LA SFIDA DELLA VALUTAZIONE

A CURA DI
NORBERTO BOTTANI E DANIELE CHECCHI

Saggi di Norberto Bottani, Eric Hanushek,
Andreas Schleicher, Maria Teresa Siniscalco,
Piero Cipollone, Luciano Abburrà, Daniele Checchi,
Olaf Koeller, Melania Rudin e Lucrezia Stellacci

SOCIETÀ EDITRICE IL MULINO

I lettori che desiderano informarsi sui libri e sull'insieme delle attività della Società editrice il Mulino possono consultare il sito Internet: **www.mulino.it**

ISBN 978-88-15-24075-0

INDICE

CAPITOLO SECONDO

THE COST OF IGNORANCE

As many of you know, I have flown in from California in the United States. I am going to give you the bird's-eye view of the world, first starting with a 10,000 metre view of the world that you get from the airplane. I will then try to come down closer to Earth and to talk a little bit more about what the beginning view means for policies. I actually have three very simple things that I want to say.

The first thing I am going to say is that the quality of education is very valuable. Here I will try to put in some perspective what achievement and knowledge and skills of a population mean for the future of nations in general and specifically of Italy.

Secondly, I am going to say that as far as we know the quality of teachers is the most important aspect of schools. You can almost say it is the only thing that counts. There are subtle differences of interpretation, but what we know in research is that teacher quality that counts for huge differences in student achievement.

And thirdly, I am going to say (and here we start to get a little bit more controversial) that there are institutions and incentives that are the key policies that we are going to be able to employ to improve our teachers' quality.

The overall message is very easy to say: The rewards to improving our schools are very, very large, but the policies that are needed are politically difficult. Nonetheless, unless we want to continue facing economic difficulties, we have to change the direction of our schools in order to improve student outcomes.

Questo capitolo è di Eric Hanushek.

Let me take our plane up to 10,000 metres. From here, I want to respond very forcefully to an argument that I hear often in the United States, namely, «while we do not do so well on PISA, it is not really that important; it does not really matter that we are not doing well». I hear that in the United States, and, to put it into perspective, the United States does slightly, but not enormously, better than Italy on these tests. So many of the things I say when addressing an audience in the United States are probably easily adapted to an Italian audience.

The message at this point is that complacency about performance on these examinations is badly wrong.

The picture that I want to use to make this point is found in figure 1. Looking at international test data, we obtain this simple graph. If you go along the horizontal axis, you get conditional test scores – I will explain conditional soon – and as you go along the vertical axis you get growth rates of countries. All of the entries in the figure indicate the location of countries going from Peru and South Africa and the Philippines at the bottom to Korea, Taiwan and Singapore up at the top. We can find the US somewhat above the line in the centre, and somewhere in the big knot of countries in the middle we will find Italy too.

The term «conditional» means that we adjust statistically for the starting level of country income. The graph shows the very simple statistical relationship between test scores of countries and growth rates *after* we allow for the income the countries started with. We have to allow for the fact that, if you start at a low level of GDP, it is easier to grow because all you have to do is copy everybody else. But, if you start at a higher level, you have to invent new things and so it is harder to grow.

The test scores I am showing are a composite of all the tests that any of these countries have taken between 1964 when the first international math test was given and 2003 which is at the end of our growth period. I

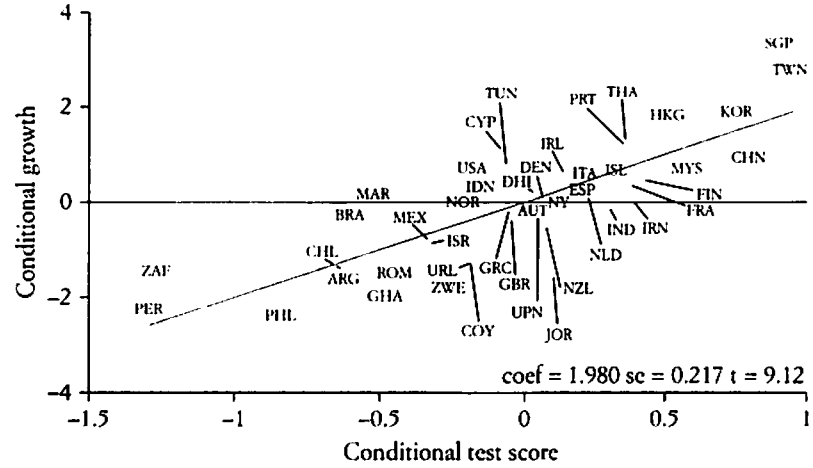

FIG. 1. The Relationship between Cognitive Skills and Economic Growth Rates.

