Paper delivered at the HSRC Conference, Birchwood, 12-13 June 2003

THE TRENDS IN INTERNATIONAL MATHEMATICS AND SCIENCE STUDY:2003

Vijay Reddy, Gerda Diedricks, Lolita Meintjies

Assessment Technology and Education Evaluation

1. The History of TIMSS

The Trends in International Mathematics and Science Study (TIMSS) is a large-scale study conducted by the IEA (International Association for the Evaluation of Educational Achievement). The IEA has organized more than 15 large-scale comparative studies including two in Mathematics, two in Science and one in both Mathematics and Science (TIMSS). These studies have been going since the 1960s and in 1995 a study called Third International Mathematics and Science Study was conducted. In 1999 a repeat of the study was conducted and it was called TIMSS-R and in 2002/2003 the name was revised and it was called Trends in International Mathematics and Science Study (TIMMS).

The primary goal of International comparative studies of achievement of mathematics and Science is to evaluate the levels of achievement of students in various countries. Even more important is to understand and explain the differences that emerge.

There are three minimum conditions, which have to be met in order for an international survey to be reliable:

- Samples of schools and students in each country must be fully representative
- Test must be as fair as possible to all countries
- Administrative procedures must be similar in all countries

These conditions are difficult, time consuming and very costly to meet and a study that went a long way in fulfilling these criteria is TIMSS.

TIMSS (Third International Mathematics and Science Study): The study was conducted in 1994/1995 at 5 grade levels and in more than 40 countries. TIMSS tested more than half a million students in Mathematics and Science and questionnaires were administered to thousands of Teachers and principals. The results were released in 1997/1998.

TIMSS-R (Third International Mathematics and Science Study – Repeat:
The Study was intended to measure the trends in achievement between 1994/1995 and 1998/1999. TIMSS-R was conducted by the International Center at Boston College and included 38 countries of which 26 participated in 1995.

TIMSS 2003 (Trends In Mathematics and Science Study):
TIMSS 2003 builds on the success of TIMSS 1995 and TIMSS 1999 by offering participants of earlier studies the opportunity to study trends in eighth grade Mathematics and Science Achievement at 3 points (1995, 1999, 2003) over an eight-year period.

2. The value and limitations of international comparative studies

The rapid increase in comparative studies (not only the IEA with TIMSS and other studies, but also by the Southern African Consortium for Measuring Quality (SACMEQ), the Monitoring Learning Achievement studies (MLA), and others) in the past few years) indicated that the demand internationally for comparative empirical data is becoming stronger. The question why international studies are carried out is often asked especially consideration the large amount of money that it cost and the possibility that other approaches may provide more cost effective information for educational improvement.

To answer this question it is necessary to look at the recipients that benefit, and how they benefit, from comparative international studies. The beneficiaries of these studies can be categorized into the following broad areas:

- Policy and decision makers worldwide who use baseline data on pupils' achievement to make informed decision on the development of curricula and the organisation and management of schooling (Howie, 2001), (Kellaghan (1996), and Postlethwaite (1999),
- Educators and other stakeholders at school level who learn from what are taught and how is it taught in other countries (Howie, 2001).
- Researchers who benefit from the exposure to the latest developments in research methodology (Beaton *et. al.*, 1999) and capacity building in countries were it does not exist (Ross, 2000).
- The community, especially this generation of parents who have become increasingly involved in the education of their children use the information to compare the quality of education that their children receive in their own country to that of other countries around the world (Howie, 2001).
- Educational institutions for whom the knowledge of the products of the schooling system is essential to be able to prepare appropriate learning programmes and for employers to be able to plan their recruitment and human resources training to meet the challenges of a global economy (Howie, 2001).

Aside from providing useful information as indicated above, education often benefits from international studies because it highlight the plight of education where there is a host of other priority areas, such as poverty, AIDS, etc. The attention of various interest groups is drawn to the outcomes of such large-scale studies, which do not always happen with other research studies, no matter how valuable they are. (Howie, 2001).

