**Beyond comparisons: is there more to international benchmarking tests than comparisons between countries?**

Anne Oberholzer and Helen Sidiropoulos
Independent Examinations Board, South Africa

**Abstract**

The IEB provides the key school-leaving examination for independent schools that register with it for that purpose. In addition, the IEB provides assessment at other levels in the schooling system. One of these services is the offering of International Benchmarking Tests in partnership with the Australian Council for Educational Research (ACER). The tests are run in English and Mathematics from Year 3 to Year 10 and in Science from Year 3 to Year 8.

The key purpose for the project is the improvement of teaching and learning in participating schools. The IEB has been involved from 2007. In 2010 the first cohort of learners who wrote the assessments in Grade 9, wrote their final NSC examinations. In this paper I will report on the initial explorative study of the IEB into the correlation between performance in the Year 9 International Benchmarking Tests and their performance in the Grade 12 National Senior Certificate examination. The analysis of the data has been done by Dr Helen Sidiropoulos, an assessment specialist at the IEB.

To understand why schools in South Africa look for an independent international benchmark, one has to understand the diversity in the South African education system which is a direct result of our history and also some difficult decisions that have had to be taken since the establishment of the democratic government in 1994.

South Africa last participated in TIMSS in 2003 and PIRLS in 2006. Results were as follows:

| Test | Subject | Year | International Average | South African Average | Position |
|---|---|---|---|---|---|
| TIMSS (03) | Mathematics | 8 | 467 | 264 | 46/46 |
| TIMSS (03) | Science | 8 | 474 | 244 | 46/46 |
| PIRLS (06) | English | 4 | 500 | 302 | 45/45 |

South Africa is again participating this year in both TIMSS and PIRLS.Given the poor results released earlier this year in performance in our Annual National Assessments, while authorities are hopeful for an improvement realistically they are not expecting an improved performance.

The situation in South African education is not uniform. There are centres of excellence where the standard of teaching and learning is very good and learner achievement is comparable to the best in the world. The difference in performance across the system by and large reflects the advantage or disadvantagein the schools in the previous political regime.

A 2010 study by Prof Charles Simkins illustrates the diversity in performance in the National Senior Certificate examination, the major school leaving examination, after 12 years of schooling. In his study of performance in Mathematics and Physical Sciences, with reference to a graph he had constructed, he noted the following:

> It shows that the bottom 75 per cent of schools produced only 17 per cent of the passes, whereas 6.6 per cent of schools produced 50 per cent of the passes.
>
> The same curve can be drawn for science passes at 50 per cent or better (figure 2). Here the bottom 75 per cent of schools accounted for 13 per cent of the science passes, while 5.5 per cent of schools produced 50 per cent. Therefore, the distribution of science passes is even more unequal than the distribution of mathematics passes.
>
> Put another way, South Africa relies on just more than 400 schools for half its mathematics passes at the 50 per cent level, and about 350 schools for half its science passes at the 50 per cent level. [1]

The high performing mathematics and science schools tend to be the previously white only schools and by and large still have a teaching force of quality. Given that there are some 5 670 secondary state schools in South Africa[2], one can see that the diversity in performance is stark.

While resources and the quality of teaching[3] are significant reasons for the stark difference in performance, the key issue, in my opinion, that hasbedevilled education in South Africa is the language policy in education. South Africa has 11 official languages. The language policy allows for mother-tongue education until the end of Grade 3 and from Grade 4, schools must choose as a language of instruction, one of the two recognised Languages of Learning and Teaching (LOLT) i.e. English or Afrikaans. These were the LOLTs of apartheid South Africa and are the home languages of the majority of white South Africans, while they are at best a second language, but frequently the third language, of many Black South Africans. As Dr Neville Alexander from Project for Alternate Education in South Africa (PRAESA) says, "those who were advantaged before apartheid remain the advantaged after apartheid"[4].

