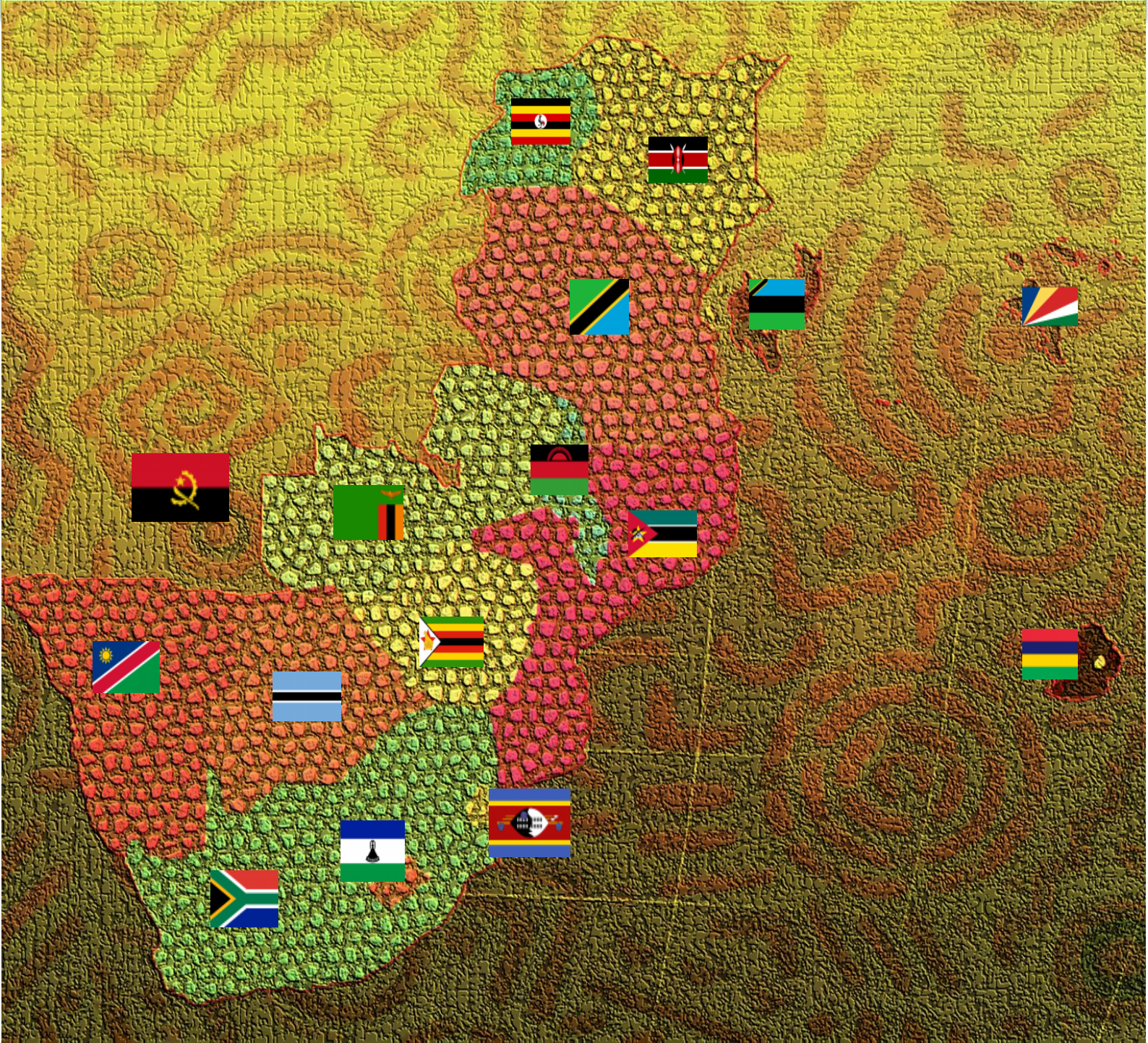




Southern and Eastern Africa Consortium for Monitoring Educational Quality

SACMEQ IV Project Technical Report



1.0 INTRODUCTION

The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SEACMEQ – formerly *SACMEQ*) was established as a capacity-building programme in the field of educational assessment. However, SEACMEQ has transformed over time into a collective developmental sub-continental organization with three main objectives to: (a) Provide training opportunities to build technical skills in research, monitoring and evaluation for Ministry of Education technocrats; (b) conduct co-operative policy research on condition of schooling and quality of basic education; and (d) disseminate research data and information to facilitate policy-dialogues among decision-makers and stakeholders. SEACMEQ strategizes to achieve its objectives by conducting large-scale international studies on the provision of basic education by the education systems of its member states.

Since the inaugural cycle in 1995, SEACMEQ (formerly *SACMEQ*) studies have provided the agenda for government actions related to *quality* of basic education and covering areas such as: educational inputs in schools; benchmark standards for educational provision; equity in allocation of resources; and achievement in literacy, numeracy, and health knowledge of Grade 6 pupils and their teachers. The fourth cycle; *SACMEQ IV*; for which this technical report is prepared commenced in 2012 and typically involved multiple activities of varying magnitude and complexity. The highlights of these activities are the following:

Project proposals – written to solicit international and national funding from development partners and member governments respectively;

Project steering committees – established at country level to oversee implementation of the study;

Policy concerns/questions – formulated to crystalize the problems to be addressed in the provision of basic education by the different systems;

Cross-national curriculum analysis – conducted to identify the different characteristics of test items and to synchronize curricular contents of participating education systems;

Test blueprint – developed as a result of curricular analysis and to ensure content validity of the tests;

Test and questionnaire item construction – aided by existing test papers, text books, syllabi, and SEACMEQ draft test and questionnaire items;

Manual for data collection – developed to ensure uniform and scientific data collection procedure in all education systems;

Pilot testing – of test instruments and questionnaires, including testing of *DataEntryManager* (DEM) system for data entry, entering and cleaning pilot test data, scaling tests, and producing and populating dummy tables;

Main data collection – preceded by finalizing data collection instruments and manuals, translating survey instruments to accommodate the languages of instruction in all participating education systems, drawing of final representative country samples, and train data collectors.

This report addresses technical issues that, in the field of research, have been accepted to be pivotal in assessing the robustness and credibility of a study. These issues are: aims, objectives and questions; target population; sampling; instrument construction and their pre-tests; translation of instruments into different languages; data collection; and data management. Although the issues mentioned will be discussed in this report, they will not necessarily be presented under individual headings, but rather under broader headers such as design, instrument development, data collection and entry etc. Since a large team of various professionals implemented *SACMEQ IV* study, their work and write-ups are frequently referred to and form part of this technical document.

2.0 SACMEQ IV GENERAL POLICY CONCERNS/QUESTIONS

Due to its collaborative networking and emphasis on capacity building through comprehensive and intensive-training programmes, SEACMEQ has uniquely nurtured a systematic approach of consulting with member governments and policy makers in order to identify and agree on the policy concerns and questions that need to be addressed through its research. The following are *SACMEQ IV* general policy concerns:

- i. What are the personal characteristics (for example, age and gender) and home background characteristics (for example, parent education, regularity of meals, home language, etc.) of standard 6 pupils that may have implications for monitoring equity, and/or that may impact upon teaching and learning?
- ii. What are the school context factors experienced by standard 6 pupils (such as location, absenteeism (regularity and reasons), standard repetition, and homework (frequency, amount, correction, and family involvement)) that may impact upon teaching/learning and the general functioning of schools?
- iii. Do standard 6 pupils have sufficient access to classroom materials (for example, text books, readers, and stationery) in order to participate fully in their lessons?
- iv. Do standard 6 pupils have access to library books and other technology related sources of information within their schools, and (if they do have access) is the use of these books and other technology related sources of information being maximized by allowing pupils to borrow the books?
- v. Has the practice of standard 6 pupils receiving extra lessons in school subjects outside school hours become widespread, and have these been paid lessons?
- vi. What are the personal characteristics of standard 6 teachers?
- vii. What are the professional characteristics of standard 6 teachers (in terms of academic, professional, and in-service training), and do they consider in-service training to be effective in improving their teaching?
- viii. How do standard 6 teachers allocate their time among responsibilities concerned with teaching, preparing lessons, and marking?
- ix. What are standard 6 teachers' viewpoints on frequency of assessment, and meeting and communicating with parents?
- x. What is the availability of classroom furniture (for example, sitting/writing places, teacher table, teacher chair, and bookshelves) and classroom equipment (for example, chalkboard, dictionary, maps, book corner, and teacher guides) in standard 6 classrooms?
- xi. What professional support (in terms of education resource centers, inspections, advisory visits, and school head inputs) is given to standard 6 teachers?
- xii. What are the personal characteristics of school heads (for example, age and gender)?
- xiii. What are the professional characteristics of school heads (in terms of academic, professional, experience, and specialized training)?
- xiv. What are the school heads' viewpoints on general school infrastructure (for example, electrical and other equipment, water, and basic sanitation) and the condition of school buildings?
- xv. What are the school heads' views on inspections, community input, problems with pupils and teachers?

- xvi. Have material and human resources (for example, classroom teaching materials, school facilities and qualified teachers) been allocated in an equitable fashion among the strata?
- xvii. What are the achievement levels and variations (among strata) of standard 6 pupils and their teachers in Reading and Mathematics?
- xviii. What are the Reading and Mathematics achievement levels of important sub-groups of standard 6 pupils (for example, gender, socio-economic status and location)?
- xix. What are the HIV and AIDS knowledge levels (for example, minimum levels, and desirable levels) of pupils and their teachers?
- xx. Do pupils, teachers and school heads have positive attitudes towards people infected with HIV and AIDS?