Another benefit for countries that participate in international studies is that it forces countries to scrutinize their curriculums (Beaton *et. al.*, 1999). This is necessary since the usefulness of results will be limited if the particular domains of subjects tested do not coincide with the curricula of the particular countries.

The main limitation of international studies should be seen in the fact that policy and educational leaders do not make the maximum use of information obtained (Fiske, 2000). For South African learners to get the maximum benefit from TIMSS 2003 the results should be analysed in such a way that useful feedback is given to stakeholders to enable them to implement it to the maximum benefit of learners and the education system.

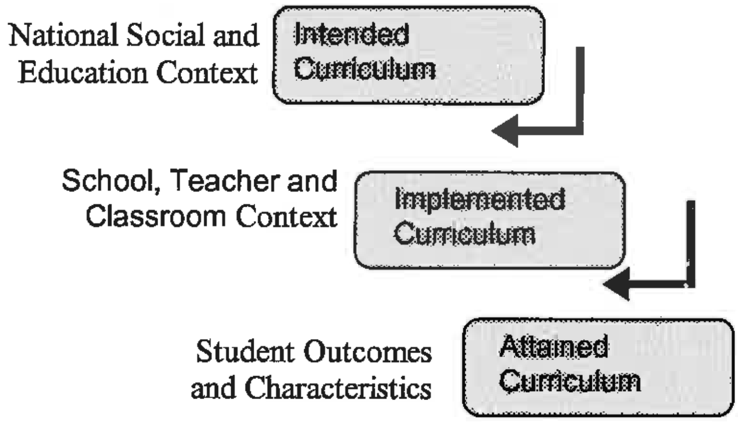
3. Conceptual framework

The central aim of IEA studies is to measure achievement in school subjects in order to learn more about the nature, content and context in which learning occur. Studies of the nature of student achievement, of the characteristics of learners themselves, the curriculum they follow, teaching methods and resources in schools explains the influences of student learning.

The frameworks of previous TIMSS studies focused on curriculum as a broad explanatory factor underlying student achievement. These studies envisaged three "levels" of curriculum, namely what society would like to see taught (the intended curriculum), what is actually taught (the implemented curriculum), and what the learners learn (the attained curriculum). Factors from the educational environment that influence educational decisions, was investigated from the perspective of these three curriculum levels.

TIMSS 2003 also uses these three levels of the curriculum as the basis of its curriculum model (see figure 1).

Figure 1: TIMSS Curriculum Model



The teaching practices of different national education systems compared with student outcomes, form the focus area for TIMMS 2003. Trend data from the TIMSS context questionnaires produce a dynamic picture of changes in the implementation of educational policy and provide a research basis for future national curriculum reform. The assessment frameworks produced for mathematics and science education were therefore designed to capture important issues for mathematics and science education today, while providing the vision necessary to shape future IEA assessment in mathematics and science,

*4. **What does TIMSS test:***

TIMSS allows researchers to:

1. Make comparisons
    - Comparing the performance of students in different countries eg. Gender, home back ground, school organization
2. Seek Explanations
    - Differences in achievement found between different students
    - Enable countries to understand their own education systems better by drawing attention to relative strengths and weaknesses
    - Identifying models and practices in other countries which may provide possible solutions to national problems

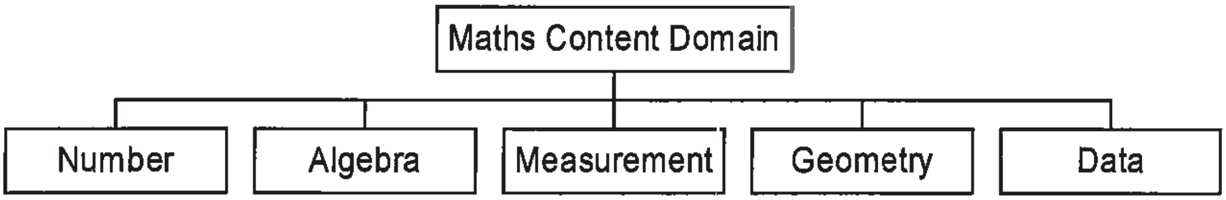
Each subject consists of two dimensions and several domains. The two dimensions and their domains are the foundation of the assessment. Each Domain has several topic areas.