The language question in South Africa is complex – in some areas e.g. city centre Johannesburg, a teacher could quite easily have a class in which the home language of learners covers six or seven languages, not all of which are official languages in South Africa. The Johannesburg city centre is home to many legal and illegal immigrants to South Africa as well as South Africans from all over the country. To determine what constitutes the mother tongue for a class of that nature is impossible. More impossible is to expect that the average teacher has sufficient command of all the languages present in his/her classroom to be able to provide mother-tongue instruction, even in Grades 1 to 3. In such instances, English is normally the 'common' language. However English is often the teachers' second or third language and hence his/her use of English as the LOLT poses its own set of problems. There are numerous reports and academic articles in South Africa that note the inadequacy of the learners and the teaching community particularly in the lower phases, to cope with the complex language issues they face.[5] While the language policy is not the only problem in education in South Africa, it is certainly a significant factor in accounting for the generally poor performance of learners.

South Africa as a country has achieved poorly in both national and international assessments. However these results do not reflect the diversity in the system and hence good schools, while seeming to be high performing in comparison to national and international norms, do not know whether the performance of their learners is comparable to countries that have performed well in international assessments. Being part of the ACER project provides individual schools with an opportunity to compare themselves against a more realistic standard in respect of what can be expected from them, given their resources and the quality of the learning and teaching at the school.

Some of the 2010 results for the IEB sample of South African schools that wrote the ACER tests suggest that there are schools in South Africa that are providing a much better standard of education than the national TIMSS and PIRLS results indicate.

Grade 8: ACER scaled score average for Maths: +/- 551 IEB sample +/- 572
Science: +/- 545 IEB sample +/- 545
Grade 4: English: +/- 485 IEB sample: +/- 512

The ACER tests hence provide a point of comparison that is more in keeping with the schools' aspirations. They also provide an incentive for improvement.

Information about the ACER tests is available from their website[6]. The ACER tests are developed by highly experienced test developers at ACER using Item Response Theory. Anecdotal reports from schools indicate that they feel that the questions are fair, well-constructed and of a very good standard.The steadily increasing enrolment for these tests managed by the IEB in South Africa certainly indicate that they believe the assessments are beneficial – from 6 800 in 2007 to 28624 in 2010.

One of the draw cards for schools is the very comprehensive and quality set of reports. These are as follows:

| Type of Report | Sections | Uses |
|---|---|---|
| Student | Question by question analysis<br>Strand analysis<br>Percentile rank comparisons<br>Scale score comparisons | Indicates strength or weakness in particular questions or a particular strand area<br>Compares a student within groups and across years |
| Group | Summary page<br>Question descriptors<br>Test results ordered by item number<br>Strand and item order<br>Strand and item difficulty | Indicates strength or weakness in the group for particular questions or a particular strand area<br>Compare groups of students from one year to the next |
| TIMSS (only for Maths and Science) | Gives an estimate for how students have performed compared to those from other countries in the TIMSS project | Compare individuals or groups of students with TIMSS country results |

The individual student reports measure and compare performance and highlight individual skills, strengths and weaknesses. Group/class reports allow schools to rank and compare student performance and identify individual, class and school-wide strengths and weaknesses. The school, regional and international comparison reports enable schools to compare and benchmark student and school performance.

The comparative framework for the school to assess their own performance consists not only of school and country comparisons, but also two international groups: the ACER IBT International Sample and in Mathematics and Science, there are comparisons against selected countries participating in the Trends in International Mathematics and Science Study (TIMSS).

The following excerpts of reports are available from the ACER website and illustrate the detail in individual learner reports which give schools enormous data not only on the strengths and weaknesses of specific learners but also the school as a whole.