have put all available tests (which could include 36 different possible scores for year-age-test combinations) together into one number for an economy. It measures, in my opinion, the cognitive skills, or the skills that are going to be important for a country. As one can see, countries are very close to the regression line. In other words, test score information explains most of the variation in growth rates around the world for all of the countries that we can look at.

I am going to contrast that with the more prevalent view that says we have to worry about school attainment; how many years of school do people get. We see the school attainment story in the World Bank's *Education for All Programme*, which basically sets the goal of getting everybody in the world up to an eighth grade education. We see it in the European Council's *Europe 2020 Vision*, which emphasises cutting down dropout rates from secondary school and increasing the attendance in tertiary education of other people. Let me show you the same picture for growth rates but by school attainment (fig. 2). There are two things that I would like to point out here. On the horizontal axis we now have school attain-

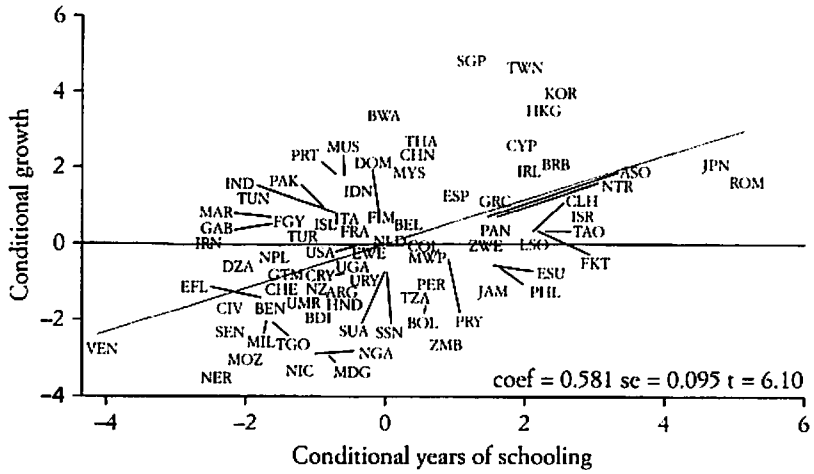

FIG. 2. The Relationship between School Attainment and Economic Growth Rates.

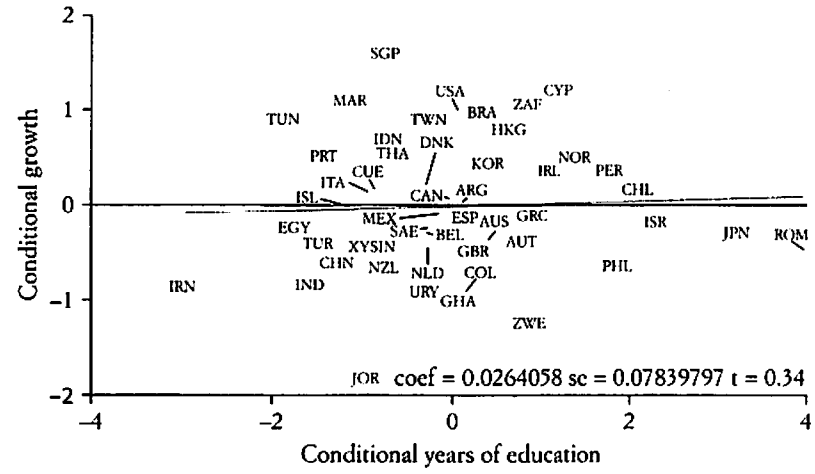

FIG. 3. The Relationship between School Attainment and Economic Growth Rates after Allowing for Cognitive Skills.

ment (measured by years of education, again, conditional upon starting income level and growth rates). You can see there is a positive relationship between school attainment, years of schooling, and growth rates, but you also see that the countries are spread quite a bit around that line; they are not as close to the line as I showed you before for cognitive skills. But let me show you the clincher to this. If, in addition to school attainment, we just add our measure of test scores to this (fig. 3), what you will see is that it is a horizontal line. There is no relationship between years of schooling and growth rates after we allow for what people know.