The two domains are:

- Content Domain – It is the specific subject matter covered

- Cognitive Domains – The expected behaviors of students as they engage with subject content.

Mathematics Framework:



a) **Numbers:**
The structure permits the direct mapping of trend items from 1995 and 1999 into the content domains defined for 2003
It includes understanding of counting and number, ways of representing numbers, relationships amoung numbers, and number systems. The number domain consists of whole numbers, fractions, integers and ratio, proportion and percent.

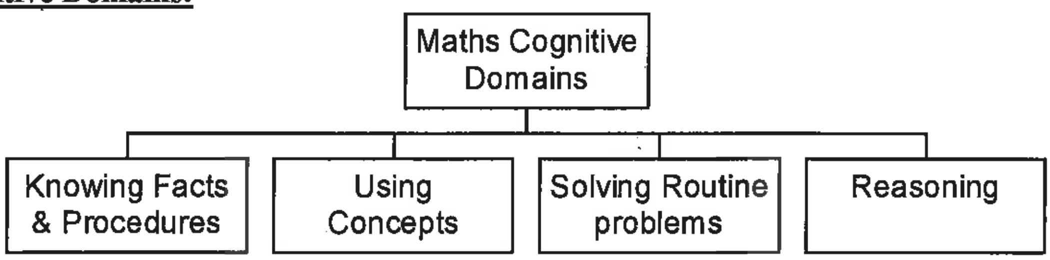
b) **Algebra:**
It includes patterns and relationships amoung quantities, using algebraic expressions to represent mathematical situations, and developing equations and formulas to solve linear equations.

c) **Measurement:**
It involves assigning a numerical value to an attribute of an object. The focus is understanding measurable attributes and demonstrating familiarity with the units and processes used in measuring various attributes

d) **Geometry:**
It includes understanding coordinate representation and using spatial visualization skills to move between two- and three- dimensional shapes. Geometry includes topics like lines and areas, congruence and similarity and symmetry and transformations

e) **Data**
It includes understanding how to collect data, organize data that have been collected by yourself or other, and displaying data in a graph and being able to interpret the results

Cognitive Domains:



Knowing facts and procedures:
Facts encompass the factual knowledge that provides the basic language of mathematics, and the essential mathematical facts and properties that form the foundation for mathematical thought.

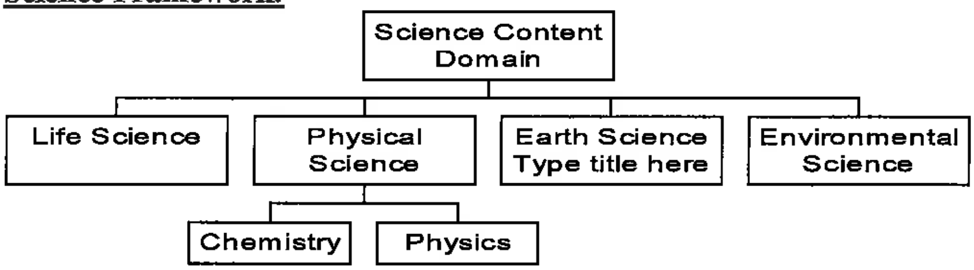
Procedures form a bridge between more basic knowledge and the use of Mathematics for solving routine problems.

Using Concepts:
For a good understanding of Mathematics; it is essential that mathematical concepts are well understood.

