EVALUATING THE IMPACT OF TIMSS-R (1999) IN LOW- AND MIDDLE-INCOME COUNTRIES

An Independent Report on the Value of World Bank Support for an International Survey of Achievement in Mathematics and Science

By Warwick B Elley

Education Consultant, New Zealand

May 2002

EXECUTIVE SUMMARY

This report outlines the impact of the Third International Mathematics and Science Survey-(Repeat) or TIMSS-R (1999), on the education systems of 18 low and middle-income countries which participated in the study. All of these countries received World Bank aid, which was paid to the International Association for the Evaluation of Educational Achievement (IEA), the organizer of the study. Without this funding assistance, most of these countries would not have been able to take part.

In order to obtain information for the evaluation, a four-page questionnaire was sent to the National Research Coordinator (NRC) of each country, and supporting information was gained from World Bank officers and by personal visits to three countries – Romania, Macedonia and Malaysia.

In their responses, NRCs claimed that the international achievement tests used in the survey were generally valid for their national curricula, that the findings were definitely relevant for their country, and that policy makers took the results seriously, particularly the international rankings in achievement. In many cases the results were a surprise, and sometimes a disappointment. Most NRCs were able to point to reforms in Mathematics and Science curricula, in teaching style and in assessment strategies, which were attributable to TIMSS-R. In many cases, it is unlikely that such reforms would have occurred without the galvanizing effect of the results of the survey.

The NRCs also claimed to have developed many new technical skills which were being put to good use within their country, and they praised the quality of the training, and most aspects of the organization of the survey. The survey appears to have made a positive contribution to expanding human capacity in these countries, many of which had no reliable data on their students' achievement levels before. Virtually all countries wished to participate in future studies of IEA, but they claimed that this was not possible without the kind of assistance given by World Bank.

The report concludes by recommending that such aid be continued, and extended to other low and middle-income countries, and that IEA continue to conduct training sessions for NRCs and their colleagues.

Table of Contents

PURPOSE

The purpose of this report is to make an independent evaluation of the impact in 18 countries, of the Third International Mathematics and Science Study of 1999, usually referred to as TIMSS-R or TIMSS (1999). The 18 countries of interest were all low- and middle-income countries which had received financial assistance from World Bank, after a request from the International Association of the Evaluation of Educational Achievement (IEA) which was responsible for initiating the TIMSS-R study. The financial assistance was forwarded to IEA Headquarters in Amsterdam, from World Bank, under its Development Grant Facility (DGF), and was used by IEA to enable these 18 countries to participate in the TIMSS-R project. It was therefore considered important to determine if the grants arranged for these countries had been usefully spent, and that participation in the study had been beneficial for educators and policy makers in those countries. In February 2002, I was contracted by IEA to undertake this evaluation

RATIONALE

Progress in education is often hampered by a lack of sound measures of the outcomes of new policies or programs. (Lockheed & Verspoor, 1991, p.138 ff). In low-income and developing countries, for instance, improvements in the quantity of education provided by governments – in terms of increased enrollments - may well obscure an accompanying decline in quality (Fuller 1991, Greaney 1996). Unfortunately, very few low and middle-income countries can provide sound measures of the quality of the outcomes of their education systems. Few can say what, or how well, the majority of their students are learning, and whether gains are being made over time, as new programs are introduced.

In an effort to create better quality measures of educational outcomes in many nations, and to provide comparative information to policy makers, about the strengths and weaknesses of education in these nations, the International Association for the Evaluation of Educational Achievement (IEA) has conducted a series of cross-national surveys of achievement. Since its inception in 1959, IEA has conducted surveys in Mathematics, Science, Reading-Literacy, Writing, Civics, and Information Technology, amongst others, and much information has been obtained by the participating nations. IEA has grown in strength and expertise, and now has 58 member nations, which contribute to, and take part in its work.

Participation in the surveys conducted by IEA requires an annual fee, for the duration of the study, primarily to meet the costs of the international offices and staff, and of conducting official meetings of the IEA General Assembly and of the National Research Coordinators for the various studies. There are expenses involved in the development and refinement of instruments, of data processing, staff training, and report writing. In addition, country representatives have costs associated with international travel to attend meetings for planning and for training sessions. In most cross-national studies conducted by IEA, up until 1990, very few developing countries took part, as the international costs proved a major barrier. Thus, the benefits of participation were restricted, in the main, to OECD countries, many of which already had relatively sound data on the output of their school systems.

In order to ensure greater involvement of poorer countries, World Bank responded positively to a request by the IEA for funding assistance designed to subsidize the participation of a group of 18 interested countries. Funds for this purpose were made available by the Bank, through its DGF, which had been established to support high-priority global/regional initiatives that could not be funded adequately through the usual country lending operations. The initial grant was US $ 40,000, per annum, for each country, to cover their international fee for each of three years during the project, and these grants were used by IEA to finance the participation of the governments of the 18 countries. Smaller amounts were made available also, to cover the costs of international staff travel, and training in data processing for these countries. The grand total of the funds made available for IEA to cover the international costs of these countries during the years 1998 to 2001 was US $ 2,975,000, which represented approximately 11% of the total costs of the project. An audited account of the details of how it was spent is not yet available, as the training sessions for secondary analysis were still going on in April 2002. However, it should be pointed out that most of these grants were designed to cover the international fee, which was standard for all countries. The 18 countries which received funding support are listed in Appendix III.

According to the questionnaire responses, shown below, it is clear that nearly all of these 18 countries would not have participated without this financial assistance. With the assurance that the grants were forthcoming, however, educators within these countries were in a stronger position to argue for funds, from public and private sources, to cover their local expenses in the project.

PROCEDURE FOR COLLECTING INFORMATION

In collecting information to assess the value of this support, I sent questionnaires to the TIMSS-R National Research Coordinators (NRCs) of each of the 18 nations (See Appendix II for a copy of the questionnaire). I also contacted World Bank officers in many of these countries for their views, and made personal visits to three of the countries – Romania, Macedonia and Malaysia, in order to make a closer case study of the impact of TIMSS-R in their education systems. I spoke also with the Executive Officer of IEA, Dr Wagemaker on several occasions, and with senior officers in the Bank, some of whom assisted with my questionnaires. I also made a personal visit to the IEA headquarters in Amsterdam. Here I was able to sight copies of national reports, of newspaper clippings outlining the media statements when TIMSS-R results were released, and financial statements about the way the IEA projects are funded.

In addition, I studied the various international reports on TIMSS-R, (Martin et al, 2000) and Mullins et al , 2000) and an IEA publication on the impact of the previous TIMSS survey of 1995, outlining the different ways that the findings of that survey had been used in the countries that participated at that time (Robitaille et al., 1998)

In brief, the major questions addressed in this report are as follows:

- How relevant was the TIMSS-R survey to the needs of each county?
- How were the results of the study received in each country?
- What changes in education policy were stimulated by TIMSS-R?
- What new competencies were developed in the participating countries?
- Will these competencies be useful in future surveys of achievement?

A copy of the full Terms of Reference for my investigation is given in Appendix I.

FINDINGS FROM QUESTIONNAIRE SURVEY

An initial letter of introduction was sent to each of the 18 National Research Coordinators (NRCs) on February 27, 2002, outlining the purpose of the study, and seeking their cooperation in completing a questionnaire. On 21 March, my questionnaire was sent to all 18 NRCs, by E-Mail, with a request that it be completed and returned within three weeks. Most NRCs responded within the required time period, but a follow-up message was required in seven cases. Sixteen of the 18 countries had replied by the time this report was compiled, and this section of the report is based on largely on their returns. One NRC had been seconded to another country for six months, and never responded to my requests. I did gain some information about the two missing cases from IEA Headquarters, and have included it, where it was relevant. In a few countries, two or more people completed the questionnaire.

The general findings produced by the questionnaires are given below, under headings which correspond closely to those of the questionnaire, and the major Terms of Reference. More detailed case studies of the three countries I visited are given after these findings. The questionnaire results are presented, in the main, without identifying the particular countries concerned. Some of the questions were open-ended, and some additional information was gained from IEA staff, Bank officers, and newspaper clippings and these too are presented without country identification.

Numbers Participating

Most of the NRCs were senior officers in their Ministry of Education or senior researchers in an associated educational research or examinations unit. Two NRCs were on the staff of a university. Most appeared to be in a position to influence policy making, and to ensure that the findings were given wide publicity.

In half of the 18 countries, there was at least one person who worked full-time on the TIMSS-R project. In the remainder, most had two or three people working at least half-time. In addition, many more were working from time to time. One NRC claimed that 237 people in his country worked on TIMSS-R at some time, either as translators, or coders, or school coordinators, or quality control officers, or data processors, etc. Another NRC counted 174 such people, and several others estimated over 100. Thus, large numbers of research workers and educators in most countries felt a sense of ownership for the project, and had an opportunity to learn about its aims and procedures.

Relevance of the Tests for the National Curriculum

If TIMSS-R was to influence policy, it was important that its tests be seen as relevant to the country's national curricula. While the IEA strategies do allow for democratic decision-making in determining the contents and emphases of the tests and questionnaires used, it is unlikely that any one set of tests could represent the whole curriculum of a diverse set of countries. Therefore NRCs were asked, during the

TIMSS-R project, and also in response to my questionnaire, to express their opinions about the content validity of the Mathematics and Science tests used in TIMSS-R.

