found only small differences in the impact of performance pay on the math achievement of subgroups in the population. Since important differential effects were identified for only one subgroup, one cannot infer that the impact of performance pay on student math learning is concentrated on any particular group of students.

I did, however, find a surprising difference in the way in which a teacher's education background affects math learning, depending on the presence of a pay-for-performance system. In countries with performance pay, teachers who have an advanced degree in pedagogy do not outperform those without such a degree (the only measure of a teacher's education available in the PISA data base). However, in countries without performance pay, students learn more in math if they have a pedagogically trained teacher. Perhaps

of a standard deviation is roughly a year's worth of learning, it might reasonably be concluded that by the age of 15, students taught under a policy regime that includes a performance pay plan will learn an additional year of math and reading and over half a year more in science. However, this conclusion depends on the many assumptions underlying an analysis based on observational data.

Although these are impressive results, before drawing strong policy conclusions it is important to confirm the results through experimental or quasi-experimental studies carried out in advanced industrialized countries. Nothing in the PISA data allows us to identify crucial aspects of performance pay schemes, such as the way in which teacher performance is measured, the size of the incremental earnings received by higher-performing teachers, or very much about the level of government at which or the manner in which decisions on merit pay are made. Studies of such matters are probably better performed within countries, taking advantage of variation in policies within those countries. The study design also does not allow one to tease out the relative importance of the incentive to existing teachers of a performance pay plan as compared to the changes that may take place in teacher recruitment when compensation depends in part on merit rather than just on a standardized pay schedule. Since much more work needs to be done on all of these questions, a wit might insist that performance pay apply to scholars as well.

*Ludger Woessmann is professor of economics at the University of Munich and head of the department of Human Capital and Innovation at the Ifo Institute for Economic Research.*