Solving Routine Problems:
This tests the student's ability to understand and interpret mathematical questions and use mathematical concepts to find solutions

Reasoning:
It involves the capacity for logical, systematic thinking. It includes intuitive and inductive reasoning based on patterns and regularities that can use to arrive at solutions

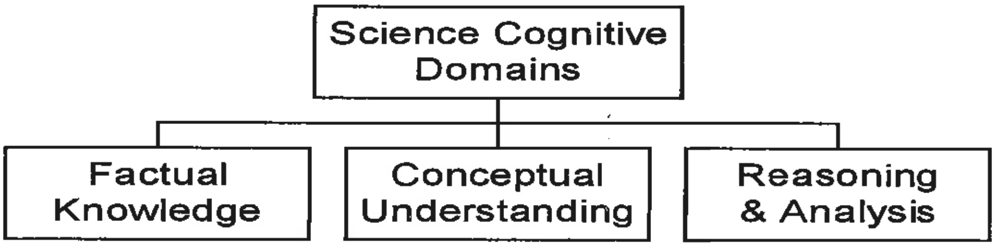
Science Framework:



Content Domain
Life Science:
It includes the understandings of the nature of and function of living organisms, the relationship between them, and their interactions with the environment.

Chemistry:
Students are assessed on their understanding of concepts related to the classification and composition of matter, the structure of matter, the properties and uses of water, acids and bases and chemical change.
Physics:
The students understanding of concepts related to energy and physical processes is assessed with regard to physical states and changes in matter, energy types sources and conversions, heat and temperature, light, etc.
Earth Science:
It is concerned with the study of the earth and it's place in the solar system and the universe. This would include topics like the earth's structure and physical features, the earth's processes, cycles and history, etc.
Environmental Science:
It is a field of applied sciences concerned with environmental and resource issues eg. Changes in population, use and conservation of natural resources and changes in the environment.

Cognitive Domains



Factual Knowledge:
It refers to the student's relevant knowledge base of relevant science facts, information, tools and procedures.

Conceptual Understanding:
Having a grasp of the relationships that explain the behaviour of the physical world and relating the observable to more abstract concepts.

Reasoning and Analysis:
It tests the students to reason, solve problems, develop explanations, draw conclusions, make decisions, and extend their knowledge to new situations

5. Methodological Famework:

Sampling:

The accuracy of the results depends on the quality of the sampling information available therefore it is imperative that all sampling procedures are explicit, practical and well documented.
The selection of valid and efficient samples is crucial to the quality and success of an international comparative study of student achievement. To ensure that our sampling was done efficiently and according to the standards set internationally; we received various resources eg. Manuals, software, training and support from the international project management.
We were required to test students at the grade 8 level, 150 schools were sampled based on certain criteria as set out by the DPC.
Within each sample school, one intact class was randomly selected and all students were tested in Mathematics and Science.The design was expected to yield approximately 7500 students per country.

There were 2 stages of sampling ie. School sampling and then class sampling. In the case of South Africa provision had to be made for the fact that provincial results were required as well. This then also had an impact on the number of schools sampled. Software was provided by the DPC to provide assistance in the sampling process. We used an application called WinW3S which was developed by the DPC. Initial information is entered into the program such as:

1. The name and particulars of the school
2. The number of Grade 8 Mathematics and Science classes per individual school
3. The number of learners within each class
4. The Mathematics and Science Teacher information

The Program then selected a random class sample within each school. As an example; If a particular school had four grade 8 classes. A random sample was drawn thus selecting one of the four classes. If, as an example, 8c was selected as the sampled class. The particular details of that class were entered as well as the Maths and Science Teacher for that particular class.

The sample size per class was not allowed to exceed 50 and in the cases that it did a sub-sample was then done in which the number of students were not allowed to exceed 40. This happened in cases where a particular school only had one grade 8-class consisting of +50 students and in extreme cases 120 students in one class.

6. Instruments and tests

TIMSS use written tests of mathematics and science to measure students' achievement and a series of questionnaires, that focus on context for students' learning in these subjects to gather information about the context in which learning took place.

The TIMSS 2003 research collected three types of data, namely

- data on students achievement,
- data on the curriculum, and
- contextual data from principals, teachers and learners

Achievement instruments

The mathematics and science tests were developed internationally in a collaborative manner. Two different types of questions (multiple-choice questions and constructed-response questions) were included in the pool of TIMSS questions. Although items focus on a particular content element it also assume knowledge or skills from other content areas. From the 2003 assessment onwards TIMSS will gradually place more emphasis on questions and tasks that offer better insight into students' analytical problem solving and inquiry skills and capabilities. To facilitate innovations in assessment instruments, investigative or production based tasks were incorporated in the tests as part of the constructed-response items.