The question-by-question analysis shows the strengths and weaknesses of individual students.

| No | Question Description | Strand | Your Answer | Result | Correct Answer |
|---|---|---|---|---|---|
| 2 | Recognise name of 4-digit number | Number | D | × | B |
| 17 | Interpret two-way table | Chance & Data | C | ✓ | C |
| 30 | Recognise 3D model in rotated position | Space | A | × | D |

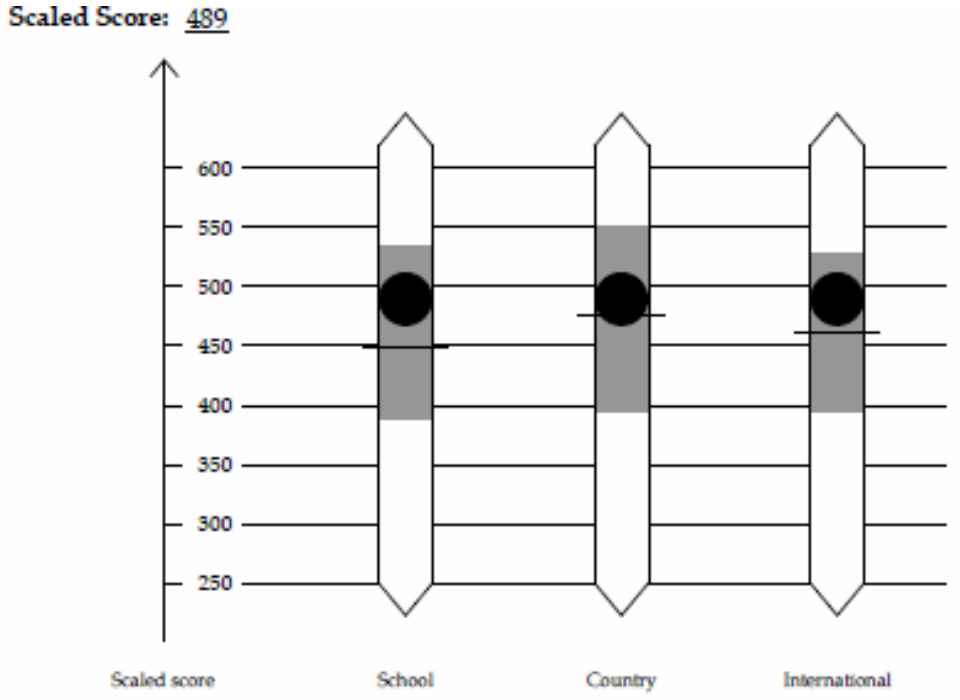
Subject areas are assessed in strands and the strand reports given teachers an indication of whether there are strands in which the learner and/or the school is particularly strong or weak. The following shows the learner report according to strands in Mathematics.

Summary of Results by Strand

| Breakdown by Strand | Number of Items in Strand | Number Correct | Percent (%) Correct | Average (%) Correct (Country) |
|---|---|---|---|---|
| Chance & Data | 5 | 5 | 100.0 | 51.8 |
| Measurement | 7 | 3 | 42.9 | 33.1 |
| Number | 20 | 18 | 90.0 | 54.2 |
| Space | 8 | 1 | 12.5 | 51.4 |
| TOTAL | 40 | 27 | 67.5 | 49.6 |

Using anchor items across year groups and across years of conduct, a scaled score is calculated for each student. This individual student scaled score comparison allows a teacher to

- accurately compare different students ;
- accurately compare a student's performance from year to year i.e. measure growth over time;
- accurately compare students to international benchmarks.