The questionnaire responses generally told a positive story. In six cases the test content was judged to reflect the official national syllabus in Mathematics "very well". This meant that they covered over 90% of the national content objectives. Only three NRCs judged the same for Science, however. In nearly every other case the tests were said to match the official syllabus "quite well". In only two countries did the NRCs judge that one of the tests (Science) did not match the syllabus well. The judgements made by NRCs on the parallel question about the actual "implemented" syllabus, were almost identical to those for the official syllabus. One NRC claimed that the tests were not a good match for the implemented syllabus, but the teachers in that country thought that they were! One NRC could not comment on this question.

It is relevant to point out that this question of the validity of the tests was studied systematically by the organizers of TIMSS and TIMSS-R, and tables showing the extent of commonality in curricula, and the extent of matching between test and curriculum in each country are given in the international reports. In Mathematics, for instance, the extent of agreement between national curricula is considerable for most countries, especially in Fractions and Number Sense, (86%) and Measurement (83%), but less so in Algebra (68%), Geometry (67%), and Data Representation (60%).(See, Mullins et al 2000, p.172). In Science the agreement levels are high in Biology and Earth Science, but lower in Chemistry and the Nature of Science. (Martin et al 2000, p.182).

More importantly, the match between actual test and national syllabus was studied by comparing the performance of students on all items in the Mathematics test, with those judged by each NRC to have been taught in the schools of their respective country, at the time of the test. In this analysis, there was virtually no difference between the two totals. In only one of the 18 countries, did the students obtain a mean score in excess of 1% more, on the items judged fair for their students, than they did on the total test of 169 Mathematics items.

Likewise, in Science, 13 of the 18 countries showed virtually the same means on the total test of 153 items and on a test of the items judged fair for their country. In one country, there was a discrepancy of 6 points, while the other four countries showed discrepancies of 2-4 points. A similar pattern was found in the corresponding analyses conducted in the TIMSS (1995) study.

One question not addressed by these analyses is that of test format. In some countries, multiple-choice tests and performance type tests are widely used. In others, they are quite unfamiliar. While some NRCs felt obliged to point out this factor, as a possible distortion, it seems unlikely to have had a major effect. In a study of the TIMSS (1995) results for eight Central and Eastern European countries, (Vari, 1997), those countries where multiple-choice tests were unfamiliar had shown the same patterns of test completion as those who were familiar with them.

In the light of the analyses described above, it is my view that the TIMSS-R tests are best viewed as good measures of general mathematical and scientific ability, rather than tests of specific information. It is not surprising, then, that policy makers in each

country, who studied the evidence, were prepared to accept the findings as valid for their students. Criticism of the content was made by very few.

National Reports

If a survey is to have local impact, there should be an authoritative report that sets out the results for that country's schools, in clear, concise terms, so that policy makers can read and judge the importance of the findings. According to the questionnaire responses, a National Report was prepared in all except one of the countries that received World Bank support.

I have examined the reports of seven of the countries concerned, and am satisfied that they fulfil the requirements of such a report satisfactorily. Most of these reports were not written in English, but with the help of NRCs and others, (and dictionaries) I was able to identify the major contents, and to study the tables and figures. Many of the NRCs also referred to summary statements they had prepared for officials and media representatives to read.

Virtually all NRCs also claimed that senior officials in their country had read the national report, or a summary of it, and in at least 11 cases, the national Minister of Education had also read the report. Understandably, some NRCs were unsure about this question. In one case, there was a new Minister, who was reluctant to accept the negative findings in his country, and who discouraged publication in the media. However, in many of the other countries, the Ministers of Education, referred to the report in public speeches about education. One NRC reported that the Minister announced the major results at the Press Conference while another spoke of briefing his Minister personally so that he could respond to questions in Parliament about the findings. Others Ministers took part in National Seminars on the findings.

Importance of the Report's Findings for Policy Makers

In all the reports I sighted, the international rankings of each country in general Mathematics and Science achievement mean scores were given prominence. And in response to a question about the importance of different aspects of TIMSS-R, this result was consistently judged as very important by policy makers. (See Table 1) It seems that all education officials are interested in where their education system was ranked, relative to those of other countries, and especially other countries of similar socio-economic and cultural characteristics. Newspaper clippings I sighted often highlighted comparisons with the results of specific countries, such as neighbors or trading partners. One national daily featured a cartoon, emphasizing the way that their country was being left behind by its neighbors. Several NRCs referred to the ranking as a good measure of quality. "It tells us how we are doing!"

For most of the low-income countries in this survey, the mean scores were below the international average, in one or both of the Mathematics and Science surveys, and such findings were often surprising, and disappointing. In the absence of any information about the relative performance of a cross-section of students, it seems that national pride encourages people to elevate the importance of minor indicators. Several NRCs mentioned the success of their top students in international Olympiad competitions in the past, but few countries had had access to data about the

achievement levels of a large representative sample of pupils. The net effect of a disappointing finding was to serve as a "wake-up call" in many of the countries. In such a context, policy makers are likely to listen to those who support new ideas for curriculum change and the reform of teaching methods. (See below) In the previous TIMSS study of 1995, one NRC referred to the results as "the most controversial issue in education in the last two decades". (Robitaille et al, 1998) In another case, the President of the country called for a "rescue package" after some disappointingly low rankings, and this was taken up seriously, after the TIMSS-R results confirmed those of 1995. Low performance often does force politicians and education officials to act.

Other topics of interest to policy makers in the TIMSS-R reports were those that related to differences in the school context – the timetables, the language of instruction, the curriculum structures, the examination policies, the national assessments, the amount of homework, the number of hours instruction, and the like. (See Table 1) I was told that at meetings of the NRCs, these matters were often compared, and syllabus statements exchanged, so that new ideas were available for those who wished to take advantage of them.

In the nine countries that took part in the previous TIMSS study, in 1995, policy makers were apparently interested in the reported trends in achievement over time, although a close study of these trends shows that only very small changes had occurred in most cases. Four years is a short period to identify systematic changes across a whole school system.

NRCs in seven countries reported that their policy makers took considerable interest in the reported student attitudes. This finding should be interpreted with caution, in my view, as there was often a negative or zero relationship between achievement levels and expressed attitudes of students. In some cultures, it is difficult to obtain honest expressions of attitude in an official looking questionnaire.

Table 1 sets out the tabulated ratings of the NRCs on the degree of importance attached to each of several features of the National Report in each country.

Table 1
Ratings on the Relative Importance of Topics in the National Report for TIMSS-R

	Of Much Interest	Considerable	Some	None	Omitted
International Ranking	15	0	0	0	1
Trends over Time	6	2	1	6	1
Content Areas	8	3	4	0	1
Student Attitudes	7	4	4	0	1
School Context	10	3	2	0	1

Amongst the "Other" responses of interest, reported for this question, were: Test format, goals of the study, factors affecting performance, questionnaires, and national assessment systems, and the role of research.

In response to a question about the extent to which the findings of TIMSS-R "had contributed to a greater understanding of education" in each country, the NRCs were very positive. In every case, NRCs claimed that the findings of TIMSS-R had contributed "a great deal" to an understanding of achievement levels in their schools, in means and distributions - relative to other countries, and within their own country. Within countries, the findings revealed differences in gender, ethnic group, location, and various school characteristics. For many countries, these issues had not been objectively studied before, on good nation-wide samples. Most NRCs said the survey had contributed a great deal also to an understanding of curricula in their country, and to an understanding of student attitudes. Most NRCs said it helped "A little" in the case of teaching methods and of education resources. "Others" referred to testing methods, importance of empirical research, and trends in achievement levels.

The general impression left by these judgements is that TIMSS-R was especially important in understanding the achievement levels of students, by allowing for analysis of within- country and between-country variations, and also in understanding the differences in curriculum content and structure, relative to those of other countries. Teaching methods were given positive, but lower ratings on this question.

WHAT USE WAS MADE OF THE TIMSS-R FINDINGS?

The following sections outline the extent to which the findings of TIMSS-R were made known to the public and the kinds of uses to which the findings are being put in improving the education of students in each country.

Publication of TIMSS-R Findings

The international results and reports were released to the world on December 5th 2000, at a press conference in Boston, USA. In most of the 18 countries in this review, the findings for their respective education systems were released, by mutual agreement, one or two days later. In many cases, there was a press conference, attended by senior education officials, and media representatives. In a few cases, this press conference was a grand affair, attended by hundreds of people, with speeches by Ministers of Education. On at least one such occasion, the Prime Minister also attended. Often the National Report was distributed at the initial press conference, and the main message was reported in major national newspapers and television. These occasions were not typical of all countries, however.