The overriding principle in the construction of achievement tests for TIMSS 2003 was to produce assessment instruments that will generate valid and reliable data. To achieve a valid assessment of the two subjects a substantial number of assessment questions were needed. To accommodate the large number of questions required in the limited testing time available (about 90 minutes per learner) the questions were divided among a set of 12 student booklets. The questions were assigned in such a way to the booklets that the combined responses of individual students provide a comprehensive picture of the achievement of the entire student population.

The items pool was grouped into 28 clusters or blocks, 14 mathematics and 14 science blocks, from which the student booklets were assembled. Each block contains only mathematics or science items. The assessment blocks are distributed across student booklets

- To maximize coverage of the assessment frameworks, and
- To ensure that every student responded to sufficient items to provide reliable measurements of trends in both mathematics and science and

At least some of the blocks need to be paired with others to enable linking amongst booklets.

Each learner completed only one of the 12 student booklets in two sessions of 45 minutes. The 12 booklets were rotated amongst the students in a class so that approximately equal proportions of students in the class responded to each booklet.

Questions types and scoring
The mathematics and science tests were developed internationally in a collaborative manner. The tests contain questions that require students to select appropriate responses (multiple-choice questions) or to solve problems and answer questions (constructed-response questions). Multiple-choice questions can be used to assess any of the behaviors in the cognitive domain. They are however less suitable for assessing students' ability to make more complex interpretations or evaluations because they do not allow for students' explanations or supporting statements. Constructed-response questions are particular well suited for assessing aspects of knowledge and skills that require explanations or interpretation of data. Each constructed-response question has its own scoring guide, developed to provide data about students' achievement as well as diagnostic information about misconceptions and common errors.

Contextual instruments
Curriculum questionnaires
Two questionnaires were designed to collect information from curriculum specialist on national level curriculum plans, issues and policies with respect to mathematics and science curricula. It collects basic information about the organisation of the science and mathematics curriculum in each country and about the content to be covered up to and including the end of grade eight. The National Research Coordinator in each country was responsible for the completion of these questionnaires.

School questionnaires
A school questionnaire was administered to the principal of each sample school. The questionnaire asked questions about enrollment and staffing, recourses available to support mathematics and science instruction, such as the availability of instructional materials and staff, school goals and the role of the principal; instruction time; home-school connections; and school climate. It is designed to take about 30 minutes.

Teacher questionnaires
The teacher questionnaires were designed for the mathematics teachers and for the science teachers that taught the participating class in each school in the sample (intact classes were assessed). These questionnaires were used to collect information on the background, beliefs, attitudes, educational preparations and teaching load, as well as details of the pedagogic approach of teachers. Both the mathematics and science teachers' questionnaires ask questions about characteristics of the class tested; instructional time, materials, and activities for teaching mathematics and science and promoting students' interest in the subjects; use of computers and the internet; assessment practices; and home-school connections. Further questions cover teachers' views on their opportunities for collaboration with other teachers and professional development, questions about the teachers self and their education and training.

Students Questionnaires
TIMSS also developed a questionnaire for students, which included questions on the student's background (home and school lives) as well as their opinions, self-perception and attitudes to mathematics and science. It further asked questions about homework and out of school

activities, computer use, home educational supports, and basic demographics. Included in the pupil and teacher questionnaires were additional questions specific to South African pupils and teachers. These questions focused mainly on issues related to language. The questionnaires required about 30 minutes to complete.

7. *Logistical Arrangements*

The design of TIMSS called for high research standards and quality data. Procedures outlined in TIMSS operational manuals have been designed to ensure that high quality, international comparable data will be available for analysis. The procedures are basically a replica from the data collection activities in the previous cycles of TIMSS with some adjustments to accommodate improvements. If a country deviate from the prescribed procedures without prior approval it runs the risk losing the ability to measure trends properly, or to compare data with other countries participating in the study. Operationally TIMSS represents a considerable challenge for a country such as South Africa.