Now, I have to give you a confession. It took me several years to understand what this picture meant. This picture means if you go to school and do not learn anything, it does not count. But that is the simple story that we have with *Education for All.* In my opinion this is because, in many cases, we have people getting eight years of schooling, but they cannot read at the end of eight years. It is the knowledge that counts.

Now let me tell you what the slope of that line means, which relates to the growth rate of nations. To an economist it is very evident, and I think to many people and to your government it is also very well known that growth is what leads to prosperity in the future. It is what counts because wages only go up if productivity goes up and if you have growth in GDP.

So let us do some experiments. What I am going to do is start out with an experiment that says, let us think of reform policies where we can change the schools and improve the performance of kids on PISA test. We are not going to specify what the policy is but simply think of a reform that started today and that in 20 years would lead to a higher level of performance on PISA. Then I am going to do is calculate what the value of GDP would be with higher performance versus lower performance into the future, assuming growth into the future follows the patterns we have seen in the past. If future growth is related to achievement the way it is in the past, we can estimate its economic value of improved achievement. This is precisely what I have done this for all of the European Union countries and what I will now describe.

Let us start out with the simple example. Let us say that Italy, or all the countries in the European Union,

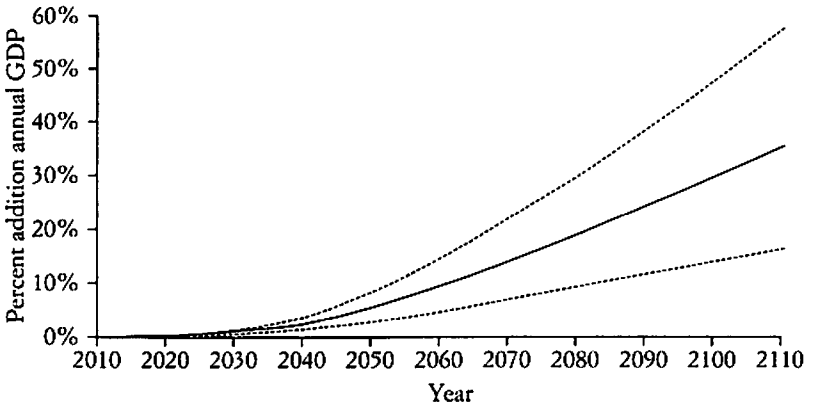

FIG. 4. Hypothetical Projections of Additions to GDP from School Improvement.

could improve their performance on the PISA test by 25 points. For Italy that means putting Italy students at a little bit above the Oecd mean or about the level of Germany. What I am going to do is project the GDP implications of that out into the future. Figure 4 provides a stylized version of the projections. The dark line here is the projection that you would get for percentage increases in GDP for each year into the future, according to the past growth that we have seen in the world and captured in the line I showed you in figure 1. I have got a couple of other lines on figure 4 just to emphasise at the beginning – there is some uncertainty in this and growth may fall someplace between the lighter lines. I am just going to concentrate on that line that represents the best guess of what happened in the past.

I am going to do a second thing. I am going to just look at the gains that would occur until the year 2090. What is 2090? That is about the life expectancy of somebody born today. Now I should say that I am going to allow for the fact that some of us, me at the beginning of the line, do not really care too much about 2090 because I am not going to see much of 2090. Consequently, I am going to discount those gains. In other words, I am going

to weight those values in the future a lot less. So I am going to calculate the present value in economist terms. The present value says what is the size of the cheque given to you today that, if you put it in the bank, you would get all these future gains. What I am going to do is just sum up all the gains into the future from this increase of 25 points in PISA test scores. It is actually a fairly simple exercise once we have the growth relationship with achievement.

The present value is 268 percent of the GDP of any of the countries from a 25 point gain in test scores. So in other words if your GDP is one trillion Euros, the value of 25 points gained over the next 20 years is 2.68 trillion dollars. For the entire EU it is a value of 32 trillion Euros, if all the countries in the EU could gain 25 points on PISA. For Italy, it is 3.8 trillion Euros if, in fact, you could put in place a policy that over the next 20 years would lead to a 25 point improvement on PISA. Seems like a worthy number!