NRCs were asked in their questionnaire to rate the extent of coverage in the various media and in publications that might reach members of the public, teachers, and other education officials. Eight NRCs claimed that the newspaper coverage of the findings had been "extensive"- which implied front-page treatment, editorial comment, and the like. One national newspaper printed all the released test items, in batches, over several editions, presumably to increase public awareness. Four NRCs said that newspapers gave "some" coverage and three NRCs described it as "little". In a few cases, the findings, provoked large headlines, usually when the results were a shock, or surprisingly good. Six countries reported that there had been "some" parliamentary debate on the findings. There was "some" coverage reported on radio and TV in a few countries. One NRC claimed to have spoken about the study six times on national

television. Apart from these public media, there was considerable publicity given to the results in teachers' newsletters, and in academic journals. NRCs were concerned to focus their attention more on getting the important messages to teachers, without whose cooperation, the proposed changes in curriculum, teaching style and assessments would be much less productive. Thus, in several countries the National Reports were sent to all schools, and two NRCs claimed that they used the TIMSS-R results as the basis for teacher workshops on the need for change. Three countries also reported large conferences on the findings, involving hundreds of educators.

In a follow-up question most NRCs claimed that there had been some public discussion of the findings, in both Mathematics and Science, but this outcome is difficult to quantify. One World Bank officer reported that the results were a "shock" to educators, while another claimed that it "caused alarm bells to ring".

Changes in Curriculum Policy

As many NRCs stressed the importance of TIMSS-R in allowing for a study of the differences in curriculum, and as most countries discovered that they had achievement means below the international average, it is not surprising that the TIMSS-R findings gave rise to a considerable amount of curriculum reform. This was particularly so in countries which were participating for the first time. Thus, eight NRCs claimed that there had already been changes made in the Mathematics curricula, as a direct or indirect result of the TIMSS-R findings. Several other NRCs reported that changes had taken place previously, as a result of the earlier TIMSS (1995) survey, while a few reported that the 1999 results gave further point and purpose to these reforms.

The major kinds of changes reported in Mathematics curricula related to more precise statements of objectives, specific statements of standards, introduction of sections on Transformations, Data Representation and Probability, more contact time each week and relocating into Grade 8, topics that had been taught later. Others referred to more real-life problems, or stressed the importance of "application". One country staged a large workshop in which teachers were shown many examples of the kind of "higher order thinking" that TIMSS-R tests assessed. Participants were then required to develop new examples of the same kind, in an apparent effort to help them revise their strategies for questioning students in class. One NRC claimed that many of the teachers still regarded Mathematics as a set of rules and algorithms to master, rather than a course in quantitative problem solving.

In regard to Science, ten NRCs reported changes in curriculum, following the release of the TIMSS-R findings. Most of these changes related to more emphasis on practical investigations, more time in the timetable, re-location of topics, and more stress on 'The Nature of Science and Scientific Inquiry'. This topic had received little or no attention in several countries in the past. In some countries, Science is taught without laboratory investigations, often because of the cost involved. TIMSS-R findings induced some to redouble their efforts to overcome such a barrier.

In at least two cases, NRCs reported that new Teachers' Guides and new textbooks for students, in both Mathematics and Science, were being written, or proposed. In a few cases, new texts were already being used in the schools, following changes which occurred after the release of the 1995 TIMSS survey. I was able to compare the new

with the old in two countries and was reassured that the revisions were educationally sound and likely to lead to greater student interest and learning.

In response to a question about likely future changes in curriculum, 12 of the NRCs reported that they were either definite, or likely to occur. It should be kept in mind that this report is being prepared only 15 months after the results of TIMSS-R were first released. It is encouraging, also, to note that in several countries, the Ministry resolved to set up task forces or special panels, to study the policies of particular countries where performance levels were high, before any changes were made in their own national policies. After all, rapid changes in policy are rare in education systems, and often with good reason.

One advantage of the fact that most NRCs held senior positions in their countries was that they were able to exert some influence on the reform process. Thus, 11 of the NRCs reported that they were directly involved in curriculum revision teams, and were thus able to bring to bear the knowledge they had gained from studying curricula in other participating countries, and from analyzing the strengths and weaknesses of their own students. In this connection, the international reports had spelled out the relative performance of students in each country in each sub-test and content area of the curriculum, and these profiles were highlighted in several of the reports I read.

Other Changes in Policy

Apart from the actual and proposed changes in curriculum, many NRCs reported a number of other policy changes. Eight NRCs claimed that there would be changes in teaching methods, following the TIMSS-R findings. Most of these changes referred to a greater emphasis on activity based learning – more problem-solving in Mathematics, and more practical investigations, and more critical thinking in Science. These are important reforms, in my view, which would not have easily been supported without an external stimulus such as an international ranking can provoke.

Such reforms also imply other changes in the pre-service and in-service training of teachers. In at least two countries, there has been a concerted nation-wide effort on the part of the TIMSS-R research team, to re-train teachers of Mathematics and Science, in Grades 7-9, and in each case the TIMSS-R scores were used as a reason for change. Reformers are stressing the importance of reducing the amount of teacher lecturing, and increasing student engagement in lessons, with more discussion, questioning, experiments in class, critical thinking, valuing of student opinion, and building on students' natural curiosity.

Further changes, provoked or stimulated by TIMSS-R, were those relating to assessment. For several of the NRCs, the assessment strategies used in TIMSS-R were apparently quite unfamiliar. Eleven of the NRCs claimed that there had been change in assessment policies – new examinations, surveys of representative samples- and in assessment methods – more diverse formats, performance type items, quality control procedures, model answers, and deliberate attempts to relate specific items to content objectives. These changes, if well instituted, could well have significant, positive effects on student learning in the future.

Two NRCs were so impressed with this new knowledge about assessment that they were sharing their expertise with those of neighboring countries that did not participate in TIMSS, while others were planning to conduct regional surveys, amongst countries with similar backgrounds, using IEA methodologies. These developments seem most desirable, to this reviewer.

Another NRC spoke of the changed outlook on the role of educational research following the TIMSS findings. In a situation where educational research had been long neglected, the study had been widely quoted in policy documents. TIMSS had once again "put research on the education agenda…"

Thus, there is considerable evidence that TIMSS-R methods and findings did affect policy in most of the countries that participated. It is noticeable that the impact was greater in those countries that were participating for the first time, but it was found in some respects in virtually all countries in this review. The impact was often different, depending on the context, but it was found in every country. These conclusions about definite impact are confirmed by my experiences in the case studies below, by the kinds of impact described following the earlier TIMSS study (Robitaille, 1998) and by World Bank officers that I consulted. Several Bank officers commented on the usefulness of the TIMSS-R results in "focussing dialogue with the Minister and education officials" and another spoke of its considerable value in preparing for a major World Bank project in education.

Human Resource Development

One objective of any educational aid project is to develop the capacity of educators to extend their expertise, so that they can carry out useful projects on their own, without external assistance, in the future. What evidence is there that the officials involved in the 18 countries who received aid from World Bank, were able to benefit from the experience they gained in TIMSS-R? Could they proceed in future with greater skill and confidence, as a result of the special and on-the-job training that was designed to develop in them the various competencies required to conduct national and international surveys of this sort? There are good reasons to be positive in this respect.

In the course of the TIMSS-R project, training sessions were held for NRCs and members of their team in such technical matters as sampling, test development, coding, data processing and the like, and most NRCs claimed to benefit a great deal from these sessions. Table 2 presents the ratings of the NRCs on each of the major topics.

Table 2

Ratings made by NRCs of the special or on-the-job training during TIMSS-R

	Very Helpful	*Helpful*	*Not Helpful*
Sampling	10	5	1
Test Development	11	3	2
Questionnaire Devpt.	9	4	3
Data Management	12	3	1
Quality Control	7	8	1

It is clear from this table that most NRCs claimed to have learned much from these training experiences. In one case, the NRC claimed that he had learned little, not because the training was inadequate, but because he or his team had already acquired this expertise in earlier IEA studies.

A follow-up question, seeking information on the specific skills learned during the project confirmed the positive judgements that the training was effective and appreciated. Most NRCs listed many skills, that they and/or their colleagues acquired, including those given in the table above, and others such as planning, coding, report writing, statistical analysis, secondary analysis, working with large data sets, etc. It seems that the extra effort that IEA staff put into this kind of training was helpful for many of the countries that participated.

Whether this new-found expertise would be put to good use in future assessment surveys, depends, to some extent, on whether the officials who acquired the expertise were still employed in this kind of activity. Once again, the evidence is positive in this respect. All NRCs reported that they were still working in the field of National Assessments. In fact, most of them, or their colleagues, were participating in further IEA studies, such as TIMSS (2003) and/or the Progress in Reading Literacy Study (PIRLS). A few had been working on new national examinations, using the methods and strategies they acquired while working on TIMSS-R. Two NRCs reported plans to conduct regional surveys, and many were working on national assessments. It seems that participation has given a new stimulus to evaluating student learning in these countries, in a way that should well be beneficial to policy makers in future.

In 13 cases, NRCs reported that one or two of their "team" had changed roles, but in every case except four these people were still using their expertise in assessment research "often".

Further National Assessments

In reply to a question about their plans to participate in future national and international assessments, the NRCs made a strong positive response. All except two reported that they planned to take part in TIMSS (2003), if funds were available, and many listed other proposed surveys, to be conducted by IEA or OECD. All except five countries replied that they planned also to conduct national assessments in the form of nation-wide or small-sample surveys, while others spoke of their participation in reforming national examinations. In all such cases, it is likely that the lessons learned while working on TIMSS-R would be helpful in improving the quality of the assessments made, and the interpretations which flow from them. Several NRCs spoke of their hope to gain feedback too, on the impact of the various curricular and pedagogical reforms they were instituting, following the TIMSS surveys.