3.0 THE AIMS OF SACMEQ IV PROJECT

SACMEQ IV project represents a major increase in the scale and complexity of SEACMEQ's research and training programmes. The focus of the project was on conditions of schooling and the quality of education in fourteen school systems: Botswana, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania (Zanzibar), Uganda, Zambia, and Zimbabwe. Premised around the policy concerns listed in the preceding section, the consortium determined that the purpose of the project was to gather data and rigorously analyze them to generate information on **a)** the general conditions of schooling, **b)** the reading and mathematics achievement levels of Grade 6 learners and their teachers, and **c)** the knowledge that learners and their teachers have about HIV and AIDS.

4.0 PUPIL AND TEACHER TEST DEVELOPMENT FOR SACMEQ IV PROJECT

This section of the report describes the process of developing *SACMEQ IV* tests for pupils and teachers. Owing to the fact that SEACMEQ maintains and uses equivalent tests in all cycles of its studies, aspects of the development of the reading and mathematics tests are detailed in *SACMEQ II* technical report which is available in the SEACMEQ archives. For purposes of continuity, a description of the reading and mathematics domains, as well as reading and mathematics skills are presented. These are followed by a description of writing and analyses of new test items, and the compilation of the final items that constituted *SACMEQ IV* tests. The development of the HIV/AIDS knowledge test (HAKT) is outlined at the end of the section.

4.1 The Reading Competency Levels for SEACMEQ Studies

As mentioned above, SEACMEQ administers equivalent tests in all of its studies; and this necessitates that the test item development procedures be somewhat identical. During *SACMEQ II* and *SACMEQ III* projects, the reading test was expanded by including items from different sources. A reading test was developed and administered for the first time to teachers in *SACMEQ II* and continued in subsequent studies. By the time *SACMEQ IV* project was implemented, there were several sources from which test items were drawn. Some of the test items that were administered in the sequential Zimbabwe Indicators of the Quality of Education and the *SACMEQ I* studies were maintained. There were items adopted from the IEA's International Study of Reading and Literacy Study (IRLS) that were also used in *SACMEQ II & III*. Some test items were sourced from PASEC to allow for comparison across assessment systems. Consistent with the developmental objectives of SEACMEQ, new test items were also written by the National Research Coordinators (NRCs) according to the curricula in the respective member countries. Care was taken to ensure that the

items were representative of the three domains of Narrative Prose, Expository Prose and Documents, whose descriptions are as follows.

Narrative prose: Continuous texts in which the writer aims to tell a story – whether this be fact or fiction.

Expository prose: Continuous text in which the writer aims to describe, explain, or otherwise convey factual information or opinion to the reader.

Documents: Structured information organized by the writer in a manner that requires the reader to search, locate, and process selected facts, rather than to read every word of a continuous text (Ross, et al, 2010).

The classification of reading and mathematics skills was undertaken by the consortium at the beginning of *SACMEQ II* project. To generate hierarchies of skills, the reading and mathematics test items were first arranged in order of difficulty, and then examined item-by-item in order to describe specific skills required to provide correct responses. When items had been linked to specific skills they were placed in groups of test items such that the item in each group had similar difficulty values and shared a common “theme” with respect to the underpinning competencies required to provide correct responses. The skill levels were combined with the domains to generate test blueprints. This was followed by conducting skills audits to ascertain congruency.

Focusing on the reading test, a skills audit was done during *SACMEQ II* Project using Rasch Analysis to find out if the proposed levels of skills in the test blue print were congruent with empirically generated levels. This resulted in the identification of eight levels of competence for the reading test. A descriptive name was linked with each of the levels – in order to summarize the competencies associated with each group of test items. The first three competency levels in reading employed the prefixes “Pre-“, “Emergent”, and “Basic” in order to reflect the mechanical nature of the most elementary competencies. Another portion of “Basic Reading” was levels four and five that emphasized comprehension skills. In general, Basic Reading skills are:

- recognizing and decoding words,
- matching individual words and phrases and extracting information directly from the text and
- reading forward and interpreting adjacent pieces of information.

From the sixth level upwards the prefixes of the summary names were different and were designed to reflect deeper abilities in subject-specific competencies. “Advance Reading” competencies that were included in the last three levels are outlined as:

- interpreting and making inferences from different types of text,
- analyzing detailed texts or extended documents for underlying messages and
- reading forward and backward making judgments about the assumptions, values and biases of the author (Ross, et al, 2010).

These levels of reading skills and their descriptors have been maintained through all successive studies; including *SACMEQ IV*. The results of the skills audit are presented in **Table 1**.

The eight competency levels provide a more concrete analysis of what pupils and teachers can actually do and they also suggest instructional strategies relevant to pupils who are learning at each level of competence.

Table 1: SEACMEQ competency levels in reading

Reading Competency Levels			
	Level	Descriptor	Competency
BASIC READING SKILLS	1	Pre-reading	Matches words and pictures involving concrete concepts and everyday objects, and follows short simple written instructions. Example Test Items: <ul style="list-style-type: none"> • Locate familiar words in a short (one line) text • Match words to pictures • Follow short and familiar instructions
	2	Emergent Reading	Matches words and pictures involving prepositions and abstract concepts; uses cuing systems (by sounding out, using simple sentence structure, and familiar words) to interpret phrases by Reading forwards. Example Test Items: <ul style="list-style-type: none"> • Read familiar words and identify some new words • Use simple and familiar propositions and verbs to interpret new words • Match words and very simple phrases
	3	Basic Reading	Interprets meaning (by matching words and phrases completing a sentence, matching adjacent words) in a short and simple text by reading forwards or backwards. Example Test Items: <ul style="list-style-type: none"> • Use context and simple sentence structure to match words and short phrases • Use phrases within sentences as units of meaning • Locate adjacent words and information in a sentence
	4	Reading for Meaning	Reads forwards and backwards in order to link and interpret information located in various parts of the text. Example Test Items: <ul style="list-style-type: none"> • Interpret sentence and paragraph level texts • Match phrases across sentences • Read forwards and backwards in order to locate information in longer texts
	5	Interpretive Reading	Reads forwards and backwards in order to combine and interpret information from various parts of the text in association with external information (based on recalled factual knowledge) that “completes” and contextualizes meaning. Example Test Items: <ul style="list-style-type: none"> • Locate, interpret, and read forward to join two pieces of adjacent information • Use multiple pieces of information to interpret general purpose of a document • Paraphrase and interpret a single non-adjacent piece of information

Reading Competency Levels			
	Level	Descriptor	Competency
ADVANCED READING SKILLS	6	Inferential Reading	<p>Reads forwards and backwards through longer (narrative, document or expository) texts in order to combine information from various parts of the text so as to infer the writer’s purpose.</p> <p>Example Test Items:</p> <ul style="list-style-type: none"> • Interpret, and make inferences from different types of texts by reading backward and forward to confirm links between widely separated information pieces • Extract information from a non-traditional (left to right) document • Make judgements about an author’s intentions or purpose beyond the text content
	7	Analytical Reading	<p>Locates information in longer (narrative, document or expository) texts by reading forwards and backwards in order to combine information from various parts of the text so as to infer the writer’s personal beliefs (value systems, prejudices, and/or biases).</p> <p>Example Test Items:</p> <ul style="list-style-type: none"> • Combine several pieces of information from a range of locations in complex and lexically dense text or documents • Analyze detailed text or extended documents for an underlying message • Identify meaning from different styles of writing
	8	Critical Reading	<p>Locates information in longer (narrative, document or expository) texts by reading forwards and backwards in order to combine information from various parts of the text so as to infer and evaluate what the writer has assumed about both the topic and the characteristics of the reader -such as age, knowledge, and personal beliefs (value systems, prejudices, and/or biases).</p> <p>Example Test Items:</p> <ul style="list-style-type: none"> • Use text structure and organization to identify an author’s assumptions and purposes • Identify an author’s motives, biases, beliefs in order to understand the main theme • Link text to establish multiple meanings including analogy and allegory

4.1.1 Construction of the Reading Tests for SACMEQ IV

As a norm, new test items were generated by the NRCs and other sources. Following various analyses on the new test items, summary statistics and individual item fit statistics were presented to a team of experts for discussion. At first, **30** out of **107** reading items that passed the analysis when the class interval structure was **10** were recommended for selection. However, it was observed that the class interval of **10** might have led to having fewer items passing the residual fit analysis. It was therefore recommended that the Class Interval Structure be reduced from the default **10** to **6** in order to counter the effects of having less number of cases per class interval. In addition, it was agreed that the intraclass correlation coefficients (ICCs) for all the items be extracted and analyzed to support the fit analysis. Further, that in the absence of linking items from the previous SACMEQ studies, a skills audit for pilot items across the various domains and levels be carried out to supplement the Rasch analysis.

4.1.1.1 Residual fit analysis

As recommended above, the class intervals were reduced from **10** to **6**. This indeed increased the number of cases in each of the class interval from around **133** to **225**. When residual analysis was conducted, **37** reading items passed the analysis.

4.1.1.2 Intraclass Correlation Coefficients (ICCs)

In the next step, analyses for ICCs were done for all the items including those that failed the residual fit analysis. The results indicated that **81** out of the **107** items passed this level of analyses. The breakdown for reading items that passed is as follows: **36** new; **24** Pre-Pearls; and **21** PASEC. It can be appreciated that **34** more reading items passed according to ICCs than those that passed the residual fit analysis. **Table 2** gives a summary of the items that passed the analysis at ICC level.

Table 2: Distribution of potential reading test items for SACMEQ IV study by source

	New	PASEC	Pre-Pearls
# of test items	36	21	24

Apparently a number of new reading items could be considered for the main survey. However, if the p-value for the Chi-square analysis is to be accounted for, and later on remove all the **24** Pre-Pearls items from the reading test, then the best items were: **17** new and **15** PASEC items which had a Chi-square p-value of 5% and above.