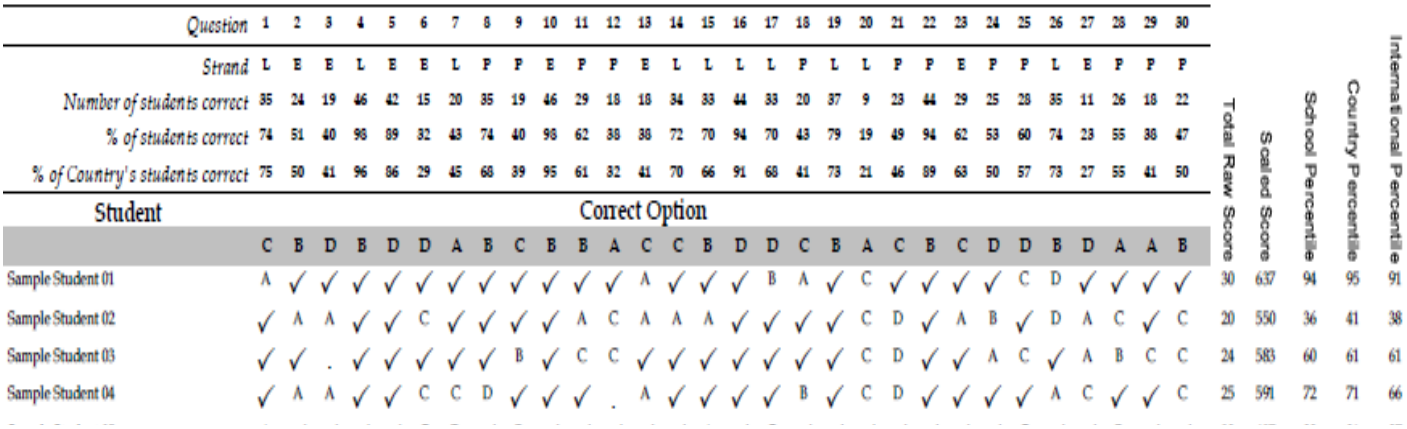



School reports are also comprehensive. The question descriptor section of the school report provides teachers with a clear idea of what concepts are assessed in each question. So while schools are not able to keep papers, teachers are able to see what aspects of the subject area are being assessed.

**Question Descriptors**

| *Question* | *Strand* | *Description* |
|---|---|---|
| 1 | Life Sciences | Understands basic properties of genetic inheritance |
| 2 | Earth Sciences | Connects information in table with picture |
| 3 | Earth Sciences | Uses information in table to answer question |
| 4 | Life Sciences | Identifies bone that is most likely to be a leg bone |
| 5 | Earth Sciences | Understands basic geological processes |
| 6 | Earth Sciences | Understands basic geological processes |
| 7 | Life Sciences | Draws conclusions from information in a diagram of a skeleton |
| 8 | Physical Sciences | Identifies lowest average temperatures from a table |
| 9 | Physical Sciences | Understands the trend in average temperatures from a table |

The school report records the responses of all learners and hence provides the school with a detailed summary of the performance of the school in each question and in each strand in comparison with the country.



| Question | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | Total Raw Score | Scaled Score | School Percentile | Country Percentile | International Percentile |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Strand | L | E | E | L | E | E | L | P | P | E | P | P | E | L | L | L | L | P | L | L | P | P | E | P | P | L | E | P | P | P | | | | | |
| Number of students correct | 35 | 24 | 19 | 46 | 42 | 15 | 20 | 35 | 19 | 46 | 29 | 18 | 18 | 34 | 33 | 44 | 33 | 20 | 37 | 9 | 23 | 44 | 29 | 25 | 28 | 35 | 11 | 26 | 18 | 22 | | | | | |
| % of students correct | 74 | 51 | 40 | 98 | 89 | 32 | 43 | 74 | 40 | 98 | 62 | 38 | 38 | 72 | 70 | 94 | 70 | 43 | 79 | 19 | 49 | 94 | 62 | 53 | 60 | 74 | 23 | 55 | 38 | 47 | | | | | |
| % of Country's students correct | 75 | 50 | 41 | 96 | 86 | 29 | 45 | 68 | 39 | 95 | 61 | 32 | 41 | 70 | 66 | 91 | 68 | 41 | 73 | 21 | 46 | 89 | 63 | 50 | 57 | 73 | 27 | 55 | 41 | 50 | | | | | |
| Student | Correct Option | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | C | B | D | B | D | D | A | B | C | B | B | A | C | C | B | D | D | C | B | A | C | B | C | D | D | B | D | A | A | B | | | | | |
| Sample Student 01 | A | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | A | ✓ | ✓ | ✓ | B | A | ✓ | C | ✓ | ✓ | ✓ | ✓ | C | D | ✓ | ✓ | ✓ | ✓ | 30 | 637 | 94 | 95 | 91 |
| Sample Student 02 | ✓ | A | A | ✓ | ✓ | C | ✓ | ✓ | ✓ | ✓ | A | C | A | A | A | ✓ | ✓ | ✓ | ✓ | C | D | ✓ | A | B | ✓ | D | A | C | ✓ | C | 20 | 550 | 36 | 41 | 38 |
| Sample Student 03 | ✓ | ✓ | . | ✓ | ✓ | ✓ | ✓ | ✓ | B | ✓ | C | C | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | C | D | ✓ | ✓ | A | C | ✓ | A | B | C | C | 24 | 583 | 60 | 61 | 61 |
| Sample Student 04 | ✓ | A | A | ✓ | ✓ | C | C | D | ✓ | ✓ | ✓ | . | A | ✓ | ✓ | ✓ | ✓ | B | ✓ | C | D | ✓ | ✓ | ✓ | ✓ | A | C | ✓ | ✓ | C | 25 | 591 | 72 | 71 | 66 |