Adaptation and translation of instruments

An international study sometimes requires adaptations because of different culture and language. Examples of required changed in the case of South Africa is the replacement of the decimal point with a comma; adaptation of American English with South African English (e.g. changing gasoline to petrol); modifying the context that might seem strange to South African students (e.g. changing robin to bird) and the replacement of common English names (e.g. Mary) with more appropriate local names (e.g. Maria or Sesule). Some of the items tested in the field test have been modified and had to be adapted for the main study. All English instruments were first modified completely before instruments were translated in Afrikaans.

Contacting schools

With the approval of the relevant officials, letters were sent to school principals requesting their participation. For each school selected in the sample a first and second replacement school was selected. Many follow up letters and telephone calls were necessary to ensure that all selected schools were informed and that the necessary information were received from them. Of the 265 schools in the sample, 251 of the sampled schools, 14 first and 2 second replacements schools took part in the study. In one case all three schools refuse to be part of the study.

Data collection

The design of TIMSS called for instruments to be sent to schools by post and administered by a staff member (school coordinator) within the school. Facts such as schools without functioning postal addresses, an unreliable postal system (which meant that it would be almost impossible to ensure that the instruments would reach their destinations) and the problem of dishonesty and cheating (which meant that the security and the integrity of the testing process could not be guaranteed), resulted in an external data collection agency (ACNielsen) being employed to collect data from the schools. Data from 255schools (out of the 264 included in the study) were collected. Most of the teachers' and school questionnaires for these schools were collected.

All the fieldworkers were specially trained by HSRC personnel to administer the instruments. Training sessions were conducted in which fieldworkers were given the background of the study so that they know the importance of correct administration procedures and so that they can answer questions at schools. Substantial training time was spent on the different administration forms as it was considered absolutely essential that the administrators were completely familiar with them and that they should be able to complete them correctly. However, despite all the planning, preparation and training, there were still some problems with the administration process (e.g. changes in application dates without informing the moderator and incorrect language information supplied by schools).

To ensure that test and background questionnaires were administered in accordance with prescribed procedures quality control observers made unannounced visits to a sample of schools on the day of testing to observe the test administration.

Coding of constructed-response questions
The use of open-ended items has advantages for research and for analysing how learners approach certain questions. However it has the disadvantages that it has to be scored or coded individually, which are time consuming and costly. The assessment instruments had 151 constructed-response questions that had to be scored. Each of the twelve booklets had between 23 and 43 constructed-response questions, A team of up to 18 teachers was trained in the coding of one booklet at a time, after which they finished all the items in that booklet that required coding. Extensive training with the coders and moderation of the standard of coding on a continuous basis definitely helped to improve the quality of the coding.

Reliability scoring of trend items
The countries that participated in TIMSS 1999 were requested to send their reliability booklets to the IEA Data Processing Center when they had completed the 1999 data collection. The student responses in these booklets have been scanned and made available electronically. By having the 2003 scorers scored the 1999 responses and comparing the results to the scores assigned by the 1999 scorers, it will be possible to estimate the reliability of the scoring of the trend items over time.

Data Cleaning:
We used a software package called WinDem for the cleaning process. There are four basic cleaning steps that we need to complete before sending tha data off to the DPC for further cleaning. These steps are:

- **Unique ID Check**: It checks the uniqueness of the primary key variable ie. Student ID
- **Column Check**: It check that valid codes were entered in the CHECK variabe
- **Validation Check**: Check the validation criterion specified in the Codebook
- **Token Check**: Checks for cases where participation indicator contradicts the data entered.

Besides the capturing of the data; it was required that an extra 25% for each of the datasets be recapured. Thus resulting in two datasets being captured for each questionnaire. The crucial part of the 25% was that the data capturing company had to make sure that these two sets of databases were entered by two different punches.

A final step in the cleaning process was then to compare the 25% captured with the main dataset. One of the rules set by the DPC was to ensure that once these files were compared we had to obtain an error rate of less than 1% if not, then the data had to be repunched.