Let me speak quickly about the causation issue. When we think about policy options, it is important that we can say «if we increase achievement, we can expect growth rates to increase». In other words, it is important that improved achievement causes improved growth as opposed just to being associated with it. For example, it could be the case that increased growth provides more income that leads to more investment in schooling and in turn to higher achievement. In other words, growth might cause achievement instead of the opposite. Questions of causation have been very controversial.

My co-author Ludger Woessmann at Munich and I have spent a lot of time looking at these issues. We have developed a variety of tests designed to eliminate the major concerns about whether there is a causal relationship. Everything we do, all the tests of whether achievement causes growth, can fail because it is hard to have definitive tests of this with these data. We have developed five or six different ways of trying to test this, all of which can fail, but all of which would fail for different reasons.

We get the same answers with each approach. So I am convinced that there is a causal relation.

But if you think there is a lot of doubt about whether this is a causal relationships, just cut the numbers in half. Say half of our observed relationship comes from the fact that higher growth causes more achievement and half comes from achievement causing growth. The projections cut in half mean that the 25 point increase would only be worth 134 percent of current GDP in Italy, as opposed to 268 percent. So causation is something to consider, but I think that the evidence is strong enough and the numbers are so large that something makes sense to try to do something to improve achievement.

Everybody envies Finland right now because it has jumped to the top of the league tables for PISA. So what would happen if everybody could be Finland in terms of achievement on the PISA test? This is a harder standard because equalling Finland is a very large leap for Italy as it is for the United States. Doing so, however, would lead to a present value that is seven times the GDP of all of the European Union countries, or 87 trillion Euros. What does it mean for Italy? 17.3 trillion Euros is the present value, or the lump sum equivalent today of getting the gains in growth from being Finland.

Figure 5 is a picture where I plot the gains for all the EU countries and the averages. We start with Finland. Finland is in an unfortunate position that it is already Finland, so it does not gain anything. But everybody else can gain the EU15 gains somewhere around 700 percent, a little bit larger for EU27. Italy (the striped column) is going to gain 11 times its current GDP, 1100 percent, in present value terms, or over 17 trillion Euros for being Finland in terms of achievement.

Let me look at another facet of this story. One part of the goals for the European Union, although one that has been deemphasised, is to get all but 15 percent of the students in each country up to level 1 on the PISA assessments for 15-year-olds. (Level 1 in science, for example, indicates students «have adequate scientific knowl-

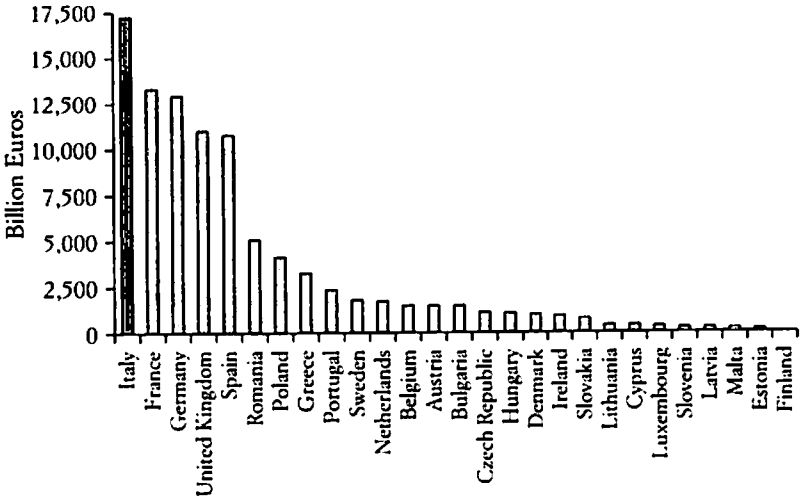

FIG. 5. Present Value of Added GDP from Bringing Student Achievement to the Level of Finland (billions of Euros).

edge to provide possible explanations in familiar contexts or draw conclusions based on simple investigations»). We can simulate what that would mean for each country to get all 85 percent of students up to level 1 on the PISA scores. The present value is almost two times GDP for the entire European Union, or 21.3 trillion Euros from the entire Union. For Italy, it is worth 5.8 trillion Euros to get to this goal of 85 percent of the students at level one on the PISA, or four times current GDP.

