This week the Minister of Basic Education released a report on the Annual National Assessment (ANA) of 2012, which amongst other things indicated that there have been considerable advances in literacy and numeracy of children in Grades 1 to 6 nationally. For instance, in Grade 3, the end of the Foundation Phase, the average language score improved remarkably since the 2011 ANA, from 35% to 52%; in mathematics the improvement was from 28% to 41%. In both cases, this was an improvement of almost half on the average 2011 score. In order to better understand what to make of the ANA results, the M&G spoke to Servaas van der Berg and Nicholas Spaull, two researchers who have done considerable research on educational performance. They are respectively Professor and Researcher in the Department of Economics at Stellenbosch University. Below are our questions and their response.

---

Q: ***Do you think the 2012 ANA results can be fairly compared to the 2011 results?***

A: Annual national testing of the sort that is done in ANA is of great importance in an under-performing educational system like ours, and the initiative to undertake the ANA is an important one. However, the ANA results for 2011 and 2012 should not be compared. It is only possible to draw valid comparisons between two tests if they are of equal difficulty. From the report it is unclear what measures were taken to ensure similar difficulty levels across years. Although the report mentions anchor items [common test items or questions between two tests], we don't know if these were used for equating purposes, i.e. to scale the tests to the same difficulty level. If this was done at all, if could only have been for Mathematics, as page 13 of the report explains that:

> "An attempt was made to link the 2012 test to the 2011 test by repeating a few of the items from 2011 in the 2012 test. This was, however, only possible for some grades and specifically with Mathematics. Language tests for Grades 4 to 6 could not link items from 2011 given that the language tests in 2011 were not demarcated into home language and first additional language" (p13).

If the tests were not equated (i.e. converted to have the same difficulty level), they cannot validly be compared. Furthermore, the ANA 2011 report explains that the HSRC created the database of item-level responses only for Grade 3 and Grade 6, thus it would have been possible to do equating for these grades only if this could have been done at all.

This there is no evidence that the tests for 2011 and 2012 were equated in a way that one could draw valid conclusions about educational progress. The large magnitude of the changes between these two years makes one sceptical that the tests were measuring at anything near the same difficulty level and that the changes documented in the report are plausible.

---

Q: ***Do you have any comments on the verification process of this year's results?***

A: From the report it would seem that there was no external, independent verification process for ANA 2012. In 2011 the HSRC verified the results of Grade 3 and Grade 6 by remarking a nationally representative sample of schools. This, year, the plan was that a different verification process would be followed, but what eventuated is not known to us.

---

Q: ***In some primary grades there were very large improvements from 2011 to 2012. For example, the Grade 3 literacy improvement from 35% in 2011 to 52% in 2012 – do you have any comment on this?***

A: Yes, we found that particularly strange. All the available evidence suggests that changes of this magnitude are simply not possible, locally or internationally.

It may help to provide some background information on the relative magnitude of these increases. The Grade 3 Literacy improvement of 17 percentage points year-on-year (a 49% increase) amounts to 0.70 standard deviations (based on the Grade 3 literacy scores from Verification ANA 2011). If one compares this to the largest improvers around the world, it would mean that South Africa has the fastest improving educational system in the world. If these results were true it would mean we have improved more in a single year than Columbia (0.52 standard deviations) did in 12 years from 1995-2007 – and Columbia was the fastest improving country of the 67 countries tested in TIMSS for this period. Or using a different cross-national survey, we have improved more in a single year than Russia did over the 2001-2006 period (0.54 standard deviations) in the Progress in International Reading and Literacy Study (PIRLS) – and Russia experienced the largest increase in student achievement of the 28 countries tested in PIRLS over this period. This is simply not possible.

One could also use local comparisons to provide a sense-check to the ANA 2011-2012 improvement. Every year the Western Cape conducts tests (Systemic Evaluations) of Grade 3 and Grade 6 students. These tests are marked centrally and not by the schools themselves. Between 2011 and 2012 there was almost no improvement in Systemic Test results in the Western Cape, yet according to the ANA results the Western Cape improved by 14 percentage points. Given that the Systemic Tests are calibrated to be of equal difficulty year-on-year, and that they are marked centrally, they are currently a more reliable indicator of true progress in learning than the ANA's and provide strong evidence that ANA is exaggerating any improvement that there may have been in learning in our schools.