The following indicators are given for each learner:

- Total Raw score
- Scaled score
- School percentile
- Country percentile
- International percentile

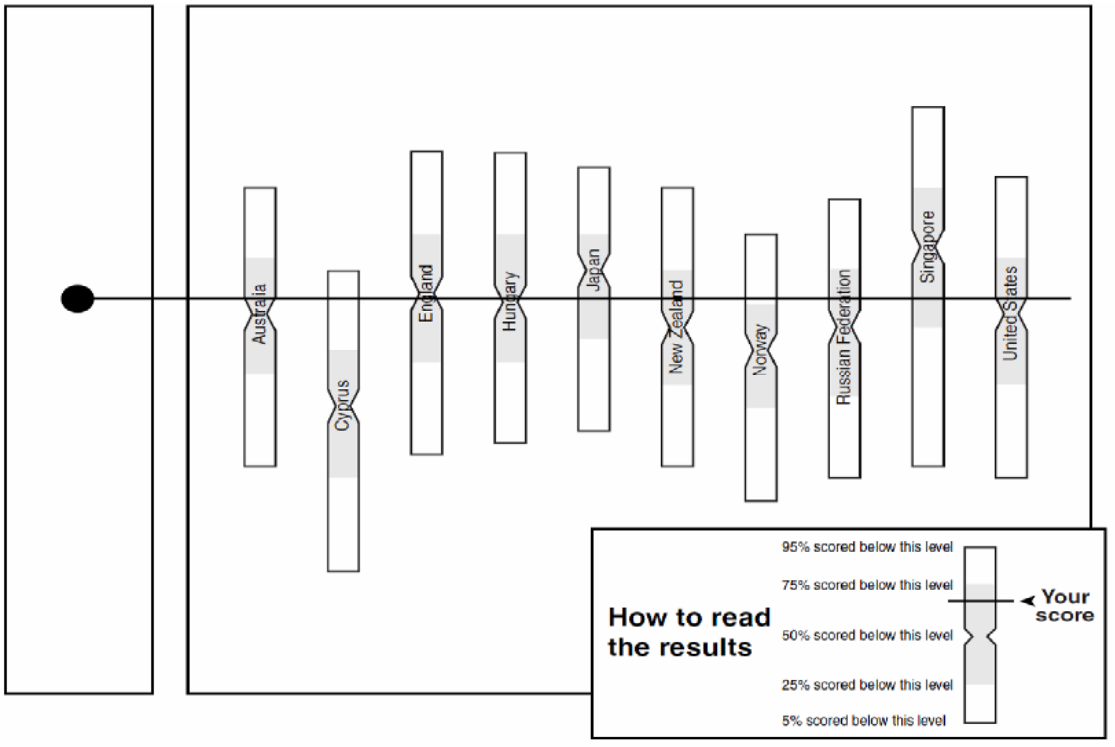
The percentile indicates the percentage of students above whom the student in question has performed. Hence the student below has scored better than 72.3% of students in the school, 67.8% of students in the country and 60.1% of students in international comparison group.