This is the story that I want to leave with you. There is a lot of uncertainty in this, of course, so again, if you think that I am way off, cut these numbers in half. The present values even cut in half still suggest that there are substantial gains to be obtained from paying attention to education. My personal view is that the gains are large enough that we should be willing to consider more radical changes than a small adjustment in class size or worrying about exactly what the math curriculum is, but instead should focus on larger things.

Let me talk a little bit about the kinds of reforms that I see from the research I do. Some of this uses international data, and some of it will rely much more on US data. I will show you why that is the case in a minute.

The initial policy option to consider is always the first one that people look at: «well, let us spend more». Figure 6 is a simple demonstration of why this is not an obvious policy to choose. This plots cumulative educational expenditures per student on the horizontal axis and PISA scores on the vertical axis. There are two lines on this. The solid line shows that there is a slight upward slope, indicating that there is a little bit of advantage to more money. Unfortunately that is an illusory, and it probably is not true. It is entirely driven by the very low performance of students in Mexico and Greece. If you do not believe that these two nations tell us much about what would happen in other countries if spending were higher, you get the second, completely flat, dashed line. In other words, outside of Mexico and Greece, there is absolutely no relationship between spending and student performance.

What I have been trying to paint here is that the future of Italy, of the United States, or of European Union countries is very different depending upon how each country's schools and the resultant skills of the population change. This aggregate performance information indicates that neither spending nor for that matter minor changes in class size, changes in curriculum, and the simple kinds of things that are politically easy to do will lead to very substantial changes in student outcomes. We tend to concentrate on these things because there is some surface plausibility to them and because doing them shows that we are trying to improve things. Unfortunately, they do not work by historical evidence.

This leads me to look in a different direction. The research consistently shows that teacher quality is the most important school input. Moreover, based upon the avail-

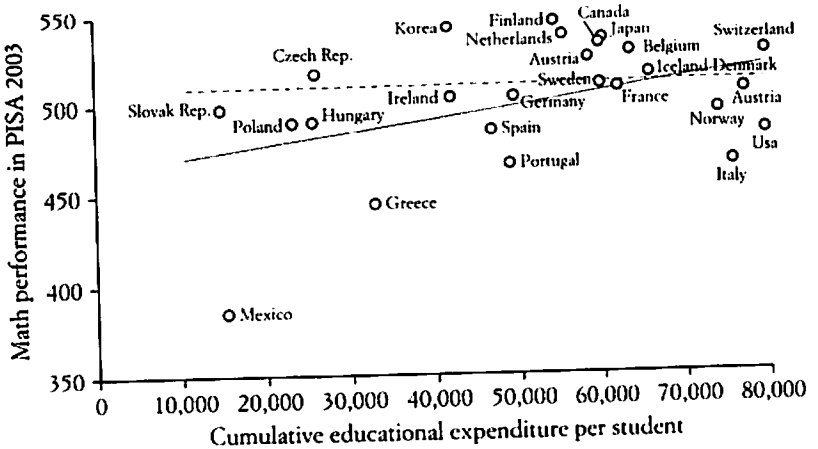

FIG. 6. Resources and Student Performance.

able research evidence, it takes improving the incentives to get good teachers in the classroom and to keep them there.

When talking about teacher quality, the starting point in the United States is that in any school, if you ask the principal of the school, if you ask the other teachers in the school, if you ask the parents in the school, you will receive a pretty reliable and systematic message about the teachers in the school. They will identify the very worst teachers and the very best teachers quite clearly, even if there is a lot of confusion in the middle. The US discussion has been how can we put in place a system that rewards the very best teachers to try to get them to stay in the school longer and that can eliminate, or move, the worst teachers on to other occupations.

The problems in the United States I think are in fact quite similar to those in Italy and many other places. The teachers' unions are not in favour of firing the bottom few percent of the teachers who are the worst. But in fact the dialogue and the discussions in the United States have been changing very rapidly on this in the last few years.

I will tell you one story that characterises this. At this time, there has been a lot of work that estimates statistically the «value-added of teachers». This value-added is the average growth in achievement of students in a classroom after you allow for everything else that goes on – families, peer influences, and the like. While there has been considerable academic discussion of various aspects of this research, the facts that there is wide variation in value-added of teachers and that it is possible to identify this in reasonable ways are now widely accepted.