Quality assurance during data collection

Other studies that can be done

In general TIMSS gets reported with a sets of percentage scores on how learners perform on different sections of work. For example there might be a percentage score for chemistry or fractions. Then there are correlations to explain how different factors influence performance. For example the relationship between teacher qualification or resources could be correlated to achievement scores.

For the study we administered in 2002, we expect low achievement scores in math and science. Nothing much has changed since 1999, when the previous studies were administered. In fact, with the curriculum restructuring process and the implementation of C2005 and an outcomes based education the scores could drop. So what does this very expensive research tell us and how does it extend our knowledge about science and math performance?

In setting up the TIMSS study, we decided to extend the study from the one that was done in 1999. The following indicates some of the other studies that needs to be done:

1. In addition to administering the study to grade 8 learners we administered the study to grade 9 learners. With a band qualification (and grades 7,8,9 forming the senior phase) the specification is what the learner should know at the end of a band. One can say with a greater degree of confidence what outcomes are achieved at the end of a band rather than at the end of a specific grade. By administering at the end of grade 9 we hope for a better indication of math and science competencies achieved at the end of a band.

2. Secondary analysis of the data sets of 1995, 1999 and 2003. TIMSS has been administered three times in South Africa. We will look at trends across the years.

3. Does the format of the question (e.g. MCQs, opened-ended questions) affect performance? There will be an analysis of performance on the MCQs and open-ended questions which test the same concept at the same cognitive level. While we know that in the MCQs with 4 options there is a one in four chance of getting right answer by guessing. This does not exist in the open-ended questions. So an analysis of responses on the different formats would indicate how response is affected by question type.

4. What sense do learners make of questions? There will be follow up on learner responses by in-depth, one-on-one interviews with learners. Student responses on the open-ended questions were sometimes very confusing. Did the learners have a clue what was asked? So with some questions (e.g……) there would be follow up and students would be asked to explain their interpretation of the questions and the meaning of their answers.

5. Interviews on responses on the question of the eclipse. Last year there was the eclipse of the sun and there were massive education and information campaigns regarding this. The eclipse was particularly publicized in Northern Province where the eclipse passed through. [state the question]. And yet students performed very poorly on this question. There would be follow up interviews with learners to determine their sense making.

6. TIMSS is administered in the English and Afrikaans languages. Afrikaans learners write a great deal in the open-ended questions, but what they write does not necessarily make sense. English second language learners write far less, but show the start of some potential understanding. [gerda to add]. Again there would be interviews with learners to look at the meanings they attribute to the questions.

Policy Implications

TIMSS is methodological rigorous and an expensive study to conduct. Having done the study, so what? We hope that since this is the first study since the introduction of C2005, it will act as baseline information reflecting the indicator of science and math scores in the school system. As the new education system settles into place, subsequent studies could be used to reflect our performance.

Another implication is a methodological one. The TIMSS International Centre suggests that TIMSS be administered with 15 year olds at the end of the grade 8 year. In view of the structure of the new curriculum is this the best year to administer the test? We hope that the performance scores at the grade 8 and 9 level would give better information to make that decision for South African students.

Concluding remarks

References

Beaton, A., Poslethwaite, T.N., Ross, K.N., Spearitt, D. and Wolf, R.M. 1999. *The benefits and Limitations of International Educational Achievement Studies.* Paris: IIEP/UNESCO.

Fiske, E.B. 2000. *Status and Trends 2000: Assessing Learning Achievement.* France. UNESCO

Howie, S. 2001. *Mathematics and Science Performance in Grade 8 in South Africa 1998/1999.* Pretoria: HSRC.

Kellaghan, T. 1996. *IEA Studies and Educational Policy. Assessment in Education,* 3(2): 143-160.

Postlethwaite, T.N. 1999. *International Studies of Educational Achievement: Methodological Ossues. Hong Kong*: Comparitive Educational Research Centre, University of Hong Kong.

Ross, K. 2000. Some background information on the South Africa Consortium for Monitoring Educational Quality. UNESCO IIEP Website.