Percentile Rank

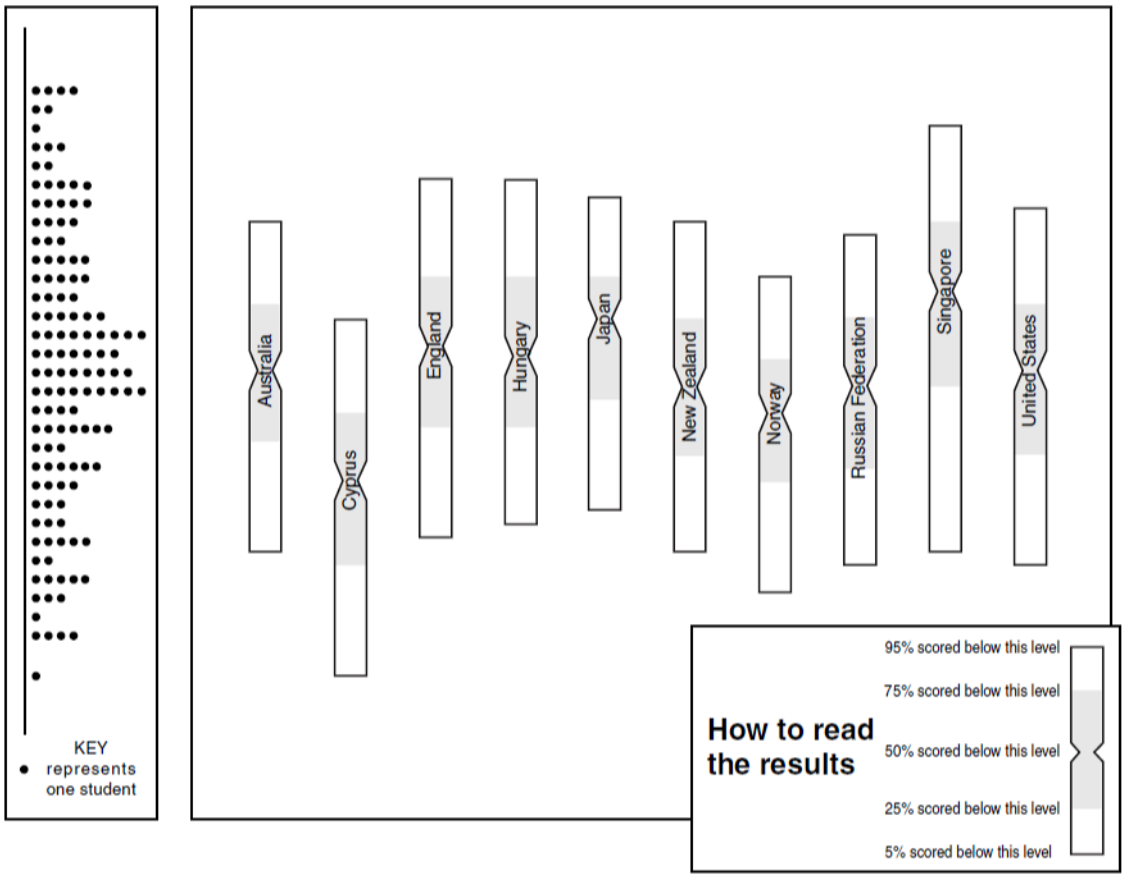
| School Percentile | Country Percentile | International Percentile |
|---|---|---|
| 72.3 | 67.8 | 60.1 |

This allows schools to assess how their top and bottom students compare against others in the country and internationally.

Because of ACER's work with TIMSS, there is a comparative graph of how individual learners compare against a selection of countries that have participated in TIMSS (Mathematics and Science). A similar report which accommodates all learners helps a school assess their standards against the same selection of countries.




Schools report against selected countries that participated in TIMSS



The information from the reports is so rich that many schools have indicated to the IEB that they have meaningfully adjustedtheir teaching program so that they can attend to the weaknesses as indicated in their reports. Most importantly, many have commented that the way the questions are asked in the IBTs militates against schools providing children with recipes of 'how to do' and hence teachers are compelled to teach in such a way that learners understand the conceptsto ensure that when the question is presented in a different form, learners still recognise the concept.