Last summer, the Los Angeles Times, the major newspaper in Los Angeles, California, hired a consultant to take the data on student performance in the schools of Los Angeles and to estimate the value-added of all of the teachers in the Los Angeles school system. Los Angeles is a very large school system with about 700,000 students. Having received the estimates from this consultant, the newspaper then published value-added scores by teacher name in the newspaper, in the «Los Angeles Times».

Now, I suspect that you felt the shockwaves here in Italy when that happened. This was a very dramatic thing. It is not one that I would actually approve of doing on a regular basis. I do not think that personnel policies should be made in the newspaper. But, as a result of this, other cities are close to doing the same. In fact, after a lengthy lawsuit, New York City newspapers did exactly the same thing. New York is the largest school system in the United States, with more than a million students. More important than the response in New York City was what happened at the same time in Los Angeles. All of a sudden the teachers' union in Los Angeles and the school system in Los Angeles stood up and said «we have to have a better evaluation system for our teachers». Everybody had said something like this for a long time, and there was no doubt about the need for this. Everybody had said this, but nobody had actually done it. Everybody said, well, since we need a better evaluation system, let us have a committee to think about this. And the committees went on forever and

ever without actually doing anything. Now there is hope for actually doing something.

You may also know the story of Washington DC, Michelle Rhee was the Chancellor of schools in Washington DC. She is a very competent young woman who was the head of the school system in Washington Cashes who pushed hard on the evaluation of teachers and the ability both to remove teachers and to reward them. Two things happened. First, she put in a place the evaluation system based on classroom observations of teachers along with statistical value-added estimates and negotiated into the contract the ability to remove teachers that were in the bottom for two years in a row on these evaluations and to reward teachers on the top. This contract that was actually put in place in Washington DC was quite novel. The second thing that happened was that the mayor of Washington DC, who was very supportive of Michelle Rhee, lost the election to continue as mayor and was thrown out of office. Michelle Rhee has now retired from being the head of the school system.

The summary is that it is hard, but there is dialogue that is continuing now about how to insure that there are high quality teachers in the classroom. In my opinion, and I have been trying to convince the teachers' unions of this, it is in the union's interests to participate in this dialogue.

It is in the union's interests for several reasons. First, the majority of teachers in the United States are really very good, but they are being harmed by being lumped in with a few bad teachers. As a result their position has been undermined. That has led, in my opinion, to low salaries for teachers in the United States. It has held down salaries, because it is politicians who decide teacher's salaries. It is not an economic market, it is a political decision. The politicians cannot go out and say we are going to throw more money, higher salaries at teachers without recognising that there are some that are good and some that are bad. The public says, «You are not really going to pay higher salaries to that (poor) teacher, are

you?» and so I think that salaries could be much higher for teachers if the unions did not insist that every teacher is great. I personally think that salaries for good teachers should be a lot higher. The average salary in the United States for a teacher is about $55,000. I think that the top teachers should be getting much over $100,000 a year, something that is justified by the economic impacts that they are having.

An alternative to rewarding good teachers and removing bad teachers is the idea that we should do a better job helping all teachers to get better. In other words, if a teacher is not doing well, the system should do what it can to make the teachers better. I am not against this. In fact I entirely agree with the idea. But I am less convinced that we can deal with our problems in such a way.

We have two conflicting forces that provide the background for this in the United States. We have the idea that professional development of teachers should be the way to go because it can take existing teachers and make them better. Along with this are some examples by people who have developed systems of professional development that seem to show that they do better. But on the other side what we do not have is much evidence that you can do this on a broad scale and reproduce it.

There have been two really well designed studies of professional development in the United States. Let me quickly describe what I think is the best set of professional development studies to date. It was done by the Institute for Education Sciences, which is the research arm of the US Department of Education. They had a design that began with a large number of schools and tried to provide very direct programs of professional development in both the teaching of early reading and of middle school mathematics. The standard argument has been that professional development has not worked well in the past because it is not directed at what the teachers need to know and it has not been intense enough. So these studies were designed to exactly deal with those issues. The reading study began with what was reputed to be the best, or one

of the best, professional development programmes for early reading, concentrating on phonemic awareness and a variety of other developmental things that seemed to be important. The mathematics program did a similar thing for teaching fractions, but, because the results are similar, I will just describe the reading research.

