



United Nations
Educational, Scientific and
Cultural Organization

Organisation
des Nations Unies
pour l'éducation,
la science et la culture

Organización
de las Naciones Unidas
para la Educación,
la Ciencia y la Cultura

Организация
Объединенных Наций по
вопросам образования,
науки и культуры

منظمة الأمم المتحدة
للتربية والعلم والثقافة

联合国教育、
科学及文化组织

Internal Oversight Service Evaluation Section

IOS/EVS/PI/93
Original: English

Evaluation of Dimensions of the Work of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) and of its Programme of Cooperation with the International Institute for Educational Planning (IIEP)

Dr. Kadriye Ercikan
Rubab Arim
Maria Oliveri
Deborah Sandilands

August 2008

The views and opinions expressed in this document are those of the authors and do not necessarily represent the views of UNESCO or IOS who has contracted the study and assured adherence to evaluation standards. The designations employed and the presentation of material throughout this document do not imply the expression of any opinion whatsoever on the part of UNESCO concerning the legal status of any country, territory, city or area of its authorities, or concerning its frontiers or boundaries.

Table of Contents

EXECUTIVE SUMMARY	4
<i>The Scope of Evaluation</i>	4
<i>Overall Comments</i>	5
<i>Recommendations</i>	6
1. Guidelines for data use	6
2. Develop and document content and construct validity evidence as well as cultural validity evidence.....	6
3. Comparability of scores across SACMEQ tests in phases I, II, and III	7
4. Reporting reliability estimates and measurement accuracy	8
5. Data capturing	8
6. Reporting of significance level of differences	8
7. Sampling design refinements for future phases	8
8. An Independent advisory panel	9
I. OVERVIEW OF SACMEQ.....	10
SACMEQ I.....	11
SACMEQ II.....	12
SACMEQ III	13
II. METHODOLOGY.....	13
Document review and analysis	13
Meetings with key informants	14
III. THE FINDINGS	15
<i>Technical Soundness of UNESCO's outputs</i>	15
Overall.....	15
Target Population	16
Tests	17
How valid were the tests?.....	21
Were pre-tests conducted?.....	23
<i>Data</i>	24
Were the test reliabilities at an acceptable level?	26
Questionnaires.....	30
<i>Relevance and adequacy of organizational arrangements</i>	30
What UNESCO inputs have gone into the development of the various UNESCO outputs? Is the level adequate	

to produce and sustain high quality outputs?.....	30
To what extent do UNESCO outputs meet the technical needs and requirements of programme beneficiaries?	32
What are programme beneficiaries' needs and requirements with regard to monitoring and assessing learning achievement?.....	32
<i>Effectiveness of the quality control and quality assurance procedures.....</i>	<i>32</i>
Is a monitoring and evaluation system in place to enable Headquarters to monitor the progress of the programme? How effectively has the monitoring and evaluation system functioned?.....	33
<i>Results</i>	<i>34</i>
What are the expected results of SACMEQ?.....	34
What results have been achieved?	36
What factors have contributed to and/or prevented the achievement of expected results?	40
<i>Risks</i>	<i>40</i>
<i>Final Comments.....</i>	<i>41</i>
REFERENCES	44
APPENDIX A: LIST OF DOCUMENTS REVIEWED.....	45
APPENDIX B: MANAGEMENT RESPONSE	48

Executive Summary

The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) is an independent organization which functions with the SACMEQ Assembly of Ministers (made up of members from a consortium of countries) making up the SACMEQ Governing Board. In consultation among its members, the Governing Board decides in which training and research support activities SACMEQ will engage. Although the International Institute for Educational Planning (IIEP) is not a member of SACMEQ, this organization has worked closely with SACMEQ to undertake integrated research and training activities for two key purposes: (1) expand opportunities for educational planners to gain the technical skills required to monitor and evaluate the quality of basic education; and (2) generate information that can be used by decision makers to plan and improve the quality of education. Hence, the objectives of the report written by the evaluation team were intended to examine the development of SACMEQ as a capacity building effort and as an assessment system for collecting policy relevant data and to assess the impact of UNESCO's work in the area of monitoring of learning achievements based upon review of documentation provided by IIEP and SACMEQ.

Three SACMEQ surveys have been conducted. The first survey took place during the period of 1995 to 1998 in 7 countries. During 2000 to 2002, SACMEQ undertook a second follow-up survey, SACMEQ II, in 15 countries. SACMEQ III is an ongoing project consisting of 15 participating countries which started in 2005, and is expected to be completed by 2009. There is extensive documentation and information available about SACMEQ I and II. However, documentation for SACMEQ III is limited. The evaluation team considered all three phases of SACMEQ but focused detailed document analysis on SACMEQ II as the most recent survey with sufficient documentation.

The Scope of Evaluation

The scope of the evaluation was defined by the terms of reference and constrained by time, documentation and funding available for the evaluation. The evaluation focused on the technical soundness of UNESCO's outputs related to monitoring and assessing learning achievement, the relevance and adequacy of organizational arrangements, the effectiveness of the quality control and

quality assurance procedures, and assessment capacity development. The evaluation is based on discussions and meetings with key IIEP staff working on SACMEQ and SACMEQ staff, a review of documents received from IIEP-UNESCO, group discussions with country representatives in Namibia in December 2007, observations of a data cleaning capacity building activity in Namibia and a review of documentation contained in the SACMEQ website. More site visits to IIEP or SACMEQ countries were not within the scope of this evaluation due to time and funding limitations for the evaluation. In addition, the evaluation team was asked to return all the electronic documentation they were provided back to IIEP in January, 2008. Even though most of the document analyses were completed by January, this set of key documentation was not available to the evaluation team at the time of writing this Executive Summary. A key limitation of this evaluation has been the lack of access to the three key sets of tests of reading and mathematics used in the three phases of SACMEQ. Despite the evaluation team's repeated requests and explanations of the importance of reviewing of these tests, the evaluation team was not given access to use these tools in the evaluation.

The criteria used in this evaluation are based on previous research on large-scale and international assessments and guided by two sets of ethical and professional testing guidelines set out by the International Testing Commission (ITC) and the American Psychological Association (APA), American Educational Research Association (AERA), and The National Council on Measurement in Education (NCME).

Overall Comments

SACMEQ is a well established, successful cross-national data collection and capacity building effort. Overall, the technical soundness of the SACMEQ project is of good quality. The aims of the study are clearly stated and are relevant for all participating countries, and the defined target population is appropriate. The achievement tests have been built upon a solid framework and blueprinting process that were designed to address the policy questions expressed by the participating countries. The content of the SACMEQ II questionnaires (pupil, teacher, and school head) provides excellent coverage of the study's overall aims. Participants in workshops have expressed great

satisfaction with the training and technical support that they receive. Evidence from these participant comments indicates that the data they receive from SACMEQ meet their needs for policy development. SACMEQ needs to be congratulated and commended for excellent organizational arrangements, effective quality control and quality assurance procedures, and assessment capacity development in the participating countries.

This evaluation found SACMEQ to be a very successful effort in achieving the two goals stated above. Careful review of processes and procedures lead to a list of recommendations presented below for further improvements in the programme.

Recommendations

1. Guidelines for data use

Large-scale international data collection efforts are complex and need to accommodate multiple constraints and be tailored to the needs and conditions of the participating countries. Deviations from ideal situations due to such complexities result in limitations in interpretability of data that may not be obvious to data users. For example, for assessments that are intended to provide information to guide schooling and learning in schools, grade-focused target population is indeed appropriate as the target population. However, in SACMEQ, this sampling results in country data that have very different pupil age distributions which have implications on interpretation of cross-country results. Another difference across countries is their exclusion rules of pupils. We recommend to SACMEQ to provide guidelines to the participating countries and highlight the limitations of interpretability of data across countries due to differences in age distributions and exclusion rules in the SACMEQ samples. However, we recognize that significant efforts have been made by SACMEQ in challenging circumstances to ensure the highest degree of international comparability.

2. Develop and document content and construct validity evidence as well as cultural validity evidence

We recommend for SACMEQ to develop and make publicly available comprehensive documentation of content validity evidence for the reading and mathematics tests at the

beginning of assessment development in each phase. This content validity evidence needs to be revised periodically to reflect possible changes in schooling and curricula of the countries. The documentation should include the degree to which test content matches country curricula in reading and mathematics. Current documentation of the processes, procedures and outcomes of framework development, item development and the test construction processes should be expanded to include more details and more information about the involvement of countries in these processes. However, we do recognize that there are some real-life limitations in developing this evidence.

