Print this page

Measuring pupils' progress can drive economic growth

Jan 9, 2015 | Martin Gustafsson

That basic mathematical skills should be a predictor of a country's economic growth surprised even economists, writes Martin Gustafsson

THE annual release of the matric results is justifiably important for pupils and their families. It is probably less significant as an indication of trends in the schooling system as a whole. There are other indicators that serve this purpose better. By understanding and promoting the good measurement of educational quality business can contribute to educational improvement.

Recent decades have seen major advances in educational psychometrics, or the science of testing. These in turn have permitted advances in economics, specifically a better understanding of the role of basic cognitive skills, largely generated in schools, in economic development. Eric Hanushek and Ludger Woessmann's work in this regard has been particularly influential.

Examination results are poor measures of educational progress, not just in SA, as they are influenced by a particularly wide range of factors including selection effects relating to who takes the examination in a particular year and what subjects pupils choose. What economists prefer to focus on are measures derived from single tests focusing on a specific basic skill such as mathematics or reading. That basic mathematical skills in, say, grade 8, should be a particularly good predictor of a country's subsequent economic growth initially surprised even the economists. It makes sense, however, if one considers that the quality of one's scientists, entrepreneurs and technocrats depends to a large degree on their foundational skills acquired early in life.

Participation in international and standardised testing programmes, or embarking on the arduous journey of establishing a robust national testing system, is a highly politicised matter. Ironically, expectations created by corporate social investment initiatives focusing on uplifting individual or small groups of schools can increase resistance to standardised testing.

We like to think that remarkable quality improvements seen in small-scale interventions can be replicated across an entire schooling system. Experience shows that this is not possible, not so much due to financial constraints but because human talent is limited.

We do not like to face up to the painful slowness of widespread quality improvements in education. Yet the danger if we do not monitor slow improvement is that we will begin believing there is no improvement, which can lead to a vicious cycle of radical policy change in the education system, followed by disillusionment that there was no magical take-off, more radical policy shifts, and so on.

One does not have to be vague or elusive about how large a change should be pursued, as long as one has reliable measures of educational quality — in other words, if one is able to compare apples to apples. Talking about annual change in average test scores as a percentage of the standard deviation in scores across students in a starting year has become popular among economists. Doing this allows one to compare countries that use different testing systems.

Arguably the best consistent educational quality improvement displayed by any country in recent years is Brazil's 2000-09 mathematics improvement in the Programme for International Student Assessment (Pisa). Brazil's improvement here amounts to 6% of a standard deviation per year. Roughly, we can consider this a limit for any developing country.

Extremely few countries have got anywhere near this speed of improvement. Yet it is more of a slog than a dash if one considers that it would take Brazil 25 years to reach the levels seen among the less impressive non-Asian members of the Organisation for Economic Co-operation and Development today, assuming that Brazil's present momentum is sustained.

So what are the features of a reliable, standardised testing system? One of the first things the psychometricians will tell you is that even the most capable teacher is unable to design two separate tests that produce equivalent

scores. This is not easy for many educationists to accept, because it violates a cherished assumption underlying many existing examination and assessment systems.

One solution is to use exactly the same test in different years and to ensure it is never leaked. But this is difficult or impossible logistically. A further problem is that with a single repeated test one ends up testing a rather narrow range of skills.

The Trends in International Mathematics and Science Study (Timss) programme, which, like Pisa, provides more or less a gold standard with respect to testing techniques, uses a combination of test secrecy, or test security, and anchor items, or questions that are common across different tests. It tries to make different tests as equivalent as possible, but this equivalence is tested post facto, using responses from students. Adjusted scores are calculated by looking at the relationship between anchor items and other questions in the test. In fact not only does Timss compile different tests for different years, within one year it also compiles different tests for randomly selected sub-sets of students.

For SA there are only two testing systems that provide an accurate trend over time, and they point to somewhat different trends.

SA's grade 9 Timss results from 2002 and 2011 point to a large improvement off a low base. The annual improvement comes to 7% of a standard deviation per year — in line with Brazil's trend of 6% a year.

An improvement in SA should not come as a complete surprise. Large improvements are easier the lower one's starting point. As countries approach human-ability ceilings, further improvement becomes much harder. In the years before 2011 SA saw increased spending on textbooks as well as a back-to-basics swing in curriculum policy.

The fact that the government has not extracted more political capital out of the Timss trend is perhaps indicative of the limited understanding there is in SA of the significance of these testing systems.

But there are also reasons to be cautious. The Southern and Eastern Africa Consortium for Monitoring Educational Quality (Sacmeq) programme pointed to no improvement in grade 6 mathematics and language between 2000 and 2007, meaning the Timss grade 9 improvement would all have to have occurred in 2010-11, assuming both testing systems are right. Unfortunately SA missed the 2007 round of Timss which would have helped to clarify matters. Many questions will be answered when the results from the 2013 run of Sacmeq and this year's run of Timss become available.

Apart from guiding education policies, standardised measures of human capital help to guide investment. The information emerging from Brazil, for instance, bodes well for long-term investment in industries focusing on innovation within that country.

Business in SA should be watching the evolution of the Annual National Assessments programme more closely. This programme has the potential to become, in part, a provider of reliable information on small but vital improvements. But that would require a clearer separation between the monitoring side of the programme, which needs to be sample-based to be practical, and the broader, more interventionist side, whereby schools and their communities are provided with school-specific performance data. Experiences with equivalent programmes in other developing countries, as well as the highly sophisticated national testing programmes found in certain developed countries, offer a number of lessons for us.

• *Gustafsson is with the department of economics at Stellenbosch University.*

~ o O o ~