4.1.1.3 Items to be considered for the main survey

A decision was taken by the team of experts that only those new items whose psychometric values were similar to those they may replace in SACMEQ III blue-print should be selected for SACMEQ IV study. However, in the absence of link items, there were two technical options both of which were to help guide the selection of the items. The options were difficulty level of the items as shown by their location, and the skills audit.

As for the difficulty levels, the location of a piloted new reading item was compared with those of *SACMEQ III*. If the location of the new item was close or similar to a corresponding *SACMEQ III* item at that level, it became a candidate to replace the latter. However, there was a need to make sure that the new item was demanding similar or same skill(s) as that of the *SACMEQ III* item it was to replace.

A skills audit was thus conducted for all the new items that were ear-marked to replace *SACMEQ III* test items. It was established that the new items were demanding similar skills as those of corresponding *SACMEQ III* items. It was further established that at some locations there were no corresponding *SACMEQ III* items. This further reduced the number of piloted new items that would be used in the main survey. As for reading, **15** PASEC and **17** new items had similar locations and demands to those they would replace from *SACMEQ III*.

4.1.1.4 Further analyses

Considering the clustered nature of the sample, the sample size of the pupil items as the base of analyses was amended to **240**. In addition, the acceptable range for residual fit was extended to ± 3.0 in order to accommodate a few more items. For pupil reading test, a total of **77** new and PASEC items were put under analysis. Results of the residual fit analysis show that **37** (**24** new, **12** PASEC, and **1** *SACMEQ II*) items were good. However, considering the availability of items at particular locations in the link table, only **15** items (**10** New and **5** PASEC) could find locations at which to replace *SACMEQ III* items. It should be noted that many reading test items were crafted out of passages such that of the **18** passages that were in the test, only **2** (starting business and Paul's car) each had at least two items whose residual fit as well as location corresponded to particular items of *SACMEQ III*.

For the teacher reading test, **38** (**27** new, **2** *SACMEQ II*, and **9** PASEC) out of the **39** new test items exhibited good residual fit. However, only **12** of them (**11** New and **1** PASEC) had good fit as well as locations at which *SACMEQ III* items could be replaced. The results further show that out of the **12** passages, only **2** (Mugi, and Tom & Mary) each had **4** items that could fit at particular locations in *SACMEQ III* test.

4.1.1.5 Reading test plan for SACMEQ IV main study

Construction of tests for the main study is schemed to accommodate items from *SACMEQ III*, new/piloted, and PASEC piloted pools. **Table 3** gives an impression of the number of items available for both pupils and teachers reading tests.

Table 3: Pupil and teacher reading test plan for SACMEQ IV showing number of test items

Reading Test	<i>SACMEQ III</i>	New items with designated location	New items without designated location	PASEC items with designated location	PASEC items without designated location
Pupils	55	10	14	5	7
Teachers	52	11	16	1	8

Ideally, *SACMEQ IV* main study should've maintained all the characteristics of the test structure and format in order to maintain testing standards and construct validity. This implies maintaining the old test items and everything else. However, the new and PASEC test items needed to find their way into the test without compromising the test assumptions. The link table for *SACMEQ III* main study was thus used to somehow “calibrate” the new and PASEC items. In the blended test structure, the total number of test items was maintained. However, item composition from *SACMEQ III* was **50%**, new items at **30%**, and PASEC items at **20%**. With the fore going and the fact that a number of new and PASEC items were calibrated before going into the test, the test standards were ensured and maintained. **Table 4** presents a summarized picture of the test composition for *SACMEQ IV* main study.

Table 4: Reading test item composition for SACMEQ IV by source

Reading Test	<i>SACMEQ III</i>	New items	PASEC items	Total	Remarks
Pupils	28 (50%)	18 (33%)	10 (17%)	55 items	-
Teachers	26 (50%)	16 (31%)	10 (19%)	52 items	PASEC will be two items short

Although not exact, the test plan attempted to accommodate the test items according to the set proportions. It was not 100% mirror image of the intended Table of Test Specification. On the other hand, PASEC items were one short for teacher reading test. However, new piloted items would replace them as long as their residual fit statistics were good. It should also be noted that even those items whose locations could not match any *SACMEQ III* items had been considered. In **Table 5** and **Table 6**, the actual items which would constitute the pupil and teacher reading tests are provided. Where such items did not match any location in *SACMEQ III*, the same have been indicated to guide the item positioning during test assembly. The tables also show which new and PASEC test items were finally included in the reading tests.

Table 5: Pupil reading test item assembly

New Items		PASEC Items	
Item	Replacing in <i>SACMEQ III/Location</i>	Item	Replacing in <i>SACMEQ III/Location</i>
RD_16	46	RB_16	44
RA_04	14	RD_04	6
RB_16	44	RD_01	39
RD_08	13	RD_03	1
RD_17	53	RB_17	38
RC_02	33	RB_15	-0.29
RA_01	24	RB_18	-0.617
RA_10	30	RA_16	-1.148
RC_04	4	RA_17	0.352
RD_09	0.283	RD_02	-0.287
RD_10	0.337		
RB_01	-0.224		
RB_02	1.128		
RA_09	0.015		
RA_11	0.412		
RD_21	1.584		
RA_06	0.585		
RC_05	-2.449		
Total	18 items		10 items

Table 6: Teacher reading test item assembly

New Items		PASEC Items	
Item	Replacing in <i>SACMEQ III/Location</i>	Item	Replacing in <i>SACMEQ III/Location</i>
RB_06	42	RB_14	-1.282
RB_08	50	RB_15	-1.114
RB_09	33	RB_16	-0.254
RB_10	1.07	RB_17	0.114
RB_07	2.651	RB_18	-2.03
RB_11	23	RB_19	19
RC_04	31	RC_18	-3.099
RC_05	44	RC_19	-1.575
RC_09	39		
RC_10	03		
RC_11	08		
RC_12	28		

RC_13	35	
RC_14	-0.809	
RB_01	-0.061	
RB_02	-1.052	
RC_01	-2.089	
RC_02	-1.543	
Total	18 items	8 items

As shown in **Table 6**, PASEC contributed **8** teacher reading test items instead of the targeted 10. This was the case because of inadequate number of PASEC items with reasonable fit values. In order to balance the numbers, new items were added. It should also be noted that only one PASEC item (RB_19) had reasonable fit as well as a location at which to replace a *SACMEQ III* item. Consideration of the items whose location did not correspond to *SACMEQ III* items was on fit value, item level, domain, and, in the case for reading, stimulus. This was necessary in order to realize content and construct validity of the tests. Further, consideration for reading test item stimulus was made to maximize the number of items from a given stimulus.

4.1.1.6 The final test blue-print for SACMEQ IV reading tests

Having gone through the processes described above, the final test blue-prints for *SACMEQ IV* reading tests are given in **Table 7** and **Table 8**.

Table7: Pupil reading test blue-print for SACMEQ IV

Item	Stimulus	Domain*	Level	Source
1	Match: biggest	N	1	<i>SACMEQ III</i>
2	Match: point	N	1	<i>SACMEQ III</i>
3	Picture: house	D	1	PASEC
4	Picture: banana	D	1	PASEC
5	Picture: house	D	1	New
6	Picture: read bk	D	1	New
7	picture: kick ball	D	2	New
8	Picture: chair	D	3	New
8	Walrus	E	1	<i>SACMEQ III</i>
9	Walrus	E	5	<i>SACMEQ III</i>
10	Walrus	E	5	<i>SACMEQ III</i>
11	Bird	N	2	<i>SACMEQ III</i>
12	Bird	N	3	<i>SACMEQ III</i>
13	Bird	N	4	<i>SACMEQ III</i>
14	Bird	N	6	<i>SACMEQ III</i>
15	Aunt Mary	N	2	<i>SACMEQ III</i>
16	Aunt Mary	N	3	<i>SACMEQ III</i>

SACMEQ IV Project Technical Report

17	Aunt Mary	N	3	<i>SACMEQ III</i>
18	Aunt Mary	N	4	<i>SACMEQ III</i>
19	population	E	3	<i>SACMEQ III</i>
20	population	E	3	<i>SACMEQ III</i>
21	population	E	4	<i>SACMEQ III</i>
22	Maria	D	3	<i>SACMEQ III</i>
23	Map	D	3	New
24	Coffie's letter	N	3	PASEC
25	Coffie's letter	N	5	PASEC
26	Photography	D	3	<i>SACMEQ III</i>
27	Tembo	N	4	<i>SACMEQ III</i>
28	Weather	D	4	PASEC
29	Weather	D	5	PASEC
30	Oranges	D	4	<i>SACMEQ III</i>
31	Oranges	E	5	<i>SACMEQ III</i>
32	Letter below line	D	5	PASEC
33	Letter in square	D	6	PASEC
34	Bonobos	E	5	PASEC
35	Bonobos	E	5	PASEC
36	Gazelles	E	7	New
37	Gazelles	E	5	New
38	Gazelles	E	5	New
39	Mr Kimweri	N	5	New
40	Mr Kimweri	N	7	New
41	Concert	D	5	<i>SACMEQ III</i>
42	Health	D	5	<i>SACMEQ III</i>
43	Health	D	6	<i>SACMEQ III</i>
44	Smoke 6	E	5	<i>SACMEQ III</i>
45	Smoke 5	E	7	<i>SACMEQ III</i>
46	Smoke	E	8	<i>SACMEQ III</i>
47	Crowded street	N	5	New
48	Crowded street	E	6	New
49	Paul's car	N	5	New
50	Paul's car	N	5	New
51	Paul's car	N	7	New
52	Start business	E	5	New
53	Start business	E	5	New
54	Start business	E	5	New
55	Indian	E	6	<i>SACMEQ III</i>

* *D* = Document; *E* = Expository; *N* = Narrative

Table 8: Teacher reading test blue-print for SACMEQ IV

Item	Stimulus	Domain*	Level	Source
1	Picture: read bk	D	2	New
2	Picture: house	D	5	New
3	Usain Bolt 1	E	1	PASEC
4	Usain Bolt 2	E	2	PASEC
5	Weather 2	D	1	PASEC
6	Weather 1	D	5	PASEC
7	Weather 3	D	8	PASEC
8	Map 1	D	3	New
9	Map 2	D	7	New
10	Map 3	D	8	New
11	Bonobos 1	E	3	PASEC
12	Bonobos 2	E	3	PASEC
13	Bonobos 3	E	5	PASEC
14	Tom & Mary 5	N	3	New
15	Tom & Mary 2	N	3	New
16	Tom & Mary 4	N	6	New
17	Tom & Mary 3	N	7	New
18	Tom & Mary 6	N	7	New
19	Lost cat 2	D	4	New
20	Lost cat 1	D	5	New
21	Quicksand	E	5	SACMEQ III
22	Quicksand	E	5	SACMEQ III
23	Quicksand	E	5	SACMEQ III
24	Quicksand	E	6	SACMEQ III
25	Parachuting 2	N	5	SACMEQ III
26	parachuting 5	N	8	SACMEQ III
27	parachuting 6	N	8	SACMEQ III
28	Mr Kimweri 1	N	5	New
29	Smoke 4	E	5	SACMEQ III
30	Smoke 5	E	7	SACMEQ III
31	Tree	E	5	SACMEQ III
32	Tree	E	6	SACMEQ III
33	Mugi 1	N	6	New
34	Mugi 5	N	7	New
35	Mugi 4	N	7	New
36	Mugi 6	N	8	New
37	Mugi 2	N	8	New

38	Indian	E	6	SACMEQ III
39	Indian	E	7	SACMEQ III
40	Peking	E	6	SACMEQ III
41	Peking	E	8	SACMEQ III
42	Peking	E	8	SACMEQ III
43	Peking	E	8	SACMEQ III
44	Peking	E	8	SACMEQ III
45	Walrus	E	6	SACMEQ III
46	Walrus	E	7	SACMEQ III
47	Vacancy	D	7	SACMEQ III
48	Thinking	E	7	SACMEQ III
49	Thinking	E	8	SACMEQ III
50	Thinking	E	8	SACMEQ III
51	Thinking	E	8	SACMEQ III
52	Thinking	E	8	SACMEQ III

* *D = Document; E = Expository; N = Narrative*

4.2 The Mathematics Competency Levels for SEACMEQ Studies

Similar to the reading test, mathematics test was developed and administered for the first time to both teachers and pupils during *SACMEQ II* Project. This implies that the competency or skills levels for mathematics were also developed during *SACMEQ II*; prior to the construction of the test items. The test had 65 items which had two different sources. The first was the Trends in International Mathematics and Science Study (TIMSS) and the second was those constructed by the SACMEQ National Research Coordinators (NRCs) according to the curricula in the respective member countries (Ross, et al, 2006). Care was taken to ensure that the items were representative of the three domains of Number, Measurement and Space-data, described as follows;

- **Number:** Operations and number line, square roots, rounding and place value, significant figures, fractions, percentages, and ratios.
- **Measurement:** Measurements related to distance, length, area, capacity, money, and time.
- **Space-Data:** Geometric shapes, charts (bar, pie, and line), and tables of data.

The mathematics test items were first arranged in order of difficulty, and then examined item-by-item in order to describe specific skills required to provide correct responses. When items had been linked to specific skills they were placed in groups of test items such that the item in each group had similar difficulty values and shared a common “theme” with respect to the underpinning competencies required to provide correct responses. The skills audit for the mathematics tests resulted in the identification of eight levels of competence.

The results of the skills audit are presented in **Table 9 (a) & (b)**. A descriptive name was linked with each of the levels – in order to summarize the competencies associated with each group of test items. The first three competency levels mathematics employed the prefixes “Pre-“, © SEACMEQ 2017

“Emergent”, “Basic” and “Beginning” in order to reflect the mechanical nature of the most elementary competencies. The first four competencies were designated as “Basic Mathematics” and involved mainly manipulation of basic operations and recognition of shapes and figures. They included:

- applying single- to two-step addition or subtraction of whole numbers and basic fractions,
- recognising three-dimensional shapes and number units, and
- translating simple textual/verbal information into basic arithmetic forms using whole numbers and fractions.

From the fourth level upwards the prefixes of the summary names were different and were designed to reflect deeper levels of understanding of subject-specific competencies. Levels five, six, seven and eight were designated as “Advanced Mathematics” and encompassed problem solving skills which included:

- solving multiple-operation problems using the correct order of arithmetic operations on whole and mixed numbers, fractions, ratios, and decimals,
- extracting and converting information from tables, charts, visual and symbolic presentations in order to identify, and then solve multi-step problems and
- identifying the nature of an unstated mathematical problem embedded within verbal or graphic information, and then translating this into symbolic, algebraic, or equation form to solve the problem.

The eight competency levels provide a more concrete analysis of what pupils and teachers can actually do and they also suggest instructional strategies relevant to pupils who are learning at each level of competence.

Table 9 (a): SEACMEQ competency levels in mathematics (levels 1-4)

Mathematics Competency Levels			
	Level	Descriptor	Competency
BASIC MATH SKILLS	1	Pre- Numeracy	<p>Applies single step addition or subtraction operations. Recognizes simple shapes. Matches numbers and pictures. Counts in whole numbers.</p> <p>Example Test Items:</p> <ul style="list-style-type: none"> • Count illustrated objects • Recognize basic numbers and shapes • Carry out simple single operations of addition and subtraction
	2	Emergent Numeracy	<p>Applies a two-step addition or subtraction operation involving carrying, checking (through very basic estimation), or conversion of pictures to numbers. Estimates the length of familiar objects. Recognizes common two-dimensional shapes.</p> <p>Example Test Items:</p> <ul style="list-style-type: none"> • Link simple verbal, graphical, and number forms with single arithmetic operations on whole numbers up to four digits • Recognize common shapes or figures in two dimensions • Estimate accurately lengths of simple shapes
	3	Basic Numeracy	<p>Translates verbal information (presented in a sentence, simple graph or table using one arithmetic operation) in several repeated steps. Translates graphical information into fractions. Interprets place value of whole numbers up to thousands. Interprets simple common everyday units of measurement.</p> <p>Example Test Items:</p> <ul style="list-style-type: none"> • Recognize three-dimensional shapes and number units • Use a single arithmetic operation in two or more steps • Convert in single step units using division
	4	Beginning Numeracy	<p>Translates verbal or graphic information into simple arithmetic problems. Uses multiple different arithmetic operations (in the correct order) on whole numbers, fractions, and/or decimals.</p> <p>Example Test Items:</p> <ul style="list-style-type: none"> • Convert units in two steps and count tabulated data • Analyze a visual prompt and interpret triangular shapes • Translate verbal to arithmetic form using two operations on fraction

Table 9 (b): SEACMEQ competency levels in mathematics (levels 5-8)

Mathematics Competency Levels			
	Level	Descriptor	Competency
ADVANCED MATH SKILLS	5	Competent Numeracy	<p>Translates verbal, graphic, or tabular information into an arithmetic form in order to solve a given problem. Solves multiple-operation problems (using the correct order of arithmetic operations) involving everyday units of measurements and/or whole and mixed numbers.</p> <p>Converts basic measurement units from one level of measurement to another (for example metres to centimetres).</p> <p>Example Test Items:</p> <ul style="list-style-type: none"> • Concert basic measurement units • Understand the order of magnitude of simple fractions • Conduct multiple steps with a range of basic operations in a strict sequence using an analysis of a short verbal or visual prompt
	6	Mathematically Skilled	<p>Solves multiple-operation problems (using the correct order of arithmetic operations) involving fractions, ratios, and decimals. Translates verbal and graphic representation information into symbolic, algebraic, and equation form in order to solve a given mathematical problem. Checks and estimates answers using external knowledge (not provided within the problem).</p> <p>Example Test Items:</p> <ul style="list-style-type: none"> • Perform complex and detailed mathematical tasks (involving considerable abstraction of verbal, visual, and tabular information into symbolic forms and algebraic solutions) using knowledge not supplied with the task • Use of an extended verbal or graphical prompt (involving an analysis of steps) to identify the correct sequence of calculations • Convert, and operate on, units of measurement (time, distance, and weight)
	7	Concrete Problem Solving	<p>Extracts and converts (for example, with respect to measurement units) information from tables, charts, visual and symbolic presentations in order to identify, and then solve multi-step problems.</p> <p>Example Test Items:</p> <ul style="list-style-type: none"> • Use multiple verbal order of steps with conversion of time units • Translate verbal to arithmetic form, apply units conversion with long division • Convert from mixed number fraction to decimals

	8	Abstract Problem Solving	<p>Identifies the nature of an un-stated mathematical problem embedded within verbal or graphic information, and then translates this into algebraic or equation form in order to solve the problem.</p> <p>Example Test Items:</p> <ul style="list-style-type: none"> Identify the nature of a problem, translate the information given into a mathematical approach, and then identify the correct mathematical strategies to obtain a solution
--	---	--------------------------	---

4.2.1 Construction of the Mathematics Tests for SACMEQ IV

The development of mathematics test items for SACMEQ IV study was undertaken concurrently with that of the reading tests. The procedure is therefore identical to that of reading tests which have already been detailed above. Consequently, the discussions under this section will be brief and rather dominated by empirical presentations. A total of **43** out of **93** new test items for mathematics were submitted to the team of experts for consideration. These were items that passed the analysis when the class interval structure was 10. However, a Bonferroni adjustment had not been effected.

4.2.1.1 Residual fit analysis

Following the reduction of the class intervals from **10** to **6** which resulted in the number of cases in each of the class interval increasing from around **133** to **225**, fifty one (**51**) mathematics test items passed the analysis.

4.2.1.2 Intraclass correlation coefficients (ICCs)

The results of the analyses for ICCs shown in **Table 10** shows that **66** out of the **93** mathematics items were suitable for the study. The breakdown of these mathematics items is as follows: **37** new; **16** TIMMS, and **13** PASEC. Fifteen (**15**) more items for mathematics passed according to ICCs than with the residual fit analysis.

Table 10: Distribution of potential mathematics test items for SACMEQ IV study by source

	New	PASEC	TIMMS
# of test items	37	13	16

If the p-value for Chi-square analysis is to be accounted considered, and later on remove all the 16 TIMMS items from Mathematics, then the best mathematics items are: **22** new and **10** PASEC. Meaning that out of the remaining **50** mathematics items, a total of **32** each had a Chi-square p-value of **5%** and above.

4.2.1.3 Items to be considered into the main survey

Only **5** PASEC and **14** new items for mathematics had similar locations and demanding similar skills at the level they would be placed in *SACMEQ III* test which had a total of **49** items.

4.2.1.4 Further analyses

The results show that for mathematics pupil test, out of the **71** items **43** passed the residual fit analysis. In the **43** group of passing items, **34** are new while **9** are PASEC items. However, **19** items (**15** New and **4** PASEC) can go into the final test because of the value of the residual fit and availability of replacement items at the locations in the link table.

Thirty seven mathematics test items for teacher test were analyzed. Results show that **36** (**25** new and **10** PASEC) of the items passed the residual fit analysis. However, only **11** of the items (**6** New and **5** PASEC) exhibited good residual fit and were at locations at which there were *SACMEQ III* items to replace.

4.2.1.5 Mathematics test plan for SACMEQ IV main study

Construction of mathematics tests for the main study is also schemed to accommodate items from *SACMEQ III*, new/piloted, and PASEC piloted pools. **Table 11** gives an impression of the number of items available for both pupils and teachers mathematics tests.

Table 11: Pupil and teacher mathematics test plan for SACMEQ IV showing number of test items

Mathematics Test	<i>SACMEQ III</i>	New items with designated location	New items without designated location	PASEC items with designated location	PASEC items without designated location
Pupils	49	15	18	4	5
Teachers	42	6	19	5	5

However, the final item composition from *SACMEQ III* will be **50%**, new items at **30%** and PASEC items at **20%**. **Table 12** shows a summarized picture of the mathematics test composition for *SACMEQ IV* main study.

Table 12: Mathematics test item composition for SACMEQ IV by source

Reading Test	SACMEQ III	New items	PASEC items	Total	Remarks
Pupils	25 (50%)	14 (29%)	10 (20%)	49 items	PASEC will be one item short
Teachers	21 (50%)	12 (29%)	9 (21%)	42 items	

In **Table 13** and **Table 14**, the actual items which would form the pupil and teacher mathematics tests have been provided.

Table 13: Pupil mathematics test item assembly

New Items		PASEC Items	
Item	Replacing in SACMEQ III/Location	Item	Replacing in SACMEQ III/Location
MD_05	4	MB_19	38
MC_21	40	MD_16	49
MD_03	5	MD_15	39
MA_23	47	MC_09	25
MC_07	13	MB_08	-1.723
MA_25	30	MD_14	0.137
MC_18	27	MA_09	0.735
MD_12	22	MB_20	0.689
MB_02	3	MC_15	0.349
MD_13	26		
MC_26	41		
MD_18	15		
MD_04	28		
MA_08	2		
MB_13	18		
Total	15 items		9 items

Table 14: Teacher mathematics test item assembly

New Items		PASEC Items	
Item	Replacing in <i>SACMEQ III</i> /Location	Item	Replacing in <i>SACMEQ III</i> /Location
MB_02	34	MB_15	14
MC_04	2	MB_20	1
MB_03	3	MB_19	16
MB_07	23	MB_22	22
MC_20	17	MC_14	24
MB_05	38	MC_08	-0.607
MC_03	-5.029	MC_15	-0.505
MB_01	-1.702	MC_23	-0.405
MB_04	-0.606	MC_09	0.075
MC_21	-0.338		
MC_06	1.020		
MB_26	2.485		
Total	12 items		9 items

4.2.1.6 The final test blue-print for *SACMEQ IV* mathematics tests

The final test blue-prints for *SACMEQ IV* mathematics tests are given in **Table 15** and **Table 16**.

Table 15: Pupil mathematics test blue-print for *SACMEQ IV*

Item	Stimulus	Domain*	Level	Source
1	Which symbol	N	1	<i>SACMEQ III</i>
2	895-783	N	1	NEW
3	360 + 35	N	1	NEW
4	Missing no. in series	N	1	<i>SACMEQ III</i>
5	Largest Number	N	2	NEW
6	_ x 5 = 15	N	2	PASEC
7	Sides in figure	SD	2	NEW
8	11 chairs	SD	2	<i>SACMEQ III</i>
9	Cylinder flat	SD	2	<i>SACMEQ III</i>
10	Length of line	M	2	<i>SACMEQ III</i>
11	Shaded fraction: triangle	SD	3	NEW
12	Dowish bananas	N	3	NEW

SACMEQ IV Project Technical Report

13	Meaning of $20 < 35$	N	3	NEW
14	Three times 9.5	N	3	<i>SACMEQ III</i>
15	Tabu's home	N	3	<i>SACMEQ III</i>
16	Pattern of tiles	N	3	<i>SACMEQ III</i>
17	Open day for Kingogo	N	3	<i>SACMEQ III</i>
18	Bicycles sold: Graph	SD	3	NEW
19	Same fraction	N	4	<i>SACMEQ III</i>
20	3142ml in litres	M	4	<i>SACMEQ III</i>
21	Highest temperature	SD	4	<i>SACMEQ III</i>
22	Time on clock	M	4	<i>SACMEQ III</i>
23	Numbers of complete pattern	N	4	NEW
24	Subtract from 900	N	4	<i>SACMEQ III</i>
25	Maputo-Ressano train	M	5	NEW
26	Mia triangle	SD	5	<i>SACMEQ III</i>
27	Joe 78 pages	N	5	<i>SACMEQ III</i>
28	Lusaka to Chisamba	M	5	<i>SACMEQ III</i>
29	$11/8 + 11/4$	N	5	NEW
30	Bus driver	M	5	<i>SACMEQ III</i>
31	Vegetable plots	N	5	PASEC
32	Ac airplane	M	5	PASEC
33	Scale in sketch	SD	5	<i>SACMEQ III</i>
34	Price list total cost	SD	6	<i>SACMEQ III</i>
35	John ride	N	6	PASEC
36	House plan	M	6	PASEC
37	reduce $28/42$	N	6	NEW
38	Toy cars	SD	6	NEW
39	Travel: 10km	M	6	PASEC
40	Mary Geometry score	N	6	PASEC
41	Statements	N	6	PASEC
42	Acute angle	SD	6	NEW
43	Area of shape	SD	6	NEW
44	Days, hours, mins	M	7	<i>SACMEQ III</i>
45	80g small packets	M	7	<i>SACMEQ III</i>
46	$5 \frac{3}{5}$ same value	N	7	<i>SACMEQ III</i>
47	Height of a fence	N	7	<i>SACMEQ III</i>

48	Cakes slices eaten	SD	8	PASEC
49	Bus driver	M	8	NEW

* *M* = Measurement; *N* = Number; *SD* = Space/data

Table16: Teacher mathematics test blue-print for SACMEQ IV

Item	Stimulus	Domain*	Level	Source
1	People in two villages	N	1	New
2	Shapes_Rectangle	N	2	New
3	895 - 783	N	2	New
4	Teacher groups	N	2	PASEC
5	Fraction shaded circle	SD	3	New
6	Time on clock	M	4	New
7	Two lines	SD	4	PASEC
8	Angles: 180	SD	4	New
9	John ride	N	4	PASEC
10	Ann & Tom pencils	N	4	PASEC
11	Numbers of complete path	N	4	New
12	Travel 10km	M	4	PASEC
13	House plan	M	5	PASEC
14	Faces & corners	SD	6	New
15	Area of shaded part	SD	6	SACMEQ III
16	Vegetable plots	N	6	PASEC
17	Fuel tank 35 litres	N	6	SACMEQ III
18	Girls to boys	N	6	SACMEQ III
19	Ratio side perimeter	SD	6	SACMEQ III
20	Joe, Mary, Lilian	N	6	SACMEQ III
21	Length of E	SD	6	New
22	AB parallel to CD	SD	6	SACMEQ III
23	9060mm in metres	M	6	New
24	False equation	N	7	SACMEQ III
25	Two athletes	M	7	SACMEQ III
26	triangular faces	SD	7	New
27	$x/2 < 7$	N	7	SACMEQ III
28	24.56/0.004	N	7	SACMEQ III
29	Rectangle with 6 squares	SD	7	PASEC
30	Shadow of a tree	N	7	SACMEQ III
31	200 sheets thick	N	7	SACMEQ III
32	x proportional to y	N	7	SACMEQ III

33	x & y missing	N	7	<i>SACMEQ III</i>
34	Enni paint	N	7	<i>SACMEQ III</i>
35	Height of a fence	N	7	<i>SACMEQ III</i>
36	Straight line	SD	7	<i>SACMEQ III</i>
37	Oranges left	N	8	<i>SACMEQ III</i>
38	5 square perimeter	M	8	<i>SACMEQ III</i>
39	Water tank	N	8	<i>SACMEQ III</i>
40	North mountain	N	8	<i>SACMEQ III</i>
41	Henry & Yao jump	M	8	PASEC
42	Chart-School drop	SD	8	New

* *M* = Measurement; *N* = Number; *SD* = Space/data

4.3 Construction of the SEACMEQ HIV and AIDS Knowledge Test

In 2005, the then SACMEQ’s Governing Body (the SACMEQ Assembly of Ministers of Education) expressed concern about the impact of HIV and AIDS on the conditions of schooling and the quality of education. At the time, most SEACMEQ ministries of education were implementing curricular aimed at increasing pupils’ knowledge about HIV and AIDS. Knowledge about HIV and AIDS had been identified as a key determinant of the decision-making ability and behavior of young people in SEACMEQ countries. In this respect, SEACMEQ Ministers asked SEACMEQ National Research Coordinators (NRCs) to develop a valid SEACMEQ HIV and AIDS Knowledge Test that would be suitable for administration to Grade 6 pupils (Primary 6 pupils) and their teachers.

The SEACMEQ HIV and AIDS Knowledge Test (HAKT) is therefore a test that was designed to assess the knowledge that Grade 6 pupils (and their teachers) in Southern and Eastern Africa have about those aspects of HIV and Aids that are covered in education programmes delivered by schools.

4.3.1 Defining knowledge sub-topics for the HAKT

In order to make an assessment of the knowledge that Grade 6 pupils have about HIV and AIDS for the *SACMEQ III* Project, it was decided that a special-purpose knowledge test should be administered at the same time as the other main data collection. Though this test was prepared with Grade 6 Pupils in mind, it was found necessary to consider administering the same to teachers who play an important role in the HIV and AIDS Curriculum delivery in schools. The HAKT was later to be administered in its entirety to pupils and teachers in *SACMEQ IV*.

The first key decision in the construction of the HAKT was that it should be based on the knowledge that Grade 6 pupils were expected to acquire about HIV and AIDS from official education programmes presented in schools. Therefore, it was necessary to construct a test blueprint (or test framework) that defined the main dimensions of knowledge about HIV and

AIDS which were explicitly identified by Ministries of Education in official school curricula, teaching and learning materials, and teachers' guides.

The framework that was used to guide the initial stages of construction was derived from a report of the International Bureau of Education (IBE, 2006) that identified four thematic areas that were touched upon to varying degrees in many official descriptions of the contents of HIV and AIDS education programmes. After consultations with international experts, the SACMEQ NRCs decided to concentrate the HAKT on an assessment of IBE's Theme 1: "Basic Knowledge for Protecting and Promoting Health". Once this decision had been made they turned their attention towards defining the main Knowledge Topics within Theme 1 that would be assessed by HAKT. The SEACMEQ curriculum analysis identified five Knowledge Topics that were being covered by HIV and AIDS education programmes in SEACMEQ school systems. They are listed as follow:

- i. **Definitions/Distinctions:** Pupils should know basic information (simple facts, terminology, and acronyms) about HIV, AIDS, and ARV; and they should also understand the causal linkage between HIV and AIDS;
- ii. **Transmission Mechanisms:** Pupils should know the three principal pathways for HIV transmission (unprotected sexual intercourse, blood, and mother to baby);
- iii. **Avoidance Behaviours:** Pupils should understand that it is possible to avoid HIV infection, and that certain behaviours can offer protection against HIV infection (such as "The ABC Rule": Abstinence, Be Faithful, and Condoms);
- iv. **Diagnosis and Treatment:** Pupils should understand the basic elements of HIV infection – especially with respect to diagnosis, treatment, and the possibility of a cure;
- v. **Myths and Misconceptions:** Pupils should be able to identify and evaluate myths (widely accepted, but false beliefs) and misconceptions that circulate in some communities about HIV and AIDS.

There was need to conduct further review of existing literature to guide the process of test item development.

4.3.2 Review of literature on HIV and AIDS knowledge tests

Following the designation of the main HAKT Knowledge Topics and Sub-Topics, the NRCs reviewed evaluation materials that could provide information for constructing test items for the HAKT. The available documents fell into three broad groups:

- i. **Textbooks Tasks:** that were included in textbooks and teaching materials;
- ii. **Teacher Guide Questions:** that were given to pupils by teachers during, or following, classroom instruction;
- iii. **Survey Questions:** that were found in data collection instruments used by external agencies as part of evaluation studies and health monitoring programmes.

The NRCs reviewed Textbook Tasks in booklets and pamphlets on HIV and AIDS published by Education Ministries, Health Ministries, and a wide range of Non-Governmental Organizations. They then considered examples of Teacher Guide Questions that were mostly contained in

teacher manuals and individually-produced teacher lesson plans. Another document that was reviewed was the UNAIDS’ guidelines on the construction of core indicators (UNAIDS, 2007) for guidance on the on the survey questions that could be used to collect data, and the manner in which these data should be employed in order to calculate national scores on the indicator. This process resulted in the preparation of a total of **43** Knowledge Sub-Topics which have been listed in **Table 17**. For each of the **43** Knowledge Sub-Topics there were **2** test items in the final version of the HAKT – which resulted in a total of **86** items.

4.3.3 Choice of the type of test items for the HAKT

The first step in developing test items for the HAKT was to decide upon the exact format of the test items. It was agreed that a closed response design would be more manageable. The NRCs started by developing test items in a multiple choice format with four response options per item. This proved to be very difficult because three plausible distractor responses for each question proved hard to draft. When items with a correct response and three distracters were prepared it was noted that they presented a very large reading comprehension load for respondents. Following a great deal of discussion, experimentation, and consultation it was decided that the HAKT items should be prepared in “true/false” format, and that each item should be carefully read to the respondents twice before they answered “true” or “false”.

The true/false item format in which respondents were asked to read the questionnaires twice before answering offered three important advantages. First, the length and linguistic complexity of each test item could be reduced. Second, more items could be used within a given time period in order to obtain a full coverage of the **43** SEACMEQ Knowledge Sub-Topics. Third, the approach of reading each item to the respondents twice before they responded addressed the issue that some respondents had limited reading skills.

4.3.4 Outline of the HAKT

As already indicated, **86** true/false HIV and AIDS test items were developed for the SEACMEQ projects. The distribution of the items according to the curricula review is shown in **Table 17**.

Table 17: Knowledge Topics and Sub-Topics and item allocation for the HAKT

Knowledge Topics	Knowledge Sub-Topics	Number of Items
1. Definitions / Distinctions (10 items)	1. Terminology: HIV	2
	2. Terminology: AIDS	2
	3. Terminology: ARV	2
	4. HIV and AIDS Links: Causal Direction	2
	5. HIV and AIDS Links: Sickness	2
2. Transmission Mechanisms	6. Body Fluids: Blood	2
	7. Body Fluids: Saliva	2

(28 items)	8. Body Fluids: Semen	2
	9. Body Fluids: Sweat	2
	10. Body Fluids: Vaginal	2
	11. Hygiene: Food	2
	12. Hygiene: Water	2
	13. Hygiene: Smoke/Dirt	2
	14. Hygiene: Cough/Spit	2
	15. Mother to Child: Pregnancy and Birth	2
	16. Mother to Child: Breast Milk	2
	17. External Sources: Needles	2
	18. External Sources: Animals/Insects	2
19. External Sources: Bites	2	
3.Avoidance Behaviours (24 items)	20. Sexual Conduct: Abstain (“A”)	2
	21. Sexual Conduct: Be Faithful (“B”)	2
	22. Sexual Conduct: Condoms (“C”)	2
	23. Sexual Conduct: Delay	2
	24. Sexual Conduct: Appearance Healthy	2
	25. Sexual Conduct: Uninfected Person	2
	26. Sexual Conduct: Additional Risks	2
	27. Social Contact: Proximity	2
	28. Social Contact: Touch	2
	29. Social Contact: Sharing Food	2
	30. Social Contact: Sharing Objects	2
31. Social Contact: Additional Risks	2	
4.Diagnosis and Treatment (16 items)	32. Diagnosis : Medical Test	2
	33. Diagnosis : Virus Detection	2
	34. Diagnosis : Opportunistic Infections	2
	35. Diagnosis : Appearance (HIV)	2
	36. Diagnosis : Appearance (AIDS)	2
	37. Treatment: ARV Medication	2
	38. Treatment: Doctor’s Medicine	2
	39. Treatment: Cure	2
5.Myths and Misconceptions (8 items)	40. Myths: Healing	2
	41. Myths: Witches	2
	42. Misconceptions: Avoidance Possibility	2
	43. Misconceptions: Certainties	2
Total =	86	

The **86** HAKT test items covered **43** curriculum topics, and they were focused on assessment of “the basic knowledge about HIV and AIDS that is required for protecting and promoting health”. These topics were grouped into five main areas: definitions/terminology; transmission mechanisms; avoidance behaviours; diagnosis/treatment; and myths/misconceptions. Transmission Mechanisms (**28**) and Avoidance Behaviours (**24**) were the topics to which the highest numbers of items were allocated. Myths and Misconceptions (**8**) had the lowest allocation of test items.

4.3.5 Analytic framework for scoring and evaluating the HAKT

The SEACMEQ NRCs accepted that knowledge about HIV-AIDS was a necessary, but not sufficient requirement to ensure that young people would adopt behaviours that would protect and promote their own health and the health of others. However, they also recognized that ignorance about HIV and AIDS could never provide a sound foundation for wise behaviour. Following extensive discussion of these matters, they agreed that it would be appropriate for each SEACMEQ country to aim to deliver HIV and AIDS prevention education programmes at the primary school level that would permit (a) all Grade 6 pupils to reach the minimal knowledge level on the HAKT, and (b) a least half of the Grade 6 pupils to reach the desirable knowledge level on the HAKT

Therefore the performance of pupils and teachers on the HAKT was to be evaluated by applying three scoring procedures:

- i. **Transformed Scores**- scaled HAKT scores that were transformed to an overall SACMEQ mean of **500** and standard deviation of **100**.
- ii. **Minimal Knowledge Scores** – dichotomous scores that indicated whether or not respondents had mastered at least **50%** of the officially-designated curriculum that was assessed by the HAKT.
- iii. **Desirable Knowledge Scores** – dichotomous scores that indicated whether or not respondents had mastered at least **75%** of the officially-designated curriculum that was assessed by the HAKT.

The HAKT scaled scores that represented **minimal knowledge** needed to identify pupils that had demonstrated that they understood at least **50** percent (or more) of the officially-designated curriculum that was assessed by the HAKT. In a hypothetical situation in which **100** “true/false” test items are employed in a pupil test, it would be expected that a pupil located at the minimal knowledge level would be able to: (a) use knowledge to provide **50** correct answers for **50** items, and (b) use guessing to provide **25** correct answers for the other **50** items. That is, we would expect that a person at the minimal knowledge level would be able to provide **75** correct responses to the **100** test items. In the case of the HAKT the scaled score that corresponded to **75** percent correct responses was used as the cut-off point to identify those students that had reached the minimal knowledge level. Using the same line of reasoning, the HAKT scaled score that represented the cut-off point for **desirable knowledge** on this hypothetical **100**-item test corresponded to **87.5** percent correct responses.

4.3.6 Development of administration procedures for the HAKT

The NRCs were adamant that the research results obtained from an administration of the HAKT would yield valid information across all SEACMEQ countries. It was therefore important that the test administration procedures should be standardised so as to guarantee that all testing sessions in all schools in all SEACMEQ countries were exactly identical, and that they conformed to the specifications provided by the SEACMEQ researchers. The main HAKT administration requirements were therefore developed and are listed below:

- i. **Practice Items:** Before commencing the HAKT, the respondents are to be given an opportunity to complete and discuss six practice items so that they fully understood the presentation and the meaning of test items in “true/false” format.
- ii. **Presentation of Test Items:** It was important that responses to HAKT items are not influenced by the reading comprehension skills of the respondents. Therefore each item of the HAKT is to be completed in a three-step fashion as follows: (a) an item is to be read slowly and carefully by Data Collectors to the respondents (while they followed the written wording of the item in the printed version of the HAKT), (b) the item is re-read, and (c) the respondents are invited to answer whether the item was true or false. All respondents in the same testing session are to proceed together (one item at a time) through the HAKT until the last item is completed.
- iii. **Supervision:** Two Data Collectors are to be present in the room throughout the testing session. One of these persons is to be responsible for training the respondents with the practice questions and for managing the administration of the HAKT. The other person is to move around the room to ensure that all respondents answered all test items during the testing session.
- iv. **Time Allocation:** The time allocation for the HAKT is one hour - with the provision of up to an extra 15 minutes that can be allowed so as to ensure that: (a) the HAKT is not a “speed test”, and (b) sufficient time is available in schools where respondents could have some extra discussion about the practice items.
- v. **After the Testing Session:** When all respondents have completed the HAKT and handed their completed booklets to the Data Collectors, around 30 minutes are to be spent on reviewing and discussing the HAKT items. The Data Collectors are to read each test item and provide the correct answer. Pupils are to be invited to ask questions and to discuss test items that need further clarification.

4.3.7 Item analysis of the HAKT in the SEACMEQ main data collection

The final 86-item version of the HAKT was integrated into the Pupil and Teacher data collection booklets for SEACMEQ projects, and the administration procedures were explained in detail in the Data Collector Manual (SACMEQ, 2007). The main SEACMEQ project data collection usually takes place between September and October in all SEACMEQ school systems. When data from all countries have been cleaned, merged and weighted, the final scaling and scoring phases are completed; preliminary data analyses are conducted to check the performance of test items.

In order to construct a scale of knowledge about HIV and AIDS that could be used to make cross-national comparisons among SEACMEQ countries, it is important to ensure that the technical characteristics of test items would permit them to be combined into a reliable and valid scale. The characteristics of the HAKT items were analyzed by using the RUMM software system (Andrich, Sheridan & Luo, 2003) which is based on the Rasch Model of measurement. The Rasch item analyses provided information about the reliability of the HAKT scaled score derived from the final set of items, and the technical characteristics of each item. These analyses allowed the NRCs to fine-tune the HAKT by (i) the removal of poorly-fitting test items, and (ii) the removal of biased test items.

4.3.8 Removing poorly-fitting test items from the HAKT

One important goal of the Rasch analysis of test items is to confirm that a final set of test items that are used to construct overall scaled scores for individual respondents can be combined into a scale that represents a single dimension (sometimes referred to as a “latent variable”). This analysis proceeds by checking the congruence (or “fit”) between: (a) the response patterns for test items that are provided by ability-defined groups of respondents, and (b) the response patterns that can be predicted from the Rasch theoretical model by using information about the ability levels of the groups and the difficulty levels of test items.

The analysis is undertaken as an iterative process. The first step requires the identification of items that do not fit the Rasch theoretical model, followed by the removal of the “worst fitting” item and the recalculation of scaled scores. The next step checks the remaining items for lack of fit and, if necessary, action may be taken to remove another item. The analysis continues until the final set of test items: (a) can be used to calculate an overall scaled score that reliably represents a single dimension, and (b) provides a valid coverage of the skills that are required to demonstrate meaningful levels of performance on the dimension assessed by the scaled scores.

A total of five test items were removed from the original set of **86** HAKT items because of a lack of fit with the Rasch Model. Two of these (Item Labels **64** and **86**) were associated with knowledge topics covering body fluids (saliva and sweat). Another two items (Item Labels **66** and **81**) were concerned with tests for HIV infection. The fifth item (Item Label **14**) was linked with misconceptions about mother to child HIV transmission.

The removal of these five test items did not cause major imbalances in the coverage of the Knowledge Topics and Sub-Topics assessed by the HAKT. In addition, the shorter version of the HAKT resulted in an improvement in the reliability of scaled scores from **0.81** (using **86** test items) to a very respectable value of **0.84** (using **81** test items).

It is important to note that the fit analysis simply demonstrated that a valid scaled score could be obtained by omitting a small number of items and thereby increasing reliability. It did not prove that the removed items were of poor quality – it only suggested that the removed items probably belonged in some other set of items that defined a different dimension. For this reason, other analyses of HAKT data conducted at the individual item level to identify knowledge gaps in HIV and AIDS curricula included all **86** test items.

4.3.9 Removing biased test items from the HAKT

An important requirement of the Rasch Model is that test items should not be “biased” – in the sense that different groups of socially-defined respondents (for example, gender groups or national groups) with similar knowledge levels should demonstrate similar performance on each test item.

If different groups of respondents with similar knowledge levels appear to experience different levels of difficulty for a particular test item, then this item is said to be “biased” because it displays the characteristic of “Differential Item Functioning” (or “DIF”). In this research the different groups examined in the DIF analyses were defined by Gender (2 groups) and school system (15 groups).

4.3.9.1 The first DIF analysis (Gender)

This was conducted by examining whether any items favoured groups defined by Gender (boys or girls). Only one item (Label **17**) related to sharing objects (toilet seats) was detected with significant Gender DIF and was therefore excluded from the calculation of the HAKT scaled scores for all respondents.

One can only speculate about the reason why significant levels of DIF occurred in this analysis for this particular item. Perhaps the worldwide tendency of mothers to warn daughters about the “dangers” of sitting on toilet seats generates differences in the ways in which girls and boys respond to this kind of item?

4.3.9.2 The second DIF analysis (Country)

This was conducted by examining whether any items favoured groups defined by Country (the then 15 SEACMEQ countries). The item removals for this analysis were country-specific in that where significant DIF occurred for one country (or several countries) then that item was removed for that country (or those countries).

It was expected that this analysis would lead to significant levels of DIF for several items because of the difficulty of maintaining stable item difficulty levels for a knowledge test that has been administered in three different languages (English, Portuguese, and Kiswahili). It was hypothesized that some HIV and AIDS concepts might be somewhat easier to understand if they are taught in certain languages. Further research is certainly warranted in this area.

Four items (Labels **12**, **29**, **36** and **68**) were detected with significant DIF across Countries and were also excluded from the calculation of the HAKT scaled scores for all respondents.

In addition, some items that presented significant DIF for specific countries were then excluded for the respondents of these countries. The excluded items have been listed in **Table 18** under separate language groups.

4.3.10 Final selection of the test items

The final number of items per country that were included in the calculation of scaled HAKT scores ranged from a low of **69** items (Zanzibar) to a high of **76** items (Lesotho, Mozambique, Namibia, Swaziland and Zambia).

Table 18: Excluded Items by Language Group and School System

<u>English</u>	
Botswana:	Item Label: 51
Kenya:	Item Label: 10, 54
Lesotho:	No Exclusions
Malawi:	Item Label: 47
Mauritius:	Item Label: 2, 13, 32, 39, 44, 76
Namibia:	No Exclusions
Seychelles:	Item Label: 39
South Africa:	Item Label: 39
Swaziland:	No Exclusions
Uganda:	Item Label: 65
Zambia:	No Exclusions
Zimbabwe:	Item Label: 9

Portuguese

Mozambique: No Exclusions

Kiswahili

Tanzania: Item Label: 3, 13, 46

Zanzibar: Item Label: 2, 8, 39, 44, 51, 54, 65

It is very important to emphasize here that in a traditional (or Classical) approach to test construction it would only be possible to make valid comparisons across countries if the same test items were used to calculate total test scores. However, this is not a problem for the Rasch Model described in this report provided that: (i) the test items used in the analysis fit the Rasch Model, and (ii) there are sufficient common items across countries to ensure that a common scale and origin can be established. These two conditions applied in this research and therefore it was possible to make valid comparisons of HAKT scaled scores across SEACMEQ countries. The test items were banked for use in subsequent SEACMEQ projects.

5.0 THE STUDY POPULATION FOR SACMEQ IV PROJECT

5.1 Desired target population

The desired target population definition for *SACMEQ IV* Project was exactly the same (except for the year) as was employed for the *SACMEQ II* and *III* Projects. This consistency was maintained in order to be able to make valid cross-national and cross-time estimates of “change” in the conditions of schooling and the quality of education.

The desired target population definition for *SACMEQ IV* Project is as follows:

“All learners at Grade 6 level in 2013 (at the first week of the eighth month of the school year) who were attending registered mainstream (primary) schools.”

5.2 Excluded target population

One of the rules followed by SEACMEQ for ensuring valid data in large-scale studies is that no more than **5** percent of the learners in the desired target population may be excluded from the defined target population. Like in *SACMEQ II* and *III*, special schools which provide education to learners with severe educational needs were excluded from *SACMEQ IV* sample. “Small” mainstream schools which had less than **15** learners enrolled in Grade 6 in 2013 were also allocated to the excluded population to reduce data collection costs – without the risk of leading to major distortions in the study population.

5.3 Defined target population

The “defined target population” was constructed by removing the “excluded target population” from the “desired target population”. In **Table 19** the numbers of schools and learners in the desired, defined and excluded populations have been presented.

Table 19: Desired, Defined, and Excluded Populations for SACMEQ IV Project

School System	Desired		Defined		Excluded		Excluded Pupils %
	Schools	Pupils	Schools	Pupils	Schools	Pupils	
Botswana	802	45,556	676	43,414	126	2,142	4.70
Kenya	22,505	790,111	18,947	757,705	3,558	32,406	4.10
Lesotho	1,385	52,212	1,072	49,590	313	2,622	5.02
Malawi	5,561	380,539	4,513	378,396	1,048	15,720	4.13
Mauritius	308	22,630	298	22,249	10	381	1.68
Mozambique	4,185	406,777	3,841	400,607	344	6,170	1.52
Namibia	1,145	55,727	948	53,108	197	2,619	4.70
Seychelles	26	1,365	25	1,364	1	1	0.07
South Africa	17,280	929,341	13,156	886,073	4,124	43,268	4.70
Swaziland	571	32,134	498	30,913	73	1,221	3.80
Tanzania	-	-	-	-	-	-	-
Uganda	15,428	888,655	12,585	846,662	2,843	41,993	4.73
Zambia	7,253	370,800	5,971	353,413	1,282	17,387	4.69
Zanzibar	285	32,444	261	32,176	24	268	0.83
Zimbabwe	5,638	349,816	4,836	337,367	802	12,449	3.56
SACMEQ IV	82,372	4,358,107	67,627	4,193,037	14,745	178,647	4.10

From the last column of **Table 19** it can be observed that the excluded population of learners was less than the stipulated 5 percent for each country to meet the SEACMEQ criteria for accuracy in large-scale assessment data.

6.0 SAMPLING PROCEDURE FOR SACMEQ IV PROJECT

6.1 Sampling and sample characteristics

A two-stage sampling design was employed. In the first stage schools in the defined target population were sampled on a “Probability-Proportional-to-Size” (PPS) basis from sampling frames that individual countries submitted to the SEACMEQ Coordinating Centre. In the second stage of sampling learners were sampled from all the Grade 6 classes in each of the sampled schools using Simple Random Sampling. Computer-generated random numbers were used to facilitate the sampling of pupils. Twenty five (25) learners (minimum cluster size) were sampled where the total number of all enrolled Grade 6 learners at the time of data collection was greater than 25. Where the number of Grade 6 learners was 25 or less than 25 in a school, all the Grade 6 learners were included in the sample.

For a detailed account of how the sampling of schools and learners was carried out, including the software that was used in the SACMEQ IV project the reader may refer to Ross and Saito (in press). The numbers of schools and learners in the planned and actually achieved sample have been presented in **Table 20**.

Table 20: Planned and achieved samples for SACMEQ IV project

School System	Schools		Pupils	
	Planned	Achieved	Planned	Achieved
Botswana	188	188	4,700	4,562
Kenya	230	224	5,750	5,325
Lesotho	185	182	4,426	4,378
Malawi	126	126	3,400	3,223
Mauritius	153	133	3,825	3,321
Mozambique	200	189	5,000	4,820
Namibia	292	285	7,300	7,261
Seychelles	25	25	1,364	1,303
South Africa	305	295	7,625	7,117
Swaziland	145	145	3,625	3,592
Tanzania	-	-	-	-
Uganda	245	245	6,125	5,261
Zambia	160	154	4,000	3,816
Zanzibar	126	126	3,150	3,150
Zimbabwe	213	190	5,250	5,089
SACMEQ IV	2,593	2,507	65,540	62,218

6.2 Response rates, design effects, effective sample sizes

The size and the quality of the sample are critical to the accuracy of the research. The response rate, the design effect and the effective sample size are some of the characteristics that SEACMEQ monitors in all the projects. The response rates, design effects and effective sample sizes for SACMEQ IV project have been presented in **Table 21**.

Figures in the first two columns under the heading “Response Rate (%)” in **Table 21** are the response rates for schools and learners, respectively. The third, fourth and fifth columns under the heading “Design Effects” are numbers (ratios) that indicate the amount of “sampling error” associated with the two-stage sample for each of Reading, Mathematics and HAKT estimates. Columns six, seven and eight under the heading “Effective Sample Sizes” are numbers of sample units (learners) in a simple random sample that would give the same level of accuracy as the two-stage sample that was used in the study for each of Reading, Mathematics and HAKT.

Table 21: Response rates, design effects, effective sample sizes for SACMEQ IV

School System	Response Rate (%)		Design Effect			Effective Sample Size		
	Schools	Pupils	Reading	Maths	HAKT	Reading	Maths	HAKT
Botswana	100.0	97.1	8.83	8.41	7.84	423	444	476
Kenya	97.4	92.6	10.67	9.63	9.56	286	317	319
Lesotho	98.4	98.9	10.89	8.14	6.60	368	493	608
Malawi	100.0	94.8	6.99	3.63	6.03	180	347	209
Mauritius	86.9	86.8	4.56	5.24	5.48	370	322	307
Mozambique	94.5	96.4	10.20	14.33	8.97	339	241	386
Namibia	97.6	99.5	4.82	5.46	5.71	1114	983	940
Seychelles	100.0	95.5	11.95	8.34	9.25	92	132	119
South Africa	96.7	93.3	7.53	7.36	6.43	510	525	602
Swaziland	100.0	99.1	5.33	4.64	6.14	427	491	371
Tanzania	-	-	-	-	-	-	-	-
Uganda	100.0	85.9	5.02	4.48	4.37	210	235	241
Zambia	96.3	95.4	7.80	5.30	6.54	359	528	429
Zanzibar	100.0	100.0	3.25	2.52	3.21	630	814	639
Zimbabwe	89.2	96.9	7.97	7.56	5.86	364	384	495
SACMEQ IV	96.7	94.9	7.56	6.79	6.57	405	447	439

The following observations can be made from **Table 21**:

Response rate in surveys refers to the percentage of the total sample units that were planned who actually participate in the study. The SEACMEQ rule is that the overall response rate for both the schools and the learners should not be less than 90%. Only Kenya and Mauritius fell negligibly short of meeting this standard.

Design effect is a number (ratio) which indicates the amount of “sampling error” that is introduced by the use of a clustered (two-stage) sampling method in relation to the “sampling error” that would result if a simple random sample of the same size had been used. Alternatively, the “design effect” is the ratio of the variance (of the sample mean) for a multi-stage sample to the variance for a simple random sample of the same size. Generally, the inaccuracy associated with a multi-stage sample is many times greater than the inaccuracy associated with a simple random sample of the same size. The design effects for the three tests in all countries lie between 2 and 12; this is acceptable for this kind of large scale cross-national study.

Effective sample size is calculated from the design effect. It is the size of a simple random sample that would be required to give the same level of accuracy as the given multi-stage sample. Generally, the “Effective Sample Size” will be smaller than the given actual multi-stage sample.

The sample designs used in *SACMEQ IV* Project were selected so as to meet the standards set by the International Association for the Evaluation of Educational Achievement (IEA). These

standards require that sample estimates of important learner population parameters in multi-stage designs should have sampling accuracy that is at least equivalent to a simple random sample of 400 learners (thereby guaranteeing 95 percent confidence limits for sample means of plus or minus one tenth of a learner standard deviation unit). The effective sample sizes for Seychelles are very small because pupils in that country were not sampled; rather the entire population was included in the study due to its small size.

6.3 The accuracy of sampling

The accuracy of a sampling procedure is not related to the proportion of the sample size to the population, but rather to the error limit associated with the sampling technique. This error limit is determined by the sample size in simple random sampling method and by the effective sample size (ESS) in complex sampling. The effective sample size, sample size, and design effect are related by the formula:

$$ESS = \text{sample size} \div \text{design effect (DEFF)}$$

The ESS is equal to 400 at 95 percent confidence limits of ± 5 percent for percentages and ± 0.1 pupil standard deviation units for means.

6.4 Homogeneity and roh

Homogeneity