The IEB introduced these tests in Grade 9 in 2007, which means these learners wrote the National Senior Certificate (NSC) examination in Grade 12 in 2010. The IEB has compared the IBT results of the sample who wrote in 2007 with their Grade 12 results. The IEB sees this as the start of a piece of longitudinal research to see whether there is any relationship between the IBT results and the Grade 12 NSC results in selected subjects, and if so, the extent to which the IBTs have a predictive value for learners in the main school-leaving examination after 12 years of learning.

There were some 400 learners in the 2010 NSC Mathematics examination whose performance could be compared to their performance in the Grade 9 IBTs in 2007. The emerging trend is that learners who do well in the IBTs are almost certainly going to do well in the Grade 12 examination, even if their Grade 9 school results are not necessarily good. In addition, proving the well-established educational truth that no learner should be 'written-off' based on their results in tests early in their learning career, there are a number of learners who did not score well in the IBTs but did manage to score 80% or more in the end of Grade 12 NSC examination. There are a few sceptics who would suggest that the NSC examination is assessing skills and knowledge that can be taught and learnt – drill and practice - and does not require intuitive understanding of the subject area, while success in the IBTs does require more than just learnt skills and knowledge.

In the comparison of learners' scores in the Mathematics NSC and the Grade 9 IBT, the following emerged:

| Grade 9 IBT Scaled Score (S) in Mathematics of: | IEB Mathematics Grade 12 final NSC % |
|---|---|
| $S \geq 700$ | 89 |
| $680 \leq S < 700$ | 86 |
| $660 \leq S < 680$ | 77 |
| $600 \leq S < 660$ | 69 |
| $570 \leq S < 600$ | 61 |
| $530 \leq S < 570$ | 56 |
| $S < 530$ | 50 |

The performances are closely aligned, with the Pearson correlation coefficient between Grade 9 IBT mathematics performance and Grade 12 mathematics performance at 0.569 (significance: $p < 0.001$). This implies that Grade 9 Maths IBT scaled scores can be considered as a good predictor of Grade 12 mathematics performance.

The IEB conducted a further analysis of the results of learners in the 2007 IBT for English against their overall results in the 2010 NSC examinations. The study looked at learners with specific scaled scores in the English IBT in Grade 9 in 2007 and compared that against their general NSC performance in 2010. More specifically we compared their IBT English scores

against the number of distinctions (80% or higher) they achieved in the group examination of 7 subjects. While this study considered only 200 learners, an interesting trend has emerged.

| Learners with an English Grade 9 IBT 2007 Scaled Score (S) of: | NSC examination (Grade 12) Number of distinctions (80% +) |
|---|---|
| $S \geq 700$ | 5.85 |
| $650 \leq S < 700$ | 3.85 |
| $600 \leq S < 650$ | 2.65 |
| $560 \leq S < 600$ | 1.05 |
| $S < 560$ | 0.48 |

This suggests that the scaled score in the English IBT at Grade 9 is a predictor of general performance in Grade 12.

Both studies are limited in respect of the number of learners considered, namely 400 in the Mathematics and 200 in the English. However the IEB will continue with this research, tracking a greater number of learners to see whether more specific trends can be established.

We do compare performance across subjects and that research shows conclusively that performance even in mathematics and science subjects is very closely aligned to language performance, especially at the higher end of the spectrum.

| English Gr 12 | English Gr 12 (Av. %) | Life Sciences Gr 12 (Av. %) | Difference | Number of candidates | Correlation | Significance P< |
|---|---|---|---|---|---|---|
| 90%-100% | 91.24 | 87.97 | 3.27 | 59 | | 0.1 |
| 80%-89% | 83.04 | 81.39 | 1.65 | 1025 | | 0.001 |
| 80%-100% | 83.49 | 81.75 | 1.74 | 1084 | | 0.001 |
| 70%-79% | 74.09 | 74.20 | -0.11 | 2488 | | 0.001 |
| 60%-69% | 64.71 | 64.03 | 0.68 | 2748 | | 0.001 |
| 50%-59% | 55.35 | 52.23 | 3.12 | 1341 | | 0.001 |
| 40%-49% | 45.91 | 42.05 | 3.86 | 340 | | 0.001 |
| 30%-39% | 36.38 | 28.97 | 7.41 | 37 | | not |
| 0%-29% | 23 | 23.67 | -0.67 | 3 | | not |
| 0%-100% | 67.64 | 66.66 | 0.98 | 8041 | 0.842 | 0.001 |