Furthermore, construct validity evidence is at the core of appropriateness and meaningfulness of assessment data. In addition to statistical construct validity evidence, we recommend to SACMEQ to conduct empirical studies to examine what the test questions are measuring using cognitive laboratories or think aloud protocols. These methods can be used to determine the cognitive processes students engage in when constructing their responses. This approach can result in a more realistic determination of pupils' interpretation of test items and the cognitive processes they engage in when they respond to the test items, as well as provide empirical evidence to skill level identification. In addition, construct validity evidence needs to be expanded to include documentation about the readability and appropriateness of questionnaire items.

Finally, in an international assessment such as SACMEQ cultural validity evidence is essential for appropriate and meaningful interpretation of data within and across countries. In addition to the statistical evidence, such as country by item parameter interaction, which are currently estimated and documented, we recommend obtaining cultural validity evidence that includes documentation from sensitivity review panels for the participating countries and documentation and analyses of cognitive processes for examinees taking tests in different languages.

3. Comparability of scores across SACMEQ tests in phases I, II, and III

One of the objectives of SACMEQ is to make comparisons concerning changes in educational quality and educational achievement using data from SACMEQ I, SACMEQ II, and SACMEQ III. Ensuring comparability of test scores across these phases is important for meeting this objective. This

degree of comparability needs to be examined empirically and documented. Although we acknowledge that significant efforts have been made to ensure comparability among SACMEQ phases, examining and demonstrating comparability of scores across different phases should be planned for the upcoming phases at the beginning of these assessment phases. Linking items need to be selected based on content as well as psychometric consideration for these items.

4. Reporting reliability estimates and measurement accuracy

We recommend reporting of reliability of tests as well as item response theory based standard error of measurement for a selected set of scores on different parts of the scale to allow examination of measurement accuracy at different score points. Such information is a very basic requirement for examining properties of data created by the SACMEQ tests.

5. Data capturing

We recommend conducting a systematic feasibility study to explore different types of data capturing instead of manual data entry of data. There are costs associated with equipment and software needed for any type of automated data capturing. However, there are potential tremendous savings in time, cost, and in accuracy of data that could be achieved by using an automated data capturing method. This possibility of a more advanced technology needs to be explored in a systematic way.

6. Reporting of significance level of differences

Differences among groups such as gender, ethnic, or regional groups, may be the focus of key policy decisions for some countries. Inappropriate interpretations of differences for such groups pose a risk of inappropriate interpretations if statistical significance levels of these differences are not taken into account. The evaluation team recommends that all future SACMEQ country reports include statistical significance levels for comparison groups.

7. Sampling design refinements for future phases

We recommend SACMEQ to refine the sampling design collaboratively with the participating countries to establish more comparable exclusion rules across countries and to obtain larger

sample sizes. These refinements will move SACMEQ in the direction of obtaining more precise country results as well as improve comparability of results across countries.

The exclusion rules were similar, but not identical for all countries. Countries varied in exclusion of “special schools” and private schools. These variations contribute to incomparability of data for different countries. In addition, exclusion of certain types of schools results in students attending these types of schools being excluded from the SACMEQ studies. Such exclusions affect the representativeness of the data for the participating countries and do not allow the use of the data from informing policy decisions that involve the excluded students.

8. An Independent advisory panel

We recommend to UNESCO Headquarters to create an international technical advisory board consisting of psychometricians, education policy experts, and statisticians among others, with international reputation. The board should have an arms-length relationship with SACMEQ and should review activities, processes, and products and provide a report (on a yearly basis) to UNESCO Headquarters. However, we do recognize that the creation of such a panel is a difficult task due to potentially high transaction costs.

Evaluation of Dimensions of the Work of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) and of its Programme of Cooperation with the International Institute for Educational Planning (IIEP)

The first part of this evaluation report contains an overview of SACMEQ, and the three SACMEQ projects which were carried out between 1995 and 2007. The second part presents a brief overview of the methodology and the third part summarizes findings from the evaluation of the technical soundness of UNESCO's outputs related to monitoring and assessing learning achievement. This is followed by evaluations of the relevance and adequacy of organizational arrangements, the effectiveness of the quality control and quality assurance procedures, and assessment capacity development.

The evaluation is based on discussions and meetings with key IIEP staff working on SACMEQ, a review of documents received from IIEP-UNESCO, group discussions with country representatives in Namibia in December 2007, observations of a data cleaning capacity building activity in Namibia and a review of documentation contained in the SACMEQ website (<http://www.sacmeq.org/>). A full list of documents used in the preparation of this report is provided in Appendix A at the end of this report.

I. Overview of SACMEQ

The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) is a consortium of countries that works closely with the International Institute for Educational Planning (IIEP) to undertake integrated research and training activities for two key purposes: (1) expand opportunities for educational planners to gain the technical skills required to monitor and evaluate the quality of basic education, and (2) generate information that can be used by decision makers to plan and improve the quality of education. The first two educational policy research projects undertaken by SACMEQ ("SACMEQ I" and "SACMEQ II") were designed to provide detailed information that could be used to guide planning decisions aimed at improving the quality of education in primary

school systems. The first survey took place during the period of 1995 to 1998 and was conducted in order to collect data on the conditions of schooling and the quality of education in 7 countries, covering around 1100 schools and 20,000 pupils. The Ministries of Education involved were Kenya, Malawi, Mauritius, Namibia, Zambia, Zanzibar, and Zimbabwe. During 2000 to 2002, SACMEQ undertook a second follow-up survey in 15 countries, covering around 2500 schools and 45,000 pupils. The participating countries were Botswana, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania (Mainland), Uganda, Zambia, Tanzania (Zanzibar), and Zimbabwe.

SACMEQ III is an ongoing project consisting of 15 participating countries started in 2006, and is expected to be completed by 2009. There is extensive documentation and information available about SACMEQ I and II. However, documentation for SACMEQ III is limited.

SACMEQ I

SACMEQ was modeled after a similar study which had been conducted in Zimbabwe in 1991. The purpose of SACMEQ I was to collect data to assist countries with policy decisions regarding quality of education. This data was intended to help understand the relationship between educational inputs and learning outcomes, in particular, whether educational inputs were cost effective and to collect data to assess whether the countries' resource allocations were resulting in equitable distribution of educational provisions.

For each country, the administrative organization involved a National Centre (NC) consisting of a full-time research team responsible for project work within the country. There was a sub-regional co-ordination centre (SCC) with a small full time staff responsible for the co-operative development of all aspects of the project and for ensuring the implementation of the work in each of the countries according to the agreed plans.

The SACMEQ I assessment was developed to measure reading at the Grade 6 level. Reading was selected because of its foundational importance for learning in all subjects. This assessment covered the general domains of narrative, expository, and document reading. Background

questionnaires for pupils, teachers, and school heads were designed to elicit information on a wide range of aspects of the educational and social environments of pupils, the general conditions of schooling, and the levels of educational inputs to schools in the sub-region.

The results of the surveys were intended to be used by ministries of education to examine the implications for policy-making in their countries. Data gathered in the first cycle of the study was intended to serve as baseline and allow countries to examine whether the inputs, processes and performance have improved, remained about the same, or decreased in future data collection efforts.

The results of SACMEQ I were reported in a series of SACMEQ Policy Research Reports. One report titled “The quality of education: some policy suggestions based on a survey of schools” was prepared for each country’s results. Each report presents the research results and policy suggestions that emerged from the implementation of SACMEQ’s initial educational policy research project.

SACMEQ II

The second SACMEQ assessment was designed as a follow-up to the first, and was begun in 2000. The SACMEQ II Project had two main aims. First, to provide comparative information concerning changes in the quality of education that had occurred since the SACMEQ I Project in 1995. Second, to generate research-based policy advice to guide decisions focused on improving the quality of education. SACMEQ’s mission was stated as “to undertake integrated research and training activities that will: (a) expand opportunities for educational planners to gain the technical skills required to monitor, evaluate, and compare the general conditions of schooling and the quality of basic education; and (b) generate information that can be used by decision-makers to plan the quality of education.”

There are important differences between the first and second SACMEQ assessments. First, SACMEQ II was designed to assess both reading and mathematics literacy. Second, both teachers and pupils were assessed in these areas. Another significant difference was that three countries translated the assessment from English into their own languages: namely Mozambique translated the

assessment into Portuguese and Tanzania (Mainland and Zanzibar) translated it into Kiswahili.