Apart from international and local comparisons, the results for ANA 2012 do not appear internally consistent. If the results were calibrated to be of similar difficulty in each grade (which is necessary for inter-grade comparison), how is it possible that the Grade 1 mathematics average in 2012 was 68% but the Grade 3 average was only 41%, just two grades later? The performance further deteriorates to 27% in Grade 6 and a dismal 13% in Grade 9 (for which test

results are presented for the first time). Are these tests of equal difficulty for their grade? If so, it would indicate much better performance in the lower than the higher grades. Yet it would seem that there was no inter-grade linking of items, which is necessary to ensure that difficulty levels are similar. This is made explicit in the report: "There was no deliberate attempt to include questions to assess the degree to which the assessment standards of earlier grades had been achieved" (p67). Thus one cannot compare the results of one grade with the next, or say that performance is deteriorating as the grades progress.

To put it simply, it is not possible to compare two grades or two points in time unless the difficulty level of the tests are comparable. This can be determined by using Rasch analysis, a technique which requires some items (test questions) to be common across two tests that are being compared so that these can serve as anchors to calculate the difficulty levels of other items and put them on the same scale. After calculating the Rasch scores one can equate the difficulty levels of tests and adjust the marks accordingly.

---

Q: ***What are your thoughts on the Annual National Assessments in general – should they be abolished?***

A: Most certainly not. They must just be improved. The Annual National Assessments are an important and worthwhile endeavour and are needed to improve the quality of education in South Africa. The introduction of these tests is one of the most important advances in educational policy in recent years, as it provides a source of information for teachers, students, parents and policy makers that was absent before. Without a testing system like ANA it is not possible to determine which schools need what help, or to allow us to diagnose and remediate learning problems early enough such that they do not become insurmountable deficits. ANA provides information to teachers about the level they should assess at, and the level of cognitive demand that should aim at. It can provide objective feedback to parents about their children's performance, which is essential for them to know how the school system serves them and what learning deficits they may have. Parents and children have a right to know this, and poor and illiterate parents doubly so.

The real problem in our system is the failure of most students to master foundational numeracy and literacy skills in primary school, which then spills over into secondary schools. However, for the ANAs to provide the information on performance in schools, they need to be reliable indicators of learning across grades and over time. To this end the Department should put in place an independent verification process, and tests should adhere to international guidelines for standardised testing. The fact that ANA's results from 2011 and 2012 are incomparable is highly unfortunate. This means that schools, teachers and parents are getting erroneous feedback. Thus the 2012 ANA results, compared to those of 2011, creates an impression of a remarkable improvement in school performance which did not really occur. This would make it so much more difficult to really induce the improvement in behaviour at the classroom level that is central to real advances in learning outcomes.

The results of two international comparisons of educational achievement in 2011 (which South Africa took part in) will be released next week – TIMSS, which tested Mathematics and Science in secondary schools, and PIRLS, which tested Reading in primary schools. These will provide a better indication of any improvement or deterioration over the last six years. We have not seen these results, but would be greatly surprised if they point to anything like the improvement that the 2012 ANA report indicates.

---

Q: ***The national average for Grade 9 mathematics was 13%, with only 2% of students performing above 50%. Do you have any comments on this?***

A: Compared to the other results reported in the ANA 2012 report, this really stands out for how badly it portrays learning achievement in our schools. However, it is very difficult to know what this performance should be compared to. If this test was set at the appropriate level for Grade 9, does this indicate that the Grade 1 tests, which showed a 68% performance level in mathematics, was too easy? Or does it indicate that progressively, over the course of primary school, our children fall further and further behind the appropriate performance levels? We simply do not know, and until the DBE makes more information available on how difficulty levels were determined and linked across grades, we would not be able to tell.

**A note on Rasch:** Rasch is a technique of measurement, based on Item Response Theory (IRT), which is used to compare the difficulty levels of items in tests, and to rank students on such tests according to the same metric. Thus a student's ability at a point in time can be considered to be at a level where such a student has a fifty-fifty probability of correctly answering questions at that difficulty level.

One particularly useful application of Rasch is to equate the difficulty level of tests. If some test items are common across two tests, this can be used to calibrate the difficulty level of items in both tests, and adjustments can be made to test performance levels to make them comparable. This technique is widely used in testing systems throughout the world, and is, for instance, very common in Australia.