Organization of TIMSS-R

NRCs were asked to rate the way in which IEA had organized and administered the TIMSS-R study. International studies, which involve people from diverse languages, cultures and traditions of research, are not easy to conduct, especially when all are affected by the kinds of decisions made about instruments and time schedules. These

matters often require consensus decisions. Moreover, the difficulty of communication, in various languages, and meeting deadlines, can put strains on the participants which are often greater than found in smaller, local surveys. Nevertheless, most NRCs were clearly very positive about their experience, and very few had any criticisms to make. One NRC was unhappy about the tight deadlines, while another mentioned the problems experienced by non-English speaking NRCs in large forums, and the lack of documentation for some decisions, but these were exceptions. Table 3 lists the various aspects of the project, and the way in which NRCs responded.

Table 3
NRC's Ratings of the Organization of TIMSS-R

	Excellent	*Very Good*	*Adequate*	*Not Adequate*
Leadership	10	4	2	0
Decision-Making	7	4	5	0
Communication	10	3	3	0
Deadlines	8	3	4	1
Data Processing	10	4	2	0

Table 3 shows that 45 of the 80 judgements made were "Excellent" and 18 more were "Very Good". The fact that most countries have signed up for another round of TIMSS confirms that the NRCs and those who pay the local costs, believe that IEA's efforts are well worth supporting. As one World Bank paper reports, TIMSS (1995) "received unprecedented visibility and international acclaim for its technical competence, and value to educational policy makers". (World Bank, 1998) Such a positive judgement appears to be supported by the NRCs with respect to TIMSS-R as well.

Reliance on World Bank Assistance

NRCs were asked to estimate the percentage of local and international expenses that were covered by their own government, and by the funding support received from World Bank, via the IEA headquarters. As outlined earlier, World Bank responded to a request by IEA for a bulk grant to support the participation of all these low and middle-income countries. The fund was requested to cover the international fee, plus assistance in travelling to international meetings and workshops for training. Local expenses, relating to staff salaries, printing of tests, administration, scoring, and coding of results, overheads, etc, were included in local costs. The Bank contribution covered approximately 11% of the total expenses of the project.

In almost every country, the responses from NRCs indicated that all the local costs were met by the national government, (or in exceptional cases, by some voluntary or philanthropic source within the country). In three cases, a small percentage of local costs were also provided by World Bank, through its resident office.

In the case of international costs, NRCs reported, in every case, that 100% of the international costs were covered by the World Bank grants to IEA.

Furthermore, all NRCs except one, expressed the view that their country could not have participated without the World Bank assistance, and they were just as consistent in claiming that they could not participate in such a survey in the future without a similar grant. The one exception appeared to be due to a cautious attitude, attributable more to lack of knowledge, than to a wish to be independent.

In the few countries that I visited or which I am familiar with, it was clearly a major struggle to raise enough money just to meet the local costs. Thus it is not surprising to learn that all except one of these low-income countries would wish the assistance to continue. When funds are scarce, in a low-income country, it is usually difficult for educational research to be given high priority.

CASE STUDY NO.1: ROMANIA

I visited Romania from 4th to 6th April, in order to meet the National Research Coordinator of the TIMSS study and her team in the Center for Evaluation, in Bucharest, and to study the way that the study was received in that country. The NRC, Gabriela Noveanu, is a Senior Researcher within the Center, which is a unit within the Institute for Educational Sciences. Gabriella had been the NRC for her country in the TIMSS (1995) study, also, so the experience of conducting a national survey was not as unfamiliar for her as it was to some of the other NRCs in countries that received World Bank assistance.

The Institute for Educational Sciences is a branch of the Ministry of Education, but is not well resourced to carry out research and development activities. Gabriella had to seek outside funding to cover the local costs of participating in TIMSS studies, and was successful in getting a sports organization -"Infosmart" - to fund the cost of printing the required 4000 test papers and questionnaires for TIMSS-R. Much of the sorting and stapling of papers was done by means of voluntary help, and Gabriella's husband, an IT specialist, performed much of the data processing for the studies, over and above his other work, outside the Ministry of Education.

Gabriella was able to produce good quality reports of both TIMSS surveys, but it was not until February 2002 that she was able to secure a small fund to have them printed. After visiting her office, and speaking to her colleagues, it was clear to me that she worked under very penurious conditions. Without World Bank aid, it was extremely unlikely that Romania could have taken part in TIMSS-R.

In addition to the NRC, Gabriella, (who specializes in Mathematics) and her husband Dragos, (who was responsible for data entry and analysis), I spoke to several other members of the Center who assisted Gabriella in the Science aspects of the study, in translations, editing, and logistics. I had interviews with senior officials of the Ministry of Education, including Professor Ioan Neascu, the Director General for Evaluation, and with Liliana Protesa, the Director-General of Primary and Secondary Education in Romania. Everyone I spoke to responded well to my questions, and consistently praised the importance and quality of the study.

Relevance of the Study

In the opinion of the NRC, the final tests used in the international study matched the content objectives of the Romanian Grade 8 curriculum "quite well" in both Mathematics and Science. The official curriculum was not significantly different from the curriculum that was implemented in the classroom, according to the officials I spoke to.

As reported in the International Report in Mathematics (Martin et al. 2000, p 172), the topics included in the Mathematics tests were an excellent match in Fractions, Number, Algebra and Geometry (100%), but only 20% of the topics in Probability and Statistics were covered by at least 90% of students in Grade 8. In Science, the test topics matched the curriculum very well in Biology, Chemistry, the Environment, and Earth Science (over 90%), but less well in Physics (70%) and the Nature of Science (33%). The NRC considered the test was fair for Romanian students, and when items which were deemed to be less relevant were dropped from the international tests, the effects on the overall mean scores were negligible. In Science, the mean improved by 1% (Martin et al, 2000) and in Mathematics there was no change. (Mullins et al, 2000. p. 350) Although some commentators did complain that the test format was often unfamiliar to students, a study of the completion rates in the TIMSS survey of 1995, had shown that this was probably not a serious factor. (Vari, 1997)

Release of the Findings

The TIMSS-R findings for Romania were given only limited publicity at the time of their international release. Indeed, the Minister of Education actually discouraged much media attention. Ministers of Education have changed frequently in Romania recently, and the current Minister knew little of the study, which was initiated under another Minister. Apparently the findings of the first survey, TIMMS 1995 were given more media attention. Nevertheless, the main results of TIMSS-R were presented in the newspapers and on the radio in December 2000.

Now that the national reports are published and printed, a greater publicity campaign is planned. The reports will be distributed to all secondary schools, and the teachers of Mathematics and Science at Grades 7 and 8 will be targeted.

The international rankings of Romania in both Mathematics and Science were an unpleasant surprise to most educators, as they fell below the international average in most areas. Some described it as "a shock". The findings were surprising because the past success of Romania's top students in Olympiad competitions left educators and public with the impression that Romanian standards of achievement were relatively high. Indeed, the Minister of Education, and many senior officials were at first skeptical of the results. However, the fact that both TIMSS surveys produced similar rankings for Romania, has removed some of that skepticism. The senior officials I spoke to were prepared to believe the results, and to see the importance of acting upon them. Some important changes have already taken place.

Impact of the Study in Romania

Curriculum revision in Mathematics and Science in Romania began in earnest in 1997, after the first TIMSS results had been announced. The first revisions occurred in the lower grades and have now been extended upwards. The results from TIMSS-R have given further point and purpose to the changes. In Romania, it is the responsibility of the Bureau of Educational Development to initiate curriculum revision and the NRC, Gabriella, and two other members of her team had an important role, in developing curriculum guides in Mathematics and Science. Gabriella also plays a key role in writing textbooks. She prepares these on her own initiative, in her own time, and her books are then placed in competition with those of other writers. Perhaps it is fortunate that she has been very successful in having her texts accepted, and used in the majority of schools. Perhaps it also reflects the fact that those selecting the texts can see the importance of moving towards the kind of teaching styles and topic presentations that are reflected in the textbooks.

I studied examples of the curriculum guides and the new textbooks, while I was in Bucharest, and could clearly see the impact of the TIMSS studies. The Curriculum guides referred to the TIMSS findings and presented examples of new test items, following the TIMSS models. The textbooks are attractive and present many examples of practical problems for children to tackle. The TIMSS findings suggest that Romanian children have not experienced this kind of emphasis in the past. In her own words, she gained much from her attendance at IEA meetings to help with these matters. She interacts with Mathematics specialists from other countries, exchanges curricula and other resources, and gains many new ideas for improving practices in Romania. Teaching styles in Romania tended to be rather theoretical, rather than practical and the senior officials I spoke to expressed a wish to change this emphasis in all schools.

Other changes have occurred in the time devoted to Mathematics and Science, in the sequence of topics, and in the relative emphasis given to the Nature of Science, and to experimentation in science, and to the topic Statistics and Probability in Mathematics.

Gabriella and her colleagues in the Center for Evaluation have been conducting seminars for teachers, using the TIMSS results as a starting point and reason for change. Thus, there has been enough desire for reform within the Ministry to start changes in curricula, and new textbooks, and there is a strong move to push these changes from the bottom up. It seems unlikely that these changes would have occurred without a “wake-up” call of the sort that the TIMSS projects produced.