Reading literacy was defined for SACMEQ II in a similar fashion as it had been in the original SACMEQ I assessment, in three domains: narrative prose, expository prose, and documents. SACMEQ II also included the construct of mathematics literacy in three domains: number, measurement, and space/data.

Results containing policy suggestions were prepared by the countries in a similar fashion to those from SACMEQ I.

SACMEQ III

This project started in 2005 and is ongoing. The data collection for the SACMEQ III Pilot Project occurred in late 2006. In this latest phase, a new data collection instrument HIV- AIDS Health Knowledge Instrument for pupils and teachers was added to the Reading and Mathematics tests. Given the ongoing nature of this project, there is no detailed documentation about these instruments. In our meetings with the IIEP staff and the Director of SACMEQ we found out that there is substantial overlap between SACMEQ II and SACMEQ III reading and mathematics instruments. However, we do not have any description of the HIV-AIDS Knowledge test.

II. Methodology

The evaluation team employed three key methods: (1) document review and analysis; (2) meetings with key informants; (3) and on-site observations of a capacity building workshop.

Document review and analysis

The evaluation team was provided with extensive documentation and reports on SACMEQ I and SACMEQ II electronically as well as in hard copy, at the beginning of the evaluation. Reviews of these documents allowed the evaluation team to examine the development of SACMEQ as a capacity building effort and as an assessment system for collecting policy relevant data and to assess the impact of UNESCO's work in the area of monitoring learning achievements. The documents allowed comparison of SACMEQ I and SACMEQ II characteristics and assessment of progress in terms of

goals, procedures, and use of data. Reviews of country reports provided information about how results from the data collection efforts were used and summarized. The country reports were based on report guidelines provided to countries by IIEP and had very similar structures. There were two types of reports for each country. One report summarized findings from the assessment and the surveys and the other focused on interpreting results and identifying policy implications. Both sets of reports were reviewed for the participating countries. In addition, we reviewed a set of memos and progress reports related to SACMEQ III. Appendix A provides the list of documents reviewed.

Meetings with key informants

The Lead Evaluator Dr. Kadriye Ercikan has had telephone and in person meetings with Dr. Kenneth Ross (IIEP staff working on SACMEQ) and Dr. Demus Makuwa (SACMEQ Acting Director). The in-person meetings took place in Windhoek, Namibia, where the lead evaluator Dr. Ercikan attended and observed a SACMEQ capacity building activity for three days. During this 3-day visit, Dr. Ercikan held meetings with the following key individuals involved in SACMEQ:

1. SACMEQ Director and SACMEQ Data Processing Manager

Dr Demus Makuwa (Namibia)

Dr Njora Hungi (Kenya)

2. Experienced SACMEQ National Research Coordinators

Mrs Toziba Masalila (Botswana)

Mrs 'Motseng Maema (Lesotho)

Mrs Ana Passos (Mozambique)

Dr Andre Leste (Seychelles)

Mr Meshack Moloi (South Africa)

Mr Jabulane Shabalala (Swaziland)

Mr Aminieli Mrutu (Tanzania - Mainland)

Mrs Khadija Ali Mohd (Tanzania - Zanzibar)

Mr Albert Byamugisha (Uganda)

3. Members of the SACMEQ Scientific Committee

Mrs Juliana Nzomo: Kenya (Aga Khan Foundation)

Dr Joseph Chimombo: Malawi (NCERT, University of Malawi)

Professor Servaas van der Berg (Stellenbosh University)

4. SACMEQ Consultants

Professor T. N. Postlethwaite (University of Hamburg)

Professor Patrick Griffin (University of Melbourne)

5. Namibia Ministry of Education Official

Under Secretary of Education Alfred Iilukena

6. Coordinator of IIEP's Programme of Cooperation with SACMEQ

Dr Kenneth Ross (IIEP – UNESCO)

These meetings were very informative about details of the SACMEQ processes and products. The insights gained from meeting with these key individuals are incorporated in all aspects of the evaluation.

III. The Findings

Technical Soundness of UNESCO's outputs

This part of the evaluation report focuses on SACMEQ II as it is the more recent assessment the evaluation team had documentation and information for. The structure of this section of the report corresponds to the terms of reference structure, questions and issues.

Overall

Have the aims of the study been stated clearly and are they relevant?

The aims of the study are clearly set out in the first chapter of each country's report (SACMEQ II). Some of the main objectives of SACMEQ were to undertake integrated research and training activities with the following objectives: (a) expand opportunities for educational planners to gain the

technical skills required to monitor and evaluate the general conditions of schooling and the quality of basic education, and (b) generate information that can be used by decision-makers to plan improvements in their education systems.

In order to address these policy concerns, SACMEQ II invited the ministries of education in the fifteen participating countries to meet and provide information about high priority policy concerns within their ministries. In this meeting, twenty general policy concerns expressed by the ministries were outlined, and summarized in five themes. These policy concerns formed the framework for SACMEQ II.

Target Population

Was the defined target population appropriate?

The desired target population was defined as: “All pupils at Grade 6 level in 2000 (at the first week of the eighth month of the school year) who were attending registered mainstream primary schools (government or non-government school).” This target population is appropriate as Grade 6 is part of compulsory education in all countries participating in the study.

Were the target populations comparable?

The target populations were the same, and all students were in Grade 6. However, there were large differences in age distributions of the participating pupils from different countries. For example, the average age of pupils in the Zanzibar sample was 37.1 months (or about 3 years and 1 month) older than the expected official age for Grade 6. In Zanzibar, age differences across regions were also found. Differences in ages both within a country and between countries can be a source of incomparability. Even though comparability of grade levels may indicate that the target populations were comparable, large age differences between age groups across countries have implications on interpretability of findings for policy purposes. For example, if two countries have very similar average performances for 6th graders and one has a mean age 1-3 years greater than the other, the performance of these two countries imply very different levels of learning outcome profiles. This

implies that it takes more years for pupils in one of the two countries in this scenario to get to the same learning outcome level in grade 6.

For assessments that are intended to provide information to guide schooling and learning in schools, grade-focused target population is indeed appropriate as we stated in the previous section on appropriateness of the target population. We recommend to SACMEQ to highlight the limitations of interpretability of data across countries due to differences in age distributions in the SACMEQ samples.

Tests

What was the purpose of the tests?

SACMEQ II was intended to provide information on the conditions of schooling (through questionnaires for students, teachers, and administrators), and to measure achievement in reading and mathematics in two populations (Grade 6 students and their teachers). Administrations of the tests were intended to provide information regarding changes in educational inputs over time. The program, and therefore the academic assessments which form a part of the program, was also intended to provide training in large-scale assessments for the countries involved.

In order to assess the effects of educational variables on achievement, SACMEQ II gathered data on the personal characteristics and viewpoints of pupils, their teachers and School Heads through the use of questionnaires. Questionnaires also provided information on contextual issues such as learning environments, availability of resources, professional support, educational infrastructure, operation and organization of schools, behaviour problems, and equity within and among regions.

How were the test frameworks developed?

The frameworks for SACMEQ II questionnaires and tests were developed based upon the policy concerns of the ministries of education of the participating countries. The policy concerns were converted into 75 research questions which were then placed into “dummy tables” used to guide instrument development. The identification of key policy questions to guide the questionnaire

development is an important feature that SACMEQ needs to be commended for. The country reports for SACMEQ II describe the developments of frameworks and construction of instruments at a key meeting in Durham, South Africa (April, 1998) as follows:

“The meeting operated as two parallel working groups that focused on test questionnaire construction. The test construction group completed a comprehensive analysis of the official curricula, school syllabi, textbooks, and examinations that were used in SACMEQ countries. This analysis was used to construct test blueprint as frameworks for writing a large pool of test items for pupils and teachers in both reading and mathematics. The questionnaire group concentrated on using the Dummy Tables to guide the construction of questionnaires for pupil, teachers, and school heads.”

This development process was followed by consultations and discussions with country teams to finalize the instruments. Dr. Ken Ross, IIEP SACMEQ staff, reported that construction of SACMEQ II tests covered a period of around nine months and described the instrument development process in an e-mail communication (May 19, 2008) as follows:

“The test construction process involved 15 national committees consisting of reading specialists, curriculum experts, educational researchers/planners, and experienced teachers. These committees were responsible for drawing together and structuring the curriculum materials and assessment tasks that were used to construct the SACMEQ test blueprints and to select a large set of test items for trial testing. Communications between IIEP and SACMEQ countries (via training workshops and email messages) during the whole process ensured that a balanced analysis of curricula was undertaken and operationalized.”