In one group of schools, all the teachers of early reading got 45 hours of classroom instruction before the beginning of the school year in this professional development programme. A second group of teachers got 45 hours of classroom instruction in this professional development programme plus 60 hours of classroom coaching throughout the year, about 2 hours a week of classroom coaching throughout the year. The third group just got the same old stuff that was not working. At the end of the first year, it found that any teacher who got the classroom instruction tended to change her behaviour in terms of how she presented material and what she did in the classroom. Yet, the students did just as well in the classroom that got no professional development as either of the other two special programmes. In the second year of performance, the teachers generally reverted to their old way of teaching, and student achievement was unaffected by having better professional development.

The result was very similar in the study of lower secondary school mathematics training. There was no discernible impact on the learning of students when tested in exactly the subjects that were being put into professional development.

As a result, I have become sceptical about professional development. I would like to see it work because it is obviously an easier answer to take existing teachers and make them better, but I see little evidence that it works. Similarly, I see little evidence that it matters where and how a teacher was trained going into the teaching, at least on a broad scale. Having a regular teacher training degree is no different in terms of student gains in performance than coming with essentially a regular bachelor's degree with no pedagogy and no teacher training.

I am willing to have people continue to work on how to do better teacher preparation, but my focus has become trying to find a way that we can operate a system when we don't know how to prepare new teachers, but we know how to identify and select teachers.

I would pin my hopes on the ability to identify, select, and retain good teachers. Doing that has to do with making sure that teachers are happy with their job; that their pay is reasonable; that the working conditions they have are good; and that other things that lead teaching to be an attractive job are in place. But, it also requires having in place a counselling programme to help some of the teachers that are not very effective find alternative jobs, ones that are better for them (and better for the students who get different and more effective teachers). I would not pin my hopes on trying to figure out how can I make these ineffective teachers better teachers. In sum, I would try to encourage good, aggressive, hard-working, and effective people to be in the teaching profession and to stay there, but at the same time I would try to eliminate the ones that are harming kids.

Simply put, in the United States there are a small number of teachers that are just harming kids. I believe that we know who they are; everybody in the schools knows who they are. The principal, the other teachers, and the parents all know who are the bad teachers. We just will not act on that information.

I have a little bit of evidence also that the quality of the principal is important in how well a school operates, in the selection of teachers, and so forth. If you have a performance pay system for teachers, you have to have the same performance pay system for principals because you have to reward principals for doing the same thing that the teachers are doing, for getting higher achievement. It is always common to call for a teacher reward system with performance pay for teachers without saying anything about the administrators. But a simple analysis suggests that you have to have a parallel system of rewards for the performance.

The exact form of any reward system for teachers and administrators is a bit uncertain. We do not know much about how the collaboration among teachers in the school affects the outcomes. We suspect that that is important, that teachers do in fact help each other, but we do not have much evidence on that. It does suggest that, if you have any performance rewards in schools, a portion of the rewards should be school-wide such that everybody in school gets a reward and a portion should be individually related.

It is usually on the last issue that people have disagreement, that is, whether any individual should be rewarded. There is a group of psychologists that goes around saying you should not do that because that will lead to bad behaviour. My analysis is somewhat different. It is that if you want good teachers to be willing to go into schools to teach and you want good teachers to be willing to go into schools where the other teachers might not be so good, you cannot reward every teacher on the basis of the whole aggregate performance; you have to give some individual rewards along with maybe some aggregate rewards.

3. *Some Conclusions*

As Italy and the rest of Europe faces the difficult economic problems growing out of the 2008 recession, it cannot ignore the long run. In the long run, it is the human capital of nations that drives economic growth. And it is economic growth that determines the pattern of incomes across nations.

Historically, by far the most important determinant of growth has been cognitive skills of the population. These skills are measured well by the PISA assessments. On this, Italy needs improvement.

If Italy could, say, bring its performance up to about the level of Germany, history suggests stronger growth that would have a present value of three times the cur-

rent Italian GDP. Gains of this magnitude seem sufficient to call for significant changes in the schools.

The evidence suggests that the most important part of schools is the quality of teachers. While the quality of teachers has proven difficult to change, it does not have to be so. What is needed is an evaluation system for teachers that permits rewarding the best and eliminating the worst teachers. This kind of policy, of course, has met with political resistance. But, not pursuing improvements means that Italy is willing to accept a much lower standard of living in the future.