Another area in which the impact can be seen clearly is in assessment policies. The use of a variety of item types, particularly multiple choice and performance types - and the links shown between mental processes and test outcomes has been emphasized in seminars with teachers. These matters were new to the majority of teachers, and are seen as a major step forward by the policy makers I spoke to in the Ministry of Education, as well as those in the Center for Evaluation.

During the TIMSS projects, training was provided by IEA for NRCs. This training was given very high ratings by the NRC for Romania, who claimed to have benefited from all aspects of the technology of conducting national assessments. In her own

words, she and her colleagues could not have carried out a national assessment without the training and experience they gained in the TIMSS projects.

All of the team who took part in TIMSS-R is still involved in making national assessments, so the training is being put to continued good use. The projects have clearly had a positive impact on the development of human capability in Romania.

Conclusion on Romania

The TIMSS studies have provided a clear baseline of educational quality against which progress can be measured, as the new curricula and textbooks are introduced in schools. The results of the surveys have had a definite impact on curricula in both Mathematics and Science, and in test format and in-service training of teachers. New skills have been acquired by those in the Center for Evaluation, and these skills are being put to good use in the next round of IEA surveys. Although it was clearly a major challenge for Romania to find the funds to participate, there was a clear message from the Ministry officials that Romania must continue to take part in these studies. There was a widespread consensus that the study was "extremely important" for Romania.

CASE STUDY NO 2: MACEDONIA

I visited Skopje on 8th and 9th April 1992, in order to interview the Macedonian NRC for TIMMS-R, and to meet with other education policy makers, with a view to gaining first hand feedback on the impact of the study in Macedonia. The NRC, Anica Aleksova, is the Senior Adviser for Mathematics in the Bureau of Educational Development, and the specialist in Mathematics in the Assessment Unit. These institutions are part of the state Ministry of Education. I spoke with the NRC several times, and she was very helpful in answering my questions. I had useful interviews also with Bojana Naceva, the Head of the Assessment Unit, with Mr Vladimir Mostrov, the founding Head of the Unit, and now an independent consultant, and with the Data Processor, the Desktop Specialist, and several other members of the Assessment Unit, who made some contribution to the TIMSS-R study.

TIMSS-R (1999) was the first international project in which Macedonia had participated. Thus, there was no possibility of identifying trends in student achievement levels over time. As in Romania, the Macedonian NRC conducted most of the TIMSS-R project on a shoestring, while engaged in other duties for her Bureau. According to the economic indicators collected as background data for the TIMSS-R project Macedonia is one of the two poorest countries in the survey. (Martin et al. 2000, p.25) The Government did support TIMSS-R, but was unable to provide more than about 30% of the local costs. Thus, World Bank provided 100% of the international costs, through its DGF, and about 70% of the local costs, through its local office, after an approach by the NRC to this office in Skopje.

Personnel

The NRC, Anica Aleksova, officially worked full-time on the project, although she did perform other duties as well, throughout the four years of the project. She was a member of a working party responsible for the national curriculum revisions in Mathematics, which took place from 1996 to the present; she contributed to the new National Assessments in Grade 4, and she played a large part in developing the new "Matura" Examination at Grade 12 level. As the Senior Adviser in Mathematics, within the Assessment Unit, she also had advisory functions to perform, and participated in a number of in-service courses for teachers. Meanwhile, she has been one of a group of three authors writing textbooks for the new National Curriculum, at several grade levels. Her textbooks are used by the majority of the students in the first six grades throughout Macedonian schools. She is currently involved in writing textbooks for Grades 7 and 8.

Anica received assistance in TIMSS-R, from time to time, from two other colleagues in the Assessment Unit, especially in sampling, data management, the translation and printing of the international tests for the pilot and main surveys, and the administration and scoring of the international tests. These officers worked about 50% of their time on the project. In addition, the NRC estimated that 174 other people had been involved in some small way, as School Coordinators, Test Administrators, Scorers, Translators, and Editors. Thus many educators had an opportunity to learn at first hand about the project, and to feel a sense of ownership, when the reports were released.

Relevance of the Tests

In the opinion of the NRC, the final versions of the TIMSS-R tests matched the old curriculum in Mathematics and Science "quite well", between 75 and 90% of the curricular objectives, overall. However the tests are expected to be a better match in TIMSS-R 2003 – "over 90% of the objectives"- once the new curriculum is properly completed in Grades 7 and 8, within the next two or three years. The revisions which are taking place at Grades 7 and 8 will bring the curriculum much closer to common international structures and content.

A more detailed description of the match between curriculum and test, shows that the tested topics in the "old" curriculum were covered well in Biology, Physics and Chemistry (over 85%), less well in Environmental Topics (67%), and poorly in Earth Science (25%) and the Nature of Science (0%) (Martin et al, 2001, p. 182). In Mathematics, most of the tested topics were expected to be taught in Algebra and Geometry, but Fractions, Number and Measurement were only partially covered (about 60%), and not at all in Probability and Statistics (0%).(Mullis, et al, 2000, p.188). This pattern is similar to that found in many central and east European countries. (Vari, 1997)

The International Report on TIMSS-R for Mathematics reveals that Macedonian Grade 8 students averaged 40% on all the169 items included in the Mathematics tests, and 41% (i.e. 1% more) on those 134 items identified as best fitted to their curriculum in Mathematics, at the time of the survey. The corresponding figures for the Science

tests were 44% and 48% respectively. Although many of the test items were apparently unfamiliar in format to the students, the NRC believed that the tests were fair, and sound measures of the abilities of a representative sample of Macedonian students at Grade 8. When items were omitted because they had not been taught, the results were scarcely affected at all. Apparently the tests are good measures of general ability in Science and Mathematics.

Release of the Results

Macedonian students scored below average in both Mathematics and Science. Their mean scores placed them at 30th out of the 38 countries that participated in Mathematics, and 29th in Science. These rankings were a real and unpleasant surprise to most educators, as their elite students had performed with success and distinction in international Olympiad competitions in the past, and there was a general belief that the products of their education system were of a high quality. Macedonian teachers considered they were well-qualified and confident in their ability to teach their subjects at Grade 8 level, according to their responses on the teacher questionnaires used in the study. As the results of TIMSS-R were the first of their kind, in a cross-sectional survey in many countries, it was the first time that there was any sound evidence available to test this general belief.

Five National Reports of the findings were produced by the NRC and her colleagues, and have been distributed to all 350 schools in the country. The reports covered results for the separate sciences and for Mathematics, as well as sample questions which were released after the survey. I have studied the reports, (with some help in understanding the language, from the NRC). As far as I can tell, they are very well prepared, clear and concise, with an abundance of practical examples to help teachers. Each report is about 50 to 70 pages in length. At least 550 copies of each report were printed by the Ministry.

At the time of the release, the national newspapers, TV and radio gave the findings "considerable coverage". The NRC arranged three meetings with the media, and appeared on national television about six times, to discuss the project, so the general public had ample opportunity to learn about its findings. One of these presentations took place during my visit to Macedonia.

A publication for teachers also carried an article about the project, and over the past two years, the NRC and a group of Ministry officials she trained, have conducted a series of seminars on the project, throughout the country, for Grade 5-8 teachers of Mathematics and Science subjects. All such teachers in the country had an opportunity to attend one of these seminars. No academic journals have yet published anything about TIMSS-R, but the NRC did make a presentation to staff at the University of Skopje. This meeting was apparently well-attended, and gave rise to lively discussion about the implications of the findings.

Impact of the Survey in Macedonia

There is no doubt that TIMSS-R had a considerable impact in Macedonia. Changes have already been made in curricula, national assessments, and in-service training. Senior officials in the Bureau and the Ministry are familiar with the reports, and the

Minister of Education has several times referred to it in his speeches and press conferences. Indeed, he and the Director of the Bureau of Education, both ex-physics teachers, have been cooperating in a campaign to change the format of science teaching in Macedonia, from a set of separate disciplines to one integrated subject. Their motivation to bring this about was apparently stimulated by the TIMSS-R project. There is concerted opposition, however, from staff at the University of Skopje.

Apart from the concern expressed over their international rankings, the policy makers were interested in the attitudes of students, which were relatively positive in both Science and Mathematics. However, a study I made of the pattern of results of students who claimed to like Science a lot, gave me some cause for skepticism about the Macedonian students' responses. Those students who claimed to be most positive towards science had much lower average scores (464) than those who were negative (484). This pattern was unusual, to say the least.

Policy makers were also interested in the relative strengths and weaknesses of the sub-tests in Science. Macedonian students were strong in Chemistry, but understandably weaker in Earth Science and the Nature of Science, which received little attention in schools. This profile clearly reflects the relative emphasis given to the topics in the National Curriculum. As it was the first time that Macedonia had participated in an international study, the NRC (and her colleagues who were involved in curriculum revision), were also very interested in studying the curricula, resources and teaching methods of her counterparts in other participating countries. These were frequently exchanged at meetings of NRCs, which took place about two or three times a year throughout the project.

In Macedonia, new topics are being added to the curriculum at Grade 8, in both Mathematics and Science. In in-service training, which is the responsibility of advisers in the Bureau of Development, like Anica, the findings of TIMSS-R are used as a lever, to get teachers to take seriously, the need to change their teaching styles from a predominantly lecturing style, to one of interaction, in which students are involved in many practical activities. The TIMSS-R reports showed that Macedonian teachers are accustomed to much lecturing, and very little student activity.

The textbooks which Anica has produced, and which I was able to study, reflect these changes admirably. They are much more interesting, colorful and interactive than those of a decade ago, which, paradoxically, were written by the same people. The experiences gained in TIMSS-R appear to have changed their views considerably.

In assessment policy, the impact of TIMSS-R has also been extensive. A series of national assessments has started in Grade 4, based on TIMSS-R methodology, and using the newly developed skills of the staff in sampling, test development, questionnaire development, test administration, quality control, objective scoring, data entry and analysis, and report writing. The staff of the Assessment Unit initiated this policy, in 2000, with a sample survey of student performance in Mathematics and Mother Tongue. Other subjects will be included, and the cycle will be repeated every four years, as is the case with TIMSS.

At Grade 12 level, a new version of the national "Matura" examination has been developed, again using TIMSS-R methodology, and is being trialed at present. The policy was introduced to avoid the clumsy situation which exists at present, where all schools set their own exit examinations. Because the results are unstandardized, the universities make their own assessments later, before selecting students.

Interestingly, the NRCs for TIMSS-R and for PIRLS have recently traveled to Bosnia, which is not a member of IEA, to conduct training sessions with their educational specialists, so that they can conduct national assessments in their school system. The Macedonians were able to use their new expertise to benefit others.

Human Capability Development

During the TIMSS-R project, the NRC, Anica Aleksova, and the Data Processor, Beti Lozanoska, attended various training sessions designed by IEA, for those who were unfamiliar with procedures. They also gained indirectly, through discussion with other NRCs at meetings, and through their participation in a project which they perceived as critical for their own education system. Anica learned much about the technical issues entailed in sampling, test development, questionnaire development, quality controls, and report writing. In all these respects, she rated the training as "very helpful". Beti had similar judgements to make about data cleaning, data entry and data processing. She uses the IEA manuals to good effect in other projects in the Assessment Unit.

All those who were involved in TIMSS-R are still working in the field of assessment, in TIMSS (2003), in PIRLS, in Grade 4 National Assessments, in PISA Plus, and the Matura Examination. The training they received in TIMSS-R has also been of assistance in helping them to restructure their curriculum from an outline of knowledge to be attained, to a set of descriptive statements about the objectives and skills to be developed in all students at different stages of the school system. The hope is that these various changes will have a positive impact on student learning. However, the fact that over half the teachers in the schools are over 50 years of age, suggests that the impact may be slow, in some respects.

Conclusion on Macedonia

The impact of TIMSS-R for education in Macedonia was considerable. The results provided a surprise, and now represent a baseline for comparisons over time, as new programs are introduced. The results also pointed the direction for reforms in curricula and teaching styles, and there has been a concerted effort by those in the Assessment Unit to pass on these messages to teachers. The models of test development are being put to good use within the system already, with new assessment policies in Grades 4 and Grade 12. It is clear that the impact in Macedonia has been positive, and the policy makers I spoke to wish to continue participation with IEA projects.

CASE STUDY NO. 3: MALAYSIA

I visited Kuala Lumpur, form May 6th to 8th, 2002, in order to interview the Malaysian NRC, Dr Azmi Zakaria and to learn about the impact of TIMSS-R on the education system of his country. Unfortunately, all of the other members of the TIMSS team

were out of Kuala Lumpur, working on the pilot test for TIMSS 2003. However, I had a fruitful discussion with Dr Zakaria, and was able to study a number of relevant newspaper clippings, the National Report on TIMSS-R, and another report prepared by the TIMSS team, on the proceedings of a national seminar which was conducted in Kuala Lumpur on the findings of the survey.

Malaysia represented a somewhat different stage of economic and educational development from that of the other case studies. The national GNP per capita (PPP) in US dollars was much higher ($7730), and the Grade 8 achievement means in both Mathematics and Science were also higher – both being above the international average. The level of sophistication in assessment strategies amongst staff also appeared to be much greater, as several members of the TIMSS-R team had studied measurement at Educational Testing Service in USA.

Personnel

As Director of Planning and Research since 1997, Dr Zakaria holds a senior position in the Malaysian Ministry of Education and he fills many roles besides that of NRC for TIMSS. For instance, his Division was recently responsible for preparing the 10-Year Blueprint for Education in Malaysia, and during my visit he was chairing a 2-day conference of parliamentary ministers on the relationships between education and unemployment in Malaysia. Although he held the position of NRC, he worked less than 50% of his time on the project. His Assistant Director actually worked more hours (Over 50%) on TIMSS-R. They also called on experts from other departments to help with matters of translation, curriculum, data management and the like. The 10 members of the Planning and Research Division worked on the project part-time, and many more (over 100) contributed from time to time.

Relevance of TIMSS-R for Educational Policy

Malaysia was participating in an IEA study for the first time, and they actually entered the project after it had commenced, so Malaysian educators had little control over the contents of the TIMSS-R tests. Nevertheless, they judged the mathematics test to have matched their syllabus objectives “Very well” (over 90% of objectives covered). For Science the test matched the syllabus “Quite well”. (75-90%). As some teachers in rural areas were believed to be less conscientious about covering the official syllabus, the NRC judged that the tests matched the “implemented” syllabus “quite well” in both Mathematics and Science.

In the curriculum analysis reported in the international report for mathematics, (Mullis, et al. 2000, p.172), the extent of coverage for 90% of the Malaysian Grade 8 students was 94% in Fractions and Numbers, 90% in Measurement, only 40% in Data Representation, 85% in Geometry, and 64% in Algebra. The overall coverage was estimated at 80% which was amongst the highest in the study. The figures for Science were less impressive, however. (Martin et al, 2000, p.182) While the matching was high for Biology and the Nature of Science, they were only 50% or less for Physics, Chemistry and Earth Science. In fact, the Malaysian TIMMS-R cohort had studied Science only from Grade 7 onwards, so their mean score above the international average was surprisingly good. TIMSS 2003 will have a cohort which

studied Science from Grade 4, and the Government is planning to begin the subject from Grade 1 in future.

As in all other countries, there was virtually no difference between the Malaysian mean score on the total tests and on the abbreviated forms which consisted of items judged fair for the Malaysian sample. In other words, the TIMSS-R tests were sufficiently relevant to the national syllabi for policy makers to make valid judgements about.

National Report

A national report of 96 pages was produced by the TIMSS-R team, and 500 copies were distributed to education officials, universities and teachers colleges. Although the report was written in Bahasa Malaysia, I was able to interpret much of it, and was impressed with its presentation, headings, charts and the kinds of examples given. It would clearly be of much interest to many Malaysian educators.

Dr Zakaria presented the report, and its main findings, to a meeting of the Malaysian Cabinet Committee on Education, and the Minister of Education was briefed so that he could answer questions in Parliament about the project. In fact the findings were the subject of debate in Parliament. In addition, a national seminar on the study was conducted by the TIMSS-R team. It ran for three days and was attended by about 500 educators – Ministry officials, university staff, teachers college staff, principals and teachers.

National newspapers gave the TIMSS-R results front page coverage at the time of the release, as the results were better than expected. Radio and TV also reported the subsequent debate in Parliament on the findings, but there was not extensive coverage after this.

Impact of the Study in Malaysia.

The positive results were the subject of some public discussion and congratulation, but lessons were also drawn about some apparent deficiencies. The NRC reported that Malaysian students were competent, by and large, on routine, well-drilled tasks, but much weaker when it came to application of the routines in real life problems. This was a common theme in the National Seminar.

For instance, the national report showed that Malaysian students had the highest ranking (1^{st}), internationally, on a routine subtraction task (92% of the sample correct), and performed very well on most straightforward geometry and arithmetic tasks that were released for public inspection. However, on tasks which required the students to read and analyze new information, and solve novel problems, their rankings were well below the international average.

In the National Blueprint for Education in Malaysia, which Dr Zakaria's Division prepared after the TIMSS-R study was released, the Government has declared its intention to give new emphasis to Science and Mathematics. Science will be introduced to the Grade 1 curriculum. Greater attention will be given to practical investigations in teaching and teacher-training. A new category of "Smart Schools"

has been selected for special teaching, making extensive use of computer technology, and more "hands on" learning in Mathematics, Science and Languages. This venture will expand to all schools by 2020.

No policy changes in assessment policies are planned. National Examinations already exist in Grades 6, 9, 11 and 13, and the levels of test development, administration and scaling are quite advanced.

Human Resources Development

The technical training provided by IEA was welcomed by the NRC, particularly in the area of sampling and large data set management. He reported that he learned little about instrument development personally, but those teachers who were involved in translating, editing and scoring the tests spoke positively about the variety of test item types, and the clever ways that they assessed students' application of knowledge.

Virtually all those who were involved in TIMSS-R are still working on assessment tasks, either in TIMSS (2003) or other national assessments. Their expertise is being put to good use in further data collection and analysis.

IEA was given "Excellent" ratings for all aspects of the organization of the study. They had no complaints to make.

Funding

As in most of the low income countries in this report, Malaysia funded all of its local costs, and World Bank grants covered the international fee, and travel to training sessions. Dr Zakaria reported that Malaysia would not have participated without World Bank aid, and claimed that they could not take part in future IEA studies without this kind of assistance. His judgement was that the TIMSS-R study was "Very important" for Malaysia.

The impact of the study was such that the Government needed little persuasion to repeat the study in 2003. In particular, Malaysian educators hope to see some improvement in their ranking in Science, following the new policy of moving the introduction of Science down to Grade 4, and the other changes they have instituted in Science teaching. The effects of starting Science in Grade 1 will not be felt for some time, of course.

Conclusion on Malaysia

TIMSS-R was the first IEA survey that Malaysia had participated in. The general conclusion amongst educators was that it was beneficial to Malaysia. The tests were accepted as valid for most Malaysian schools, and the findings were given good coverage in the newspapers and in Parliament. Educators learned much about their relative strengths and weaknesses. The national report has been widely circulated, and educators resolved to make several changes in curriculum and teaching style, following the National Seminar on TIMSS-R. For such reasons, the Government wishes to be take part in further IEA surveys, in order to establish "whether they are making progress".

SUMMARY AND CONCLUSIONS

This report has detailed the impact of TIMSS-R (1999) on a set of 18 low and middle-income countries, all of which received financial assistance from World Bank in order to participate. All of these countries were in the lowest half of the 38 countries that participated, in respect of GNP per capita, and all except one claimed that they could not have participated without the support they received.

Validity of the Information Gained

The main source of information for this evaluation was found in the responses made to a four-page questionnaire sent to each of the National Research Coordinators, (NRCs), so it is pertinent to ask whether their responses were honest and unbiased. If they reported that the impact was slight or that they could have managed without World Bank help, then they were unlikely to be given financial assistance by the Bank in the future. While such factors cannot be ruled out, they seem unlikely to have been a major source of distortion, for these reasons.

Firstly, the World Bank officers in most of the countries were asked to report on the extent of publicity and the educational impact, and their responses generally confirmed what the NRCs reported. Secondly, it is unlikely that many of the kinds of reforms described, in curricula, and examination policies, could have been exaggerated, as they are public knowledge in each country. Thirdly, it was clear that the judgements made about content validity of the tests were similar to those reported in the international reports. Fourthly, I was able to see at first hand, the changes occurring in three countries, and I have no doubt that the impact in these cases was considerable. Fifthly, it was clear from a few negative judgements that many NRCs were quite prepared to be honest. In many cases, the responses were made, and checked by others in the TIMSS-R team of the country concerned.

With regard to the questions about the need for financial assistance, it is obvious that most of these countries could not have participated in TIMSS-R, without such help, as they had little to spend on education, and all spent very little on research and development. (See Martin et al., 2000, p 25). With regard to the questions on human resource development, it is clear also that new skills have been developed, as new national assessments have been instituted, which require such skills, and the facts on the unusually high stability of staff can easily be verified, as most have a continuing relationship with IEA. If the NRCs or their governments had not considered their participation in international studies of this kind to have been worthwhile, they surely would not have agreed to spend scarce funds on the local costs required for further surveys. Thus, there is good reason to believe that the generally positive picture of the benefits enjoyed by the participating countries was a fair and accurate one.

At the time of completion of this report, no response had been received from two countries. However, while visiting IEA headquarters, I was able to sight reports from both countries, outlining the press conferences held at the time of the release, and the subsequent newspaper publicity. Both gave the TIMSS-R study a great deal of positive publicity. Furthermore, the World Bank officer for one of these countries spoke glowingly about the value of the TIMSS-R findings in preparing a major

human resources project for this country. The same officer recommended that such surveys be extended to other countries in the Middle East and North Africa.

Summary of Major Conclusions

To summarize the main findings, this review has shown that:

- The instruments used to measure Mathematics and Science achievement in most of these 18 countries provided fair and valid assessments of the abilities of a cross-section of their Grade 8 students. They were therefore able to provide useful information to policy makers in those countries.
- In almost every country, a national report of the results was prepared and distributed, and was read by senior officials in the education sector. There was good reason to believe that the findings were taken seriously.
- Most NRCs reported that policy makers showed most interest in the international rankings in achievement and in differences in the sub-tests. However, there was also high reported interest in the differences in the school contexts, between countries, and for those that had participated in earlier studies, in trends in achievement over time. Most NRCs claimed that they had learned a great deal about several aspects of education in their country.
- Newspapers publicized the findings widely at the time of the TIMSS-R release in December 2000. However, the study was not given major coverage by the media, after that, except in a few countries. Most NRCs reported more coverage in educational publications, and this was confirmed by Bank officers. All except two countries reported that there was some public discussion of the findings.
- Over half the NRCs reported that there had been some changes in curricula, following the release of the findings. In nearly every case, the NRC was on a team that was engaged in curriculum reform. Most expected future curriculum changes.
- Over half the NRCs reported changes in assessment policy or practice, following TIMSS-R models, and about one third referred to changes in teaching practice and in teacher training. Many of these changes appear to have been substantial, and are bound to be beneficial.
- Nearly every NRC claimed to have developed new technical skills in the course of the study, and the majority opinion was that the training was of a high quality, and was conducted in a context of immediate practical use, under supervision.
- Nearly all NRCs, and their TIMSS-R colleagues, were still using those skills in making national or international surveys. Thus the training they received was helpful in enabling them to sustain the ability to conduct research which would benefit their education system.
- All NRCs expressed a wish to participate in further IEA surveys, but all except one country would not be able to do so without continued financial support for the international expenses of participation.
- Over half the NRCs expressed the view that participation in TIMSS-R 1999, was “Extremely important” for their education system; the remainder judged it to be “Very important”. None played it down.

It is clear that the impact of participation in TIMSS-R has been substantial, and positive, for most of these countries. It seems unlikely moreover, that many of the

reforms would have occurred without the stimulus of an international ranking. There is good reason to believe that the World Bank aid was well spent on these countries.

RECOMMENDATIONS

In the light of these positive conclusions, I would like to recommend that:

1. World Bank continues to support IEA in its efforts to encourage low and middle-income countries to participate in its international surveys. It is to the advantage of World Bank, and other international bodies, to have access to good quality data on achievement levels, standardized across countries. It is also of real benefit to the educational policy makers in those countries, as we have seen.

2. IEA continue to emphasize the training of NRCs in their projects. In most cases, the NRCs clearly needed, and spoke highly of the training they received.

3. IEA make greater efforts to attract more low and middle-income countries, especially the larger countries of Africa, South America, and Asia to participate in their surveys. Countries such as China, India, Nigeria, Brazil and Argentina account for large proportions of the world's illiterate population, but figures on the levels of achievement in their schools are difficult to determine without an evaluative study such as IEA can provide.

REFERENCES

Fuller, B. (1991) *Growing Up Modern: The Western State Builds Third World Schools.* London, Routledge

Greaney, V. (1996) Reading in Developing Countries: In V Greaney (Ed.) *Promoting Reading in Developing Countries.* Newark, Delaware, International Reading Assn.

Lockheed M.E. & Verspoor, A M., (1991) *Improving Primary Education in Developing Countries.* Oxford, Oxford University Press.

Martin, M.O., Mullis, I.V.S., Gonzales, E.J., Gregory, K.D., Smith, T.A., Chrostowski, S.J., Garden, R.A., & O'Connor, K.M. (2000). *TIMSS 1999, International Science Report.* Boston & Amsterdam, IEA.

Mullins, I.V.S., Martin, M.O., Gonzalez, E.J., Gregory, K.D., Garden, R.A., O'Connor, K.M., Chrostowski, S.J., & Smith, T.A. (2000) *TIMSS 1999, International Mathematics Report.* Boston & Amsterdam, IEA.

Robitaille, D.F., Beaton, A.E. & Plomp T. (Eds.) (1998) *The Impact of TIMSS on the Teaching and Learning of Mathematics and Science.* Vancouver, & Amsterdam, IEA.

Vari, P (Ed.) (1997) *Are We Similar or Are We Different? A Study of Grade 8 in Nine Central and Eastern European Countries.* Amsterdam, IEA.

World Bank (1998). Development Grant Facility Funding: New Grant Proposal for FY 1998. Washington, DC. (Unpublished Document)

APPENDIX I TERMS OF REFERENCE

Independent Evaluation of TIMSS-R (1999)

World Bank has given support to a number of developing and lower middle income countries to help them carry out TIMSS-R, and there is a need to establish what impact this has had on the education systems in these countries.

The following issues should be addressed:

Relevance
Country issues: How relevant is TIMSS-R for policy making?
Did the test items (in the view of the Ministry of Education) represent a reasonable sample of the a) official curriculum and b) the curriculum actually delivered in schools?
How relevant is the final report for policy makers?
Is there any evidence that policy makers have read the report?
How relevant was the cross-country data reported in TIMSS for policy making?
Bank issues:
How relevant is TIMSS-R in terms of deepening understanding of education within the country?
Has the Bank team for that country actually used TIMSS-R data in its work?

Efficacy: Use of Results
Have educational policies/practices been altered primarily as a result of participation in TIMSS? E.g. Curriculum content, In-service teacher training, More involvement in international bodies working in education?

Is there evidence that TIMSS results been made public within the country?
Have they been featured in
Parliamentary debate?
Newspaper editorials?
Radio programs?
TV programs?

How has the Bank used the results for its own policy making? E.g. in sector strategies, training programs, project development?

Efficiency:
At the national level: Questions should focus on the use of resources--and could be built around the concept of efficient use of resources -manpower, time and finance-- to carry out TIMSS?
What is the view of Bank staff working in education in these low-income countries?

Sustainability:

Capacity development
Has capacity in assessment been developed in the participating countries?
Does the TIMSS-R team continue to work on national assessment issues?
Are members of the team using their skills in related areas- developing diagnostic tests?
advising examination authorities?

What is the evidence that the TIMSS initiative is sustainable?
Is the institution/team that carried out the assessment still functioning as an assessment unit?
Is there a line item budget for educational assessment?
How many of the national core team continues to work on educational assessment?
Is the team developing/carrying out a national assessment?

Institutional Development Impact:
Has the country created an institution to carry out future assessments?
Independent? Within government?
How has the assessment affected practices in the national curriculum unit?
Indirect impact: To what extent are those who previously worked on TIMSS (and who are no longer working on national/international assessment) using their TIMSS inspired assessment skills in their new institutions?

Grantee's Performance:
How did the Grantee use IDA funds e.g. training, travel, test development?
Did the Grantee complete each implementation stage on time?
Did the grantee actively encourage wide-scale release of the results?
Did the grantee make good use overall of the results?
IEA is also a grantee: There is a need to have a professional auditor in the Bank examine and comment on IEA's audit report. (This information could be included as part of the evaluation report. World Bank officers would attend to this matter.)

Performance of donors and intermediary organizations:
IEA might wish to comment and evaluate the extent to which there was a timely release of IDA funds.
IDA reporting requirements including audit, nature of communication with IDA counterparts, significance of IDA funding, use of TIMSS by IDA (including brown bags, seminars and HD week)

Perspective of Stakeholders:
Some descriptive national case studies would be useful.
Dr. H Wagemaker of IEA could comment on the importance of IDA funds for TIMSS-R.

The evaluation report should include:

Summary of key findings
Conclusions and lessons learned
Recommendations

APPENDIX II.

EVALUATION OF IDA SUPPORT FOR TIMSS-R (1999)

QUESTIONNAIRE FOR NATIONAL RESEARCH CO-ORDINATORS

To: National Research Coordinator
TIMSS-R

From: Warwick Elley
Educational Consultant.

I have been contracted by IEA to study the **impact of TIMSS-R** (**1999**) on education policy and practice in a group of countries that received funding assistance from World Bank. Your views are a key source of information in this project. Therefore, I would be very grateful if you would attempt to respond to each of the following questions. Feel free to add any comments where appropriate, or at the end. Your individual responses will be **confidential.**

INTRODUCTION

1. Please give your **name** and **official title**. _____
   _____

2. **How many people** (professionals) worked on the implementation of TIMSS-R in your country: Full-time? (Over 90% of their time) _____
   Over Half-Time? (50 – 90% of their time) _____
   Part-Time? (10 - 50% of their time) _____

RELEVANCE OF TIMSS-R

1. How well did the TIMSS-R international tests match the content objectives of your **official national curriculum**? (Mark with an "**x**")

	Very Well (Covered over 90% of Objectives)	*Quite Well* (Covered over 75% of Objectives)	*Not Very Well* (Covered less than 75% of Objectives)
Maths Test	_____	_____	_____
Science Test	_____	_____	_____

   Comment: _____
   _____

2. How well did the TIMSS-R tests match the content objectives of your **national curriculum**, as *implemented* in the schools?

Very Well / Quite Well / Not Very Well

Maths_Test_____/ _____ /_____

Science Test_____/ _____ /_____ _____

Comment: _____

_____

3. Was a **National Report** prepared for your TIMSS-R findings? (Yes/No)_____

4. To the best of your knowledge, did **Senior Education Officials** read the National Report, or summaries of it? (Y/N) _____

5. Did the **Minister of Education** read the Report (or summaries?) (Y/N)_____

6. What aspects of the **findings** included in your National Report (or other summary of findings) on TIMSS-R were of most interest to policy makers?

Of Much Interest / Considerable / Some Interest/ None

International Rankings_____/_____/_____/_____

Trends over Time _____/_____/_____/_____

Content Area Differences_____/_____/_____/_____

Student Attitudes _____ /_____/_____/_____

Differences in Context _____/_____/_____/_____

Other: _ _ _ _ _ _ _ _____/_____/_____/_____

Comment:

_____

_____

7. To what extent do you think TIMSS-R contributed to a **greater understanding** of education in your country?

A Great Deal / A Little / Not Much

Student Achievement Levels_____/_____/_____

Student Attitudes _____/_____/_____

Curriculum Emphases _____/_____/_____

Teaching Methods _____/_____ /_____

Education Resources _____/_____/_____

Other _ _ _ _ _ _ _ _ _ _____/_____/_____

Comment if you wish: _____

_____

USE OF RESULTS

1 How much **coverage** was there of TIMMS-R findings in these media?

Extensive / Some / Little / None

Newspapers _____/_____/_____/_____

Parliamentary debate _____/_____/_____/_____

Magazine articles _____/_____/_____/_____

Special radio programs _____/_____/_____/_____

TV programs _____/_____/_____/_____
Teacher publications _____/_____/_____/_____
Academic journals _____/_____/_____/_____

2. Did the TIMSS-R project promote any **public discussion** of the Mathematics Curriculum? (Yes/No) _____

3. Did the TIMSS-R project promote any **public discussion** of the Science Curriculum? (Y/N) _____

4. Did the TIMSS-R project promote any ***change*** in the Mathematics curriculum? (Y/N)_____ (Eg)_____
_____

5. Did the TIMSS-R project promote any ***change*** in the Science Curriculum? (Y/N)_____(Eg) _____
_____

6. Are there likely to be more curriculum changes **in the future**, because of TIMSS-R findings? (Y/N) _____ (Eg) _____
_____

7. To what extent have you, or your team been involved in **discussion** of curricular reform?_____

8. Can you list any **other changes** that have been implemented, or proposed, as a direct or indirect result of TIMSS-R findings?
In Teaching Method _____

In Teacher Training _____

In Assessment _____

Other Policies_____

HUMAN RESOURCES AND SUSTAINABILITY

1. How would you evaluate the **special, or on-the-job training** that you and (your TIMSS-R team) received during TIMSS-R?

	Very Helpful	*Helpful*	*Not Helpful*	*Not Applicable*
Sampling	_____	_____	_____	_____
Test Development	_____	_____	_____	_____
Questionnaire Devpt	_____	_____	_____	_____
Data Management	_____	_____	_____	_____
Quality Control	_____	_____	_____	_____
Other _ _ _ _ _ _ _	_____	_____	_____	_____

2. What particular **skills or knowledge** have you learned, or expanded, through your participation in TIMSS-R? _____
_____

3. How would you rate the **overall quality** of the organization of TIMSS-R?

	Excellent	*Very Good*	*Adequate*	*Not Adequate*
Leadership	_____	_____	_____	_____
Decision-Making	_____	_____	_____	_____
Communication	_____	_____	_____	_____
Time Deadlines	_____	_____	_____	_____
Data Processing	_____	_____	_____	_____
Other_ _ _ _ _ _	_____	_____	_____	_____

Comment on Organization: (Special Praise or Problems?):
_____
_____
_____

4. Are you still working on projects that involve **assessments**, such as examinations, national assessments, test development, advice on evaluation, etc? (Y/N)_____

5. What roles are **other professional staff** who worked on TIMSS-R, for more than 50% of their time, carrying out now?
   _____
   _____

6. If you, or someone in your TIMSS-R team, have **changed roles**, to what extent are they using the training they received from TIMSS-R?

   (Circle one) *Often / Now & Again / Seldom / Never / Not Applicable*

7 Does your country plan to make any **national assessments** of achievement in future? *With IEA? (Y/N)*_____ *With Others?*_____ *On Your Own?*_____

Comment: _____

8. Approximately what percentage of the costs of implementing TIMSS-R was contributed by your **Government**, **and/or Institution**? *Local Costs*? _____
International Costs? _____

9. Approximately what percentage of the costs were provided by a contribution from **World Bank**? *Local Costs?* _____
International Costs? _____

10. Do you think your country would have participated in TIMSS-R without financial support from **World Bank**? _____

11. Do you think that your country will fund the full costs of participating in the **next TIMSS** study? _____

12. Considering the time, effort and resources that were used in TIMSS-R, **how important** do you think the project was for your country?

(Circle one) *Extremely Important/ Very Important/ Quite Important/ Not Important*

Thank you ! Please e-mail this questionnaire to vwelley@iconz.co.nz
As I will be travelling after March 25, I would be pleased if you would also send a hard copy of your completed questionnaire to my son's home in Singapore. (Address provided)

APPENDIX III

COUNTRIES WHICH RECEIVED WORLD BANK SUPPORT FOR TIMSS-R

Bulgaria
Hungary
Indonesia
Jordan
Latvia
Lithuania
Macedonia
Malaysia
Moldova
Morocco
Philippines
Romania
Slovak Republic
Slovenia
South Africa
Thailand
Tunisia
Turkey