This consultation, review and revise process for test framework development and test construction is critical for effective and accurate data collection in the participating countries. The curriculum match between tests and school curricula for reading and mathematics determine the degree to which appropriate and meaningful interpretations of SACMEQ data can be made within and across countries.

How were the test items written?

After test blueprints had been developed, the NRCs worked in teams to either select or write all of the required test items for the SACMEQ tests. Some items were written specifically for SACMEQ, while others were selected from prior assessments such as the Third International Mathematics and Science Study (TIMSS). Items were classified according to the cells in the test blueprints. For each cell twice as many items as required were prepared so that the rejection of poor items after the trial testing did not result in a shortage of items in some cells. Test items were in multiple-choice format with four options per item.

At the end of the Durban meeting, all of the questionnaires, test instruments and draft manuals had been prepared. A reading test blueprint was prepared and mapped to three domains (narrative, expository and documents reading) and 5 skill levels. The math test was blueprinted with three domain areas (number, measurement, and space-data). Test blueprints and descriptions of the math and reading levels are given in the country reports' appendices.

The item pools were sent to all countries so that each country had the opportunity to review the items by its own curriculum specialists. Based on this, panel members made recommendations for editorial changes to existing items and for additional items to ensure that test items met the requirements of the respective national curricula. Documentation of such curriculum match constitutes very important content related validity evidence. This documentation is not available for SACMEQ, even though curricular match considerations were included in the test construction process. We recommend SACMEQ to develop and make publicly available comprehensive documentation of content validity evidence for the reading mathematics tests at the beginning of the assessment development process in each phase. This content validity evidence needs to be revised periodically to reflect possible changes in schooling and curricula of the countries.

The SACMEQ II test construction was based on the use of test items from four sources: the initial Zimbabwe testing, SACMEQ I, TIMSS, and the International Reading Literacy (IRL) Study. The SACMEQ II pupil tests were intended to be linked to and compared with reading scores from the

IRL study and mathematics scores with the TIMSS results, as well as to develop comparability with the SACMEQ I results.

The linking items between SACMEQ I and SACMEQ II tests appear to cover all content domains, but not necessarily in the proportions in which they appear in the tests. For example, while narrative items make up approximately one-third of the total items in the SACMEQ I assessment, they comprise only one-tenth of the linking items between SACMEQ I and the SACMEQ II pupil reading assessments. Given that one of the objectives of SACMEQ is to make comparisons concerning changes in educational quality and educational achievement between SACMEQ I and SACMEQ II, ensuring that linking of these two assessments is conducted accurately is important for meeting this objective. The implications of this limitation in the linking items need to be examined empirically to determine the degree of comparability between SACMEQ I and SACMEQ II scores.

Inclusion of large numbers of anchor (linking) items and having these items as content representative for both sets of tests is very challenging for all assessments. Assessment developers need to work with multiple constraints, such as testing time, fatigue, and need for changing assessments from one phase to another. As in all assessments, it is essential for test developers to conduct ongoing studies to examine accuracy, comparability and meaningfulness of scores on an ongoing basis. Our recommendation is to conduct an empirical investigation to examine the degree of comparability between the two phases of SACMEQ data.

There was item overlap between teacher and pupil tests to allow for measurement of the two groups on the same scale. For the 148-item “composite” reading test, there were 36 items that came only from the SACMEQ I pupil reading test, 52 test items that came only from the SACMEQ II pupil reading test, and 26 items that came only from the SACMEQ II teacher reading test. An additional 34 items were located in more than one test, with 9 of these items being located in all three tests, and 3 sets of items associated with pairs of tests. For the 91-item “composite” mathematics test there were 50 items that came only from the SACMEQ II pupil mathematics test, and 28 items that came only from the SACMEQ II teacher mathematics test. An additional 13 items were located in both tests.

Both the reading and mathematics teacher tests had sufficient number of overlapping items with the pupils' tests to create comparable score scales for these two groups.

How valid were the tests?

Content and construct related validity evidence

The country reports provide some evidence for construct validity of the tests. The statement is made that “the structure of the pupil tests was congruent with the content (domains) and behaviours (skills) derived from detailed analyses of the curricula, syllabi, examinations, and textbooks used in the SACMEQ countries”¹. This statement is supported to some degree by the fact that items were developed by representatives of each country, items were sent to countries for review by their curriculum specialists and for identification as “essential items” from their curriculum. However, no documentation is provided that demonstrates curriculum match between country curricula and the test content. In addition, beyond this level of content related validity evidence, no construct related validity evidence is provided. These types of validity evidence are essential for judging the quality of the assessment data and are required by the professional testing guidelines (AERA, APA, & NCME, 1999).

Cultural validity evidence

Three countries translated the SACMEQ II assessment from English into their own languages. Mozambique translated the documents into Portuguese, while Tanzania (Mainland) and Tanzania (Zanzibar) translated it into Kiswahili. In addition, countries were permitted to make adaptations to accommodate different notations. For example, changes were made in order to address the use of: “(a) a comma or a full stop for decimals, (b) a comma, a full stop, or a space for “separating” digits in numbers greater than or equal to 1000, (c) different currency units, (d) different nomenclature for grade levels, (e) different methods for expressing dates, and (f) 12 hour or 24 hour clocks for time”.

¹ From Technical Report, Chapter 2, THE CONDUCT OF THE SACMEQ II PROJECT file D04-053testrev2

Since no two countries used exactly the same conventions for items (a) to (f), it was necessary to prepare a unique set of data collection instruments and manuals for each country.

Current standards for translation and adaptation require test developers to provide evidence of test comparability in the resulting translated versions. The Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing, 1999) set out that when a test is translated from one language to another the methods used in establishing the adequacy of the translation should be described, and empirical and logical evidence should be provided for score reliability and validity of score inferences for the uses intended (Standard 9.7) and that when multiple language versions of a test are intended to be comparable, test developers should report evidence of test comparability (Standard 9.9). A description of the translation process used in SACMEQ II is presented in the Pilot Study Manual for National Research Coordinators, 1999 document. This document provides NRCs with guidelines regarding test translation and states that translations should aim to preserve equivalence of word and sentence structure difficulty, word frequency usage and use of synonyms. A basic description of the translation and adaptation processes is further given in the country reports for SACMEQ II. For example, the Zanzibar country report states the following:

“For Zanzibar, the questionnaires and the test items were translated from the English version into Kiswahili which was the language of instruction at primary level. Two different translators were used for this process. They translated the text and items separately and then compared their translation. Finally, a third translator was given the Kiswahili final version for translation from Kiswahili back into English. The back translations were deemed to be very similar to the original English version and hence the Kiswahili translations were also deemed to be of good quality².”

² Zanzibar Chapter 2 page 17

According to the Standards as set out above, the test developers have a duty to provide both empirical and logical evidence for test comparability between the different language versions. The Pilot Study Manual for National Research Coordinators, 1999 document further states that: “Other countries [besides Mozambique and Tanzania] wishing to provide their data collectors with translated copies of some questionnaires and/or the data collectors’ manual should feel free to do so.” (p.11³) These procedures may threaten comparability of measures as having questionnaires or manuals used by different countries that are different from each other may decrease countries’ ability to gather comparable data.

In SACMEQ assessments, test items were administered to pupils from a diverse set of cultures. Measurement research conducted by the two evaluators on this team and others in the measurement community indicates that pupil performance may be affected by the contexts in which test items are presented even though no previous knowledge about the context is required by the items (Ercikan, 2005; Solano-Flores & Trumbull, 2003). For example, a reading comprehension item with a passage on skiing may prove to add unnecessary cognitive burden to students who are not familiar with this sport. Therefore, cultural validity evidence is needed in order to make appropriate and meaningful interpretations of test results for tests administered to pupils in different cultures and in particular in different languages. Typically, cultural appropriateness evaluations are conducted by a broad sensitivity review panel consisting of individuals who are knowledgeable about education and the role of language and culture in cognitive processes. Cultural validity evidence needs to be added to efforts of validating SACMEQ assessments and instruments. We were not able to review test items to assess their degree of cultural sensitivity due to access restrictions to these items.

Were pre-tests conducted?

A “small-scale trial testing” of the instruments and manuals was conducted in February of 1999 in 10 schools near Harare, Zimbabwe to coincide with a planning and training meeting for NRC’s at that

³ This document is on the SACMEQ website

time. This provided NRC's with practice and experience in data collection and data entry. Data from this testing was expected to be used to make improvements to the tests, questionnaires, and manuals, with input from external consultants. The evaluation team was not able to find information about the psychometric qualities of the items on the small-scale trial or about how the information gained was used to make improvements to the tests.

A "large-scale trial testing" of instruments and manuals was conducted in August and September, 1999. The NRCs and other Ministry staff were provided with training in computer-based data entry and data cleaning techniques. The reports indicate that "more than 400 schools and 8000 pupils" from participating SACMEQ II countries were involved in the data collection. A meeting of SACMEQ II NRCs and their Deputies was held at the IIEP in October, 1999 in order to analyze the trial test data from the reading and mathematics performance of pupils and their teachers in order to select the test items for the main data collection.

Data

What kinds of analyses were carried out on the trial data?

The reports indicate that the data from the trial-testing phase "were subjected to Rasch and classical item analyses in order to detect items that did not "fit" the relevant scales, or that were "behaving differently" across subgroups of respondents defined by gender and country. The poor quality test items were rejected – keeping in mind the need to prepare a "balanced" test across skill levels and domains."

However, details of these analyses have not been provided in the reports. The Rasch and classical item analyses were also conducted a second time after the main testing. Although the data from the trial-testing phase were said to be subjected to Rasch and classical item analyses, no specific information regarding the analysis results has been found in the documentation provided to this team. The scale creation and analyses are conducted centrally at the SACMEQ headquarters, jointly with key IIEP staff. Detailed documentation of trial data as well as final data need to be created and made

publicly available. Public availability of such technical reports is very important for users of the SACMEQ data to be able to make judgments about the quality of the data.

Was the item-writing process convincing?

As noted above, each country had the opportunity to have input into the writing process, and to ensure that the country's curriculum was adequately represented in the items. It is important to involve in each country a wide range of curricular experts in the areas of reading and mathematics in the item development and selection process. This wide representation of curricular expertise enhances the quality of decisions made related to test content by country teams. In an e-mail communication (May 19, 2008), Dr. Ken Ross reported that:

“Each SACMEQ country had its own technical committee plus a special committee for test development. This second committee had representatives drawn from specialists in subject matter, experienced teachers, and education planners and researchers”

Such information about item writing needs to be documented. The membership on these technical communities and the processes they used to write and review items is critical to content and construct validity of the score interpretation from SACMEQ.

In addition to selecting test items, the NRC's determined the skill levels for the items “through a Grade 6 pupil's eyes”⁴. One potential area for improvement would have been to more adequately test the trial items on grade 6 students through cognitive laboratories or think aloud protocols which can be used to determine the processes students experience when constructing their responses. This approach can result in a more realistic determination of pupils' interpretation of test items and the cognitive processes they engage in when they respond to test items, as well as provide empirical evidence on skill level identification.

⁴ Kenya report, chapter 2 page 37

Were the test reliabilities at an acceptable level?

Specific reliability information in the form of reliability coefficients (such as an internal consistency coefficient) has not been provided in the country reports. However, the SACMEQ II reading and mathematics pupil assessments (83 and 63 items respectively) are of sufficient length to provide high degree of reliability evidence. The evaluation team has conducted reliability studies using data from some of the participating countries which demonstrated high internal consistency levels. These reliability estimates need to be calculated and reported for each country and each test in order for users of data to be informed about measurement accuracy in these tests.

It is now commonly accepted that standard error of measurement (sem) at the test level is not a good indicator of measurement accuracy at different score levels. Item response theory based scaling methodologies, such as the Rasch model used in SACMEQ, allows estimation of sem for different score levels. Given that performances on SACMEQ tests may vary a great deal across countries, measurement accuracy may also vary for these countries. Therefore, we recommend reporting of sem for a selected set of scores on different parts of the scale to allow examination of measurement accuracy at different score points and possibly for different countries.

Sampling errors were calculated based on the study design as is appropriate in large-scale assessments. The IIEP's specialized sampling software (IIEPJACK) was used to make calculations of sampling errors. However, it should be noted that sampling error is only one source of error, other sources of error, a key one being measurement error needs to be reported in order for SACMEQ data users to judge their accuracy and possible effect of these accuracies/inaccuracies on data analysis and results.

Was the data collection well conducted?

Prior to the main data collection, data collection training sessions were held at which data collection team leaders were trained. The main SACMEQ II data collection occurred for 12 of the 15 SACMEQ Ministries of Education in the period of September to December, 2000. The Mauritius data collection was completed in July, 2001 and the Malawi data collection in September, 2002. Data

collectors were provided with data collection manuals which provided instructions for processes leading up to, during, and after data collection.

Data were entered by hand in SACMEQ data collection efforts. This is a tremendously time-consuming, error-prone way of entering data. One data entry scenario is described in the technical report as below:

8 people from Zanzibar Ministry of Education were trained in the use of WINDEM for data entry. A ten person-team (university students, teachers and data centres keyboard operators) was recruited and trained in the use of WINDEM, a special data entry package used by SACMEQ to enter all data. The data entry took about 2 months. At the end of this procedure the data files were sent by email to the unit 'Monitoring Educational Quality' at the IIEP in Paris. Many consistency checks were made for many variables as well as for the identification codes used. The IIEP team had many queries. The first data files were sent to Paris in February 2001 and after nine to-ings and fro-ings the files were finally declared to be clean on 27 January 2003. (ch.2, p.21)

In addition to training individuals in the data entry effort, there were up to 31 iterations of data check and verification between the countries and the IIEP⁵. In the most recent SACMEQ III assessment, a week long workshop was spent on data clean-up training. Even though there are benefits of engaging NRCs with actual assessment data, data capturing could have been done more efficiently by using scanners. Undeniably, there is a need for training on how to use scanners and a specifically developed software, tremendous savings in time, cost, and in accuracy of data could be achieved by using this alternative method of data capturing.

⁵ Technical Report, file D04-053

How was the sampling carried out?

Our review of documentation revealed that the sampling plan followed in the SACMEQ II study consisted of the following desired target population: *“All pupils at Grade 6 level in 2000 (at the first week of the eighth month of the school year) who were attending registered mainstream primary schools.”*

The exclusion rules were similar, but not identical for all countries. For example, all countries excluded schools with too few (less than 10, 15, or 20) grade 6 students, and special schools. In addition, Namibia excluded inaccessible schools, Malawi excluded inaccessible schools and private schools, and Uganda excluded schools affected by serious military conflicts. By excluding private schools, Malawi’s results may not be strictly comparable to those of the other countries. Also, there is no unified description of “special schools” presented in the documentation, and the manner in which each country excluded schools based on this criterion may also lead to incomparability of results across countries. Such differences in exclusion rules across countries are not unique to SACMEQ and other international assessments need to accommodate logistics and conditions of the participating countries as well. Two issues need to be addressed regarding the possible incomparability created by different exclusion rates. The first is explicit discussion and warning of users of SACMEQ data of such incomparability across some countries perhaps under limitations of SACMEQ data. The second is, revisiting this issue with the participating countries in the next phase of SACMEQ assessments to establish more comparable exclusion rules.

The sample designs used in the SACMEQ II Project were intended to meet the standards set down by the International Association for the Evaluation of Educational Achievement which require pupil population parameters to have sampling accuracy at least equivalent to a simple random sample of 400 pupils⁶. Sampling included both explicit (“regions”) and implicit (“school size”) strata. A

⁶ Technical Report, file D04-053phasesrev2

minimum cluster size of 20 students⁷ per school selected by simple random sampling was chosen. The number of schools per region to be sampled was determined through the use of the intra-class coefficient (“rho”). Based on rho values of 0.3 to 0.6 in this study, the required number of schools per country ranged from 134 to 248. In the Seychelles and Zanzibar all schools were included in the defined target population.

The sampling design consisted of a two-stage design. The first stage consisted of using probability proportional to size (PPS) sampling of schools within strata and the second stage consisted of the selection of a simple random sample of a fixed number of pupils within selected schools. No replacement was allowed for absent students.

The SACMEQ II goal was to achieve overall response rates of 90% for schools and 80% for pupils. This was accomplished with the exceptions of Tanzania and Zambia which experienced loss of data within schools resulting in pupil response rates of 77% and 75%, respectively.

It is important to note that 7 of the 14 countries (Kenya, Lesotho, South Africa, Swaziland, Tanzania, Uganda and Zambia) that participated in SACMEQ II had effective sample sizes of less than 400. In the SACMEQ II Project, two school systems, South Africa and Uganda, fell far below the required target of an effective sample size of 400 pupils. In South Africa the values were 185 and 230 for reading and mathematics, respectively, and in Uganda the values were 222 and 176 for reading and mathematics, respectively.” These sample sizes are very low and have very high levels of sampling error associated with them. Such high levels of sampling error jeopardize capturing statistically meaningful patterns in the data.

⁷ Technical report, from pages 25-39 of d04-053samplingrev2

Questionnaires

Do the questions provide complete coverage for the study's overall aims?

A review of the specific research questions that guided this study and the contents of the SACMEQ II questionnaires (pupil reading, pupil math, teacher, and school head) revealed that the questionnaires provided excellent coverage of the study's overall aims. There are two issues that may need further consideration regarding the pupil questionnaire. One is the level of reading ability required to respond appropriately to some of the questionnaire items, for example, the use of the terms "corrugated iron" and "asbestos sheets". This level of vocabulary and word difficulty may be difficult for respondents to understand. It may also lead to comparability difficulties as it may be difficult to translate these words and to find synonyms of these words in other languages. In addition, pupils' may not know the answers to some questions such as regarding whether payments are made for extra tuition, and if so who makes those payments and how much is paid.

How were the background questionnaires developed?

Questionnaires were developed at the meeting of National Research Coordinators (NRCs) which was held in Durban, South Africa in April 1998, along with the achievement test items⁸. At this meeting a group of NRC's (the "questionnaire group") concentrated on using the Dummy Tables that were based on key policy issues identified by the countries to guide the construction of questionnaires for pupils, teachers, and school heads.

Relevance and adequacy of organizational arrangements

What UNESCO inputs have gone into the development of the various UNESCO outputs? Is the level adequate to produce and sustain high quality outputs?

IIEP-UNESCO provides a comprehensive set of inputs into SACMEQ. These include the following:

- A centralized coordination of activities of a large consortium;
-

⁸ Technical report, file D04-053phasesrev2

- Expertise through UNESCO staff as well as consultants who provide ongoing technical support;
- Timely capacity building efforts in the form of hands-on workshops;
- Instrument development efforts;
- Software;
- Files for computer-based data entry;
- Software tools for computer-based data cleaning;
- Software tools for selecting scientific probability samples;
- Continued support in the form of consultation;
- Manuals and training for country teams for all processes and procedures;
- Coordination of development of instruments;
- Help in development of policy questions;
- Support (model country reports) for writing country reports.

The level of UNESCO inputs is excellent for quality implementation of SACMEQ activities. The observations at the Namibia workshop in December 2007 demonstrated a very well organized workshop with utmost care for detail, respectful, and collegial atmosphere, and collaborative partnerships among country participants and with IIEP staff. The timing of the workshop was expressed as being “great” for the participants. They also expressed that the skills they gained in similar workshops was tremendously helpful for them in their engagement with other projects.

To what extent do UNESCO outputs meet the technical needs and requirements of programme beneficiaries?

The country teams Dr. Ercikan met in Windhoek, Namibia expressed utmost satisfaction with the support they received from IIEP on an ongoing basis. The support they received not only met their technical needs specific to SACMEQ, they expressed that the training they received has been helpful for them as professionals as well as for their countries in general.

What are programme beneficiaries' needs and requirements with regard to monitoring and assessing learning achievement?

The participating countries need up to date data on achievement levels of their students at key grades and quality data to determine where the most urgent needs for policy actions are. The focus on Grade 6 seems to fit the needs of the participating countries, though, data on other grades may also be desirable. There is a systematic effort in SACMEQ to identify policy issues across countries and to develop background questionnaires to best inform these questions. The participants reported that currently the data they received from SACMEQ met their needs.

Effectiveness of the quality control and quality assurance procedures

SACMEQ has highly organized and structured quality control and quality assurance procedures in place. The key elements of SACMEQ quality assurance are as follows:

- Documentation and dissemination of schedules;
- Preparation and dissemination of progress reports and upcoming activities ;
- Timely capacity building activities that correspond with the needs of the beneficiaries, for example the Data Cleaning workshop in Namibia coincided with when the countries were about to engage in data cleanup activity in their countries;
- Software to enable and check the quality of sampling;
- Software to check the data entry and clean-up procedures;

- Guidelines and training on how to write country reports;
- Verification of data accuracy at IIEP.

The country reports we have reviewed and reliability analysis we conducted have not identified any errors. Given the limited time and resources allocated to this evaluation it is not possible to determine all possible errors and possibilities of errors. However, we expect that the procedures themselves and the care that is evident in everything we have been able to review speak very favorably for the quality assurance procedures.

Is a monitoring and evaluation system in place to enable Headquarters to monitor the progress of the programme? How effectively has the monitoring and evaluation system functioned?

This evaluation is the first evaluation of SACMEQ as a separate programme by UNESCO headquarters. A previous evaluation of IIEP which included SACMEQ was conducted in 2006. Both of these evaluations have been conducted by external evaluators (two different groups) on a one-time basis during a limited amount of time. International assessments and data collection efforts can benefit from having a monitoring and evaluation system to enable UNESCO Headquarters to monitor progress of the programme on an ongoing basis. A key recommendation we would like to make as a monitoring and evaluation system from UNESCO is implementation of an ongoing external review process. This can be in the form of an international technical advisory board, consisting of psychometricians, education policy experts, and statisticians among others, with international reputation. This board needs to have an arms-length relationship with the project and will review activities, processes, and products and provide a report to UNESCO Headquarters on a yearly basis. In almost all international assessments we are familiar with, such committees and boards are an integral part of quality assurance and collaborative improvement of processes. UNESCO's support of such a committee of experts will go a long way in assuring quality and revision of procedures according to up to date professional standards.

Results

What are the expected results of SACMEQ?

The expected results of SACMEQ correspond to their two goals:

- expand opportunities for educational planners to gain the technical skills required to monitor and evaluate the quality of basic education;
- generate information that can be used by decision makers to plan and improve the quality of education.

The first goal implies development of professionals and capabilities in order for countries to develop their assessments and data collection efforts. The interviews with country managers indicated that individuals involved in SACMEQ gained valuable skills that they used in other assessments of learning. Specific impact areas of SACMEQ were in: (1) developing important technical skills related to the design and implementation of large-scale data collections; (2) application of a wide variety of computer-based techniques for the preparation, management, analysis, and reporting of educational planning data.

Similarly, the second goal implies the use of SACMEQ data to inform policy decisions. The SACMEQ projects (I, II, and III) were designed to address specific policy questions identified to be important for the participating countries, and the data resulting from the study are intended to be used to make policy suggestions. The evaluation team examined the country policy reports for four countries (Malawi, Namibia, Seychelles and Botswana) who participated in SACMEQ II for the purpose of evaluating how the SACMEQ data was used for informing policy decisions. Each of these reports contains a chapter which describes the research-based policy suggestions that have arisen from the SACMEQ II study. These chapters are intended to draw a link between the research study and potential policy development in each participating country. The policy suggestions are said to flow directly from the results of the study, and are grouped into five main categories (consultation with staff, community members and experts; reviews of existing planning and policy procedures; further

data collection; initiation of large-scale educational research studies; and investment in infrastructure and resources). Each policy suggestion is further classified according to a relevant department, the cost of the policy's implementation, and an estimated length of time required to implement it. Such an approach to classifying policy suggestions is very practical and of great potential benefit to the participating countries in determining which policy suggestions to act upon.

There is good evidence that the policy suggestions do indeed flow from the data obtained for each individual country. For example, the SACMEQ II results for Malawi indicate that students in Standard 6 were approximately 174 months of age, which exceeds the expected age by nearly 22 months. Following from that data, the first policy suggestion for Malawi was to increase awareness of the importance of starting school earlier. Further, and also related to age of students in Standard 6, 66% of the students in Malawi reported that they had repeated a grade. A policy suggestion was made to reduce class repetition and to understand the factors related to repetition, particularly in those areas where class repetition was highest.

The Honourable Nangolo Mbumba, Minister of Education for Namibia, gives further evidence of the usefulness of the SACMEQ studies to the participating Ministries of Education. In his opening speech at the *Working Meeting for SACMEQ National Research Coordinators on Computer Based Data Entry and Data Cleaning* for the SACMEQ III Project at Windhoek, Namibia in December of 2007, Minister Mbumba provided confirmation of the impact of the SACMEQ studies for his country. He indicated that the results of the SACMEQ II study were used to formulate regional operational plans that have since been incorporated into the Main Strategic Plan of Namibia's entire education sector. Further, SACMEQ results have been integrated into the overall programmes of Namibia's Ministry of Education.

The SACMEQ I Report for Malawi (Milner, Chimbombo, Banda and Mchikoma, 2001) makes explicit policy suggestions (Policy Suggestions 4.6a and 4.6b) related to analyzing the training needs of teachers and developing policy related to increasing the amount of teacher training. In a 2004 report prepared for the 47th Session of the International Conference on Education and compiled by the

Ministry of Education and the Malawi National Commission for UNESCO, it is noted that the Malawian Government had developed and was implementing strategies related to teacher education and development. Item 2.1.4.1 of the report sets out that the Government's goal is to increase the number of qualified teachers through appropriate teacher training at various levels.

Inappropriate interpretations of data can pose a risk of jeopardizing the effectiveness in informing policy decisions. The Standards for Educational and Psychological Testing set out that when average scores for groups are reported they should be accompanied by information about the sample size and shape or dispersion of the score distributions to enable audiences to make meaningful interpretations of the results. Currently, differences between groups (such as gender) are reported without any indication as to whether the differences are statistically significant. This lack of information presents a potential risk for misinterpretation of the data. The evaluation team therefore recommends that all future SACMEQ country reports provide statistical significance levels for group comparisons. This recommendation is in keeping with current best practices in educational measurement as outlined in the Standards for Educational and Psychological Testing.

What results have been achieved?

In July 2004, the SACMEQ Ministers of Education decided to allow educational planners and researchers from around the world to have access to the SACMEQ Data Archives. IIEP staff report that this was a turning point for SACMEQ and since then SACMEQ has developed a reputation for research and training excellence within Africa and many other countries.

There has been a great degree of interest in using the SACMEQ Data Archives for conducting educational policy research studies via secondary data analyses, and for the preparation of postgraduate theses. This growth in interest is evident in the following developments (reported by IIEP in a document titled SACMEQ's Impact: The Use of Its Research and Training Resources , January 2008): *“(a) in July, 2004 a Google search for the “SACMEQ” acronym provided around 250 “hits”, and (b) by May, 2007 this figure had grown to over 21,000 hits. In addition, in the new Google “Scholar” option (which targets published reports and articles) the “SACMEQ” acronym*

now results in a listing of over 200 reports and publications.” On January 25, 2008 Google search for “SACMEQ” identified 5,390 hits. This difference may very well be due to changes in the Google search algorithms. On May 19, 2008, IIEP staff reported that AltaVista search engines reported around 21,100 hits for searching on “SACMEQ” on 2 May 2008.

The following use of SACMEQ research and training resources by ministries, international organizations/bilateral, conferences, researchers/planners, and postgraduate students are identified in the IIEP (SACMEQ’s Impact: The Use of Its Research and Training Resources, January 2008):

1. Use by the SACMEQ Ministries of Education: Some Examples

Three examples of reviews of SACMEQ's impact on policy and planning within Ministries of Education are:

- Murimba, S. (2005). The impact of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ). *Prospects*, XXXV (1), pp. 91-108.
<http://www.springerlink.com/content/7126vx7082543x83/>
- Nzomo, J. and Makuwa, D. (2006). How can countries move from cross-national research results to dissemination, and then to policy reform? In K. N. Ross and I. Jurgens-Genevois (Eds.) *Cross-national studies of the quality of education: Planning their design and managing their impact* (pp. 213-228). Paris: IIEP (UNESCO).
http://www.unesco.org/iiep/PDF/pubs/G117.pdf?class=IIEP_PDF_enhp&page=G117&estat_url=http://www.unesco.org/iiep/PDF/pubs/G117.pdf
- Oduol, T. (2006). Towards the making of education policy in Kenya. *International Education Journal*, 7(4), 466-499.
<http://ehlt.flinders.edu.au/education/iej/articles/v7n4/Oduol/paper.pdf>

2. Use by International/Bilateral Organizations: Some Examples

SACMEQ resources have been used by organizations such as UNESCO, World Bank, Asian Development Bank, South African Development Bank, DFID, etc. for discussion and evaluation of policy options related to the quality of education. Some illustrative examples are presented below.

- *The 2005 EFA Global Monitoring Report (pages 46, 47, 54):*
http://portal.unesco.org/education/en/ev.php-URL_ID=35939&URL_DO=DO_TOPIC&URL_SECTION=201.html
- *The 2006 EFA Global Monitoring Report (pages 58, 61):*
http://portal.unesco.org/education/en/ev.php-URL_ID=43283&URL_DO=DO_TOPIC&URL_SECTION=201.html

- *The 2007 EFA Global Monitoring Report (pages 50, 51):*
http://portal.unesco.org/education/en/ev.php-URL_ID=49640&URL_DO=DO_TOPIC&URL_SECTION=201.html
- *The 2008 EFA Global Monitoring Report* http://portal.unesco.org/education/en/ev.php-URL_ID=49591&URL_DO=DO_TOPIC&URL_SECTION=201.html
- *The World Bank's World Development Report 2007 (page 48)* http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2006/09/13/000112742_20060913111024/additional/ch2.pdf
- *The World Bank "EdStats" Information Resources Page*
http://devdata.worldbank.org/edstats/los_app.asp
- *Asian Development Bank Review of New Challenges for Educational Development*
<http://home.hiroshima-u.ac.jp/cice/hirosato4-2.pdf>
- *Development Bank of South Africa Policy Research Studies*
<http://www.dbsa.org/other/educonference/Van%20der%20Berg&Louw.doc>
- *DFID's New "School Effectiveness and Quality" Project at Bristol University*
<http://www.edqual.org/research/school.html>

What factors have contributed to and/or prevented the achievement of expected results?

All the countries we have been able to talk to reported great success levels in achieving their expected results.

Risks

We considered 4 key factors as potential risks that might jeopardize the outcomes and achievement of the goals of SACMEQ. Some of these factors are as follows:

- Cultural appropriateness of tests;
- Accuracy of measurement and representation of skills and competencies in some countries;
- Representativeness and comparability of samples;
- Comparability of data and scores across countries;

The current documentation did not provide any cultural validity evidence for SACMEQ tests and questionnaires. Review of test questions or questionnaire items by country teams is not sufficient to ascertain cultural validity. In addition, the outcomes of the review processes for cultural, construct, or content related validity evidence need to be documented and made publicly available.

SACMEQ reports are diligent about reporting sampling error in the results. However, equally important is reporting of measurement error for different countries, for different score ranges. Both sampling error and measurement error provide information about the error involved in comparing groups within and across countries, as well as precision in interpreting relationships between educational inputs and learning outcomes.

Even though the target population is fixed in SACMEQ sampling, Grade 6 pupils, the samples may have incomparability due to two key factors. One factor is different exclusion rules (e.g., private

schools) in some countries. The second factor is the variation in ages of the Grade 6 students. Both of these can affect how the results can be interpreted within and across countries.

The fourth risk (comparability of data and scores across countries) is due to a combination of the first three risks (cultural validity, construct representation and accurate measurement of constructs and comparability of samples) and may be mitigated by addressing these issues. All four of these risk factors can be addressed by SACMEQ by providing comprehensive validity evidence that includes evidence of construct and content related validity evidence, cultural validity evidence, measurement accuracy at different score points, and evidence for consequential aspects of validity and by highlighting and describing implications of limitations in comparability of samples across countries. Such validity evidence is also required by the Standards for Educational and Psychological Testing (AERA, APA & NCME, 1999).

Final Comments

SACMEQ is a well established successful cross-national data collection and capacity building effort. Overall, the technical soundness of the SACMEQ project is of good quality. The aims of the study are clearly stated and are relevant for all participating countries, and the defined target population is appropriate. The design of the study allows for the data to be used in the manner in which it is intended. The achievement tests have been built upon a solid framework and blueprinting process that were designed to address the policy questions expressed by the participating countries. The content of the SACMEQ II questionnaires (pupil, teacher, and school head) provides excellent coverage of the study's overall aims.

The relevance and adequacy of organizational arrangements has been demonstrated and the level of UNESCO input has been shown to be adequate for quality implementation of SACMEQ activities. Participants in workshops have expressed great satisfaction with the training and technical support that they receive. Evidence from these participants indicates that the data they receive from SACMEQ meets their needs for policy development.

With regard to effectiveness of the quality control and quality assurance procedures, SACMEQ has highly organized and structured quality control and quality assurance procedures in place. The procedures themselves and the conscientiousness of the IIEP/SACMEQ staff are evident in everything the evaluation team has reviewed and speaks very favourably for the quality assurance procedures.

In terms of its goals of providing training and information to the participating countries, SACMEQ has developed a reputation for research and training excellence within Africa and in other countries. There has been a great degree of interest in using the SACMEQ Data Archives for conducting educational policy research studies via secondary data analyses, and for the preparation of postgraduate theses.

SACMEQ has taken a very key step in reacting to the high priority needs of the continent by developing and including an HIV/AIDS Health Knowledge Test. The data gathered through this instrument can provide extremely valuable avenues for educational interventions in the area of HIV/AIDS. The key people involved in SACMEQ, IIEP staff as well as the country participants, have great commitment to the project, as was evident in all conversations, discussion and meetings held.

The evaluation team has identified a few areas which require further evidence and documentation to support the valid use of and inferences from the SACMEQ data. There are potential sources of incomparability in SACMEQ. Due to the fact that the achievement test was administered in many countries and in three languages, potential sources of incomparability between the countries arise due to cultural, language and age differences between the country samples. Efforts should be made to reduce these sources of incomparability. The tests need to be reviewed for cultural and linguistic appropriateness. These validity evidence gathering procedures need to be built into instrument development activities from initial test development stages. In addition, since one of the objectives of SACMEQ is to make comparisons concerning changes in educational quality and educational achievement between SACMEQ I and SACMEQ II, and SACMEQ III ensuring that linking of these two assessments is conducted accurately will be important for meeting this objective. In order to meet the requirements of the Standards for Educational and Psychological Testing, empirical and logical

evidence should be provided for score reliability and validity of test use and inferences, and evidence of test comparability should be reported.

It has not been possible for the evaluation team to conduct a complete validity evaluation of SACMEQ. This situation has arisen because the team was not provided with copies of the SACMEQ reading and mathematics assessment items, and therefore we do not have sufficient documentation on which to provide a full evaluation of validity evidence.

Although SACMEQ has a dedicated team of professionals committed to quality control and quality assurance procedures, a key recommendation we would like to make to improve the monitoring and evaluation system is the implementation of an ongoing external review process. Such an external review process has become common practice in large scale assessments, and may take the form of an international technical advisory board, consisting of psychometricians, education policy experts, and statisticians among others, with international reputation. The board would have an arms-length relationship with the project and would review activities, processes, and products and provide a report to UNESCO Headquarters on a yearly basis. UNESCO's support of such a committee of experts would be very valuable in assuring quality and revision of procedures according to current professional standards. The key people involved in SACMEQ, IIEP staff as well as the country participants, have great commitment to the project, as was evident in all the personal conversations, discussion and meetings Dr. Ercikan has had. This enthusiasm and commitment need to be combined with ongoing input and feedback from external experts who are not involved in the day to day activities of the project.

References

- AERA, APA, NCME, (1999). *The Standards for Educational and Psychological Testing*. American Educational Research Association, Washington DC.
- Ercikan, K. (2005). Developments in assessment of student learning. In P. Winne & P. Alexander (Eds.), *Handbook of educational psychology* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Solano-Flores, G., & Trumbull, E., (2003). Examining Language in Context: The Need for New Research and Practice Paradigms in the Testing of English-Language Learners, *Educational Researcher*, 32, 2, 3-13.

Appendix A: List of Documents Reviewed

- IIEP (2003). Codebook of SACMEQ II variables. Paris: IIEP.
- IIEP (2003). Link from Tables to policy questions. Paris: IIEP.
- IIEP (2003). Link from policy questions to tables. Paris: IIEP.
- IIEP (2003). Test item link among pupil-teacher-SACMEQ I. Paris: IIEP.
- IIEP (2003). Data matrix of linked test items. Paris: IIEP.
- Ministry of Education Botswana. (2003). Pupil-level merged file for Botswana in SPSS format.
- Ministry of Education Kenya. (2003). Pupil-level merged file for Kenya in SPSS format.
- Ministry of Education Mozambique. (2003). Pupil-level merged file for Mozambique in SPSS format.
- Ministry of Education Zanzibar. (2003). Pupil-level merged file for Zanzibar in SPSS format.
- Moloi, M., & Strauss, J. (2005). The SACMEQ II Project in South Africa: A Study of the Conditions of Schooling and the Quality of Education. Harare: SACMEQ.
- Moyo, G. et al (1993). A Southern Africa proposal for monitoring progress towards attaining the goals of the EFA Jomtien Conference concerning the quality of education. Paris: IIEP.
- Nassor, S., Abdallah, M., Said Ali, O., & Salim, M. (2005). The SACMEQ II Project in Zanzibar: A Study of the Conditions of Schooling and the Quality of Education. Harare: SACMEQ.
- Nassor, S., & Mohammed, K. A. (1998). The quality of education: some policy suggestions based on a survey of schools. SACMEQ Policy Research No. 4: Zanzibar. Paris: IIEP.

Nzomo, J., et al. (2001). The quality of education: some policy suggestions based on a survey of schools. SACMEQ Policy Research No. 6: Kenya Paris: IIEP.

Onsumu, E., Nzomo, J., & Obiero, C. (2005). The SACMEQ II Project in Kenya: A Study of the Conditions of Schooling and the Quality of Education. Harare: SACMEQ.

Passos, A., Nahara, T., Magaia, F., & Lauchande, C. (2005). The SACMEQ II Project in Mozambique: A Study of the Conditions of Schooling and the Quality of Education. Harare: SACMEQ.

Ross, K.N., Saito, M., Dolata, S., Ikeda, M, Zuze, L., Murimba, S., Postlethwaite, T.N. and Griffin, P., Chapter 2 Methodology for SACMEQ II Study. Paris: IIEP.

SACMEQ (2000). All the questions and their unweighted univariate analyses for all variables contained within the data archives developed for SACMEQ's initial educational policy research project. Unpublished document.

SACMEQ NRCs (1995, January). Five "high priority" policy questions about the quality of education and their transformation into associated specific research questions and dummy tables. Unpublished working paper.

SACMEQ NRCs (1995, January). The purpose of the SACMEQ project. Unpublished working paper

SACMEQ National Coordinators. (2000). Data Collection instruments for SACMEQ II main study. Harare: SACMEQ

SACMEQ National Coordinators. (2000). Manual for NRCs. Harare: SACMEQ

SACMEQ National Coordinators. (2000). Manual for data collectors. Harare. SACMEQ

SACMEQ National Coordinators. (1999). Data collection instruments for pilot study. Harare:

SACMEQ

Voigts, F. (1998). The quality of education: some policy suggestions based on a survey of schools. SACMEQ Policy Research No. 2: Namibia. Paris: IIEP.

Appendix B: Management Response

Recommendations	If not accepted, give reasons for rejection
<p>1 We recommend to SACMEQ to provide guidelines to the participating countries and highlight the limitations of interpretability of data across countries due to differences in age distributions and exclusion rules in the SACMEQ samples.</p>	<p><i>On 1 September 2008 the recommendations will be sent to the SACMEQ Scientific Committee with a request for comment and identification of appropriate action. The SACMEQ Acting Director will follow up on actions as necessary, with guidance as appropriate from the SACMEQ Managing Committee and the SACMEQ Assembly of Ministers.</i></p>
<p>2 We recommend to SACMEQ to develop and make publicly available comprehensive documentation of content validity evidence for the reading and mathematics tests. In addition to construct validity evidence that should include documentation on the readability and appropriateness of questionnaire items, we recommend to SACMEQ to conduct empirical studies to examine what the test questions are measuring using cognitive laboratories or think aloud protocols. We also recommend obtaining cultural validity evidence that includes documentation and analyses of cognitive processes from sensitivity review panels for participating countries and for examinees taking tests in different languages.</p>	
<p>3 The degree of comparability of test scores needs to be examined empirically and documented. Linking items should be based on content and psychometric consideration.</p>	
<p>4 We recommend reporting of reliability of tests as well as item response theory based standard error of measurement for a selected set of scores on different parts of the scale to allow examination of accuracy at different score points.</p>	
<p>5 We recommend conducting a systematic feasibility study to explore different types of data capturing instead of manual data entry of data.</p>	
<p>6 We recommend that all future SACMEQ country reports include statistical significance levels for comparison groups such as gender, ethnic or regional groups.</p>	
<p>7 We recommend to SACMEQ to refine the sampling design collaboratively with the participating countries to establish more comparable exclusion rules across countries and to obtain larger sampling sizes.</p>	
<p>8 We recommend to UNESCO Headquarters to create an international technical advisory board consisting of psychometricians, education policy experts, and statisticians among others in order to review activities, processes, and products and provide a report (on a yearly basis).</p>	