| English Gr 12 | English Gr 12 (Av. %) | Mathematics Gr 12 (Av. %) | Difference | Number of candidates | Correlation | Significance P< |
|---|---|---|---|---|---|---|
| 90%-100% | 91.45 | 89.86 | 1.59 | 112 | 0.128 | not |
| 80%-89% | 83.20 | 80.44 | 2.76 | 1722 | 0.229 | 0.001 |
| 80%-100% | 83.63 | 81.02 | 2.61 | 1834 | 0.287 | 0.001 |
| 70%-79% | 74.12 | 69.53 | 4.59 | 3770 | 0.251 | 0.001 |
| 60%-69% | 65.00 | 57.57 | 7.43 | 3467 | 0.255 | 0.001 |
| 50%-59% | 55.64 | 45.56 | 10.08 | 1234 | 0.220 | 0.001 |
| 40%-49% | 45.96 | 32.08 | 13.88 | 262 | 0.262 | 0.001 |
| 30%-39% | 37.43 | 25.91 | 11.52 | 23 | 0.504 | 0.05 |
| 0%-29% | 28 | 3 | 25 | 1 | N/A | N/A |
| 0%-100% | 69.83 | 63.78 | 6.05 | 10591 | 0.666 | 0.001 |

| English Gr 12 | English Gr 12 (Av. %) | Physical Sciences Gr. 12 (Av. %) | Difference | Number of candidates | Correlation | Significance P< |
|---|---|---|---|---|---|---|
| 90%-100% | 91.44 | 84 | 7.44 | 94 | 0.097 | not |
| 80%-89% | 83.22 | 77.22 | 6 | 1309 | 0.318 | 0.001 |
| 80%-100% | 83.77 | 77.68 | 6.09 | 1403 | 0.361 | 0.001 |
| 70%-79% | 74.18 | 67.32 | 6.86 | 2735 | 0.332 | 0.001 |
| 60%-69% | 64.94 | 56.12 | 8.82 | 2431 | 0.326 | 0.001 |
| 50%-59% | 55.43 | 44.01 | 11.42 | 1054 | 0.339 | 0.001 |
| 40%-49% | 46.03 | 32.72 | 13.31 | 250 | 0.311 | 0.001 |
| 30%-39% | 37.05 | 24.15 | 12.9 | 20 | 0.246 | not |
| 0%-29% | 28 | 9 | 19 | 1 | N/A | N/A |
| 0%-100% | 69.54 | 61.39 | 8.15 | 7894 | 0.787 | 0.001 |

The correlation between the English Home Language scores and the other subjects is significant and especially from 60% and above. This year we are looking at determining what it is that differentiates a C from a D candidate in Home Language, as that appears to be a significant point.

Our work is on-going. We are particularly keen to further explore the predictive value of the IBTs. However the work this year has re-emphasised the importance of language in academic success and for South Africa, it highlights the deficiency in the current policy of language of learning and teaching.

**References**


1. The Maths and Science performance of South African public schools, Charles Simkins, CDE, 2010
2. www.southafrica.info, 2007
3. The quality and quantity of South African teachers, The Centre for Development and Enterprise (CDE), September 2011
4. Sink or Swim – video released by PRAESA
5. An aspect of language for academic purposes in secondary education: complex sentence comprehension by learners in an integrated Gauteng school, D Van Rooyen and H Jordaan, South African Journal of Education, May 2009;
   National Reading Strategy, Department of Education, 2008;
   Dr Neville Alexander and the work of PRAESA at the University of Cape Town;
   Various reports from Joint Education Trust
6. www.acer.edu.au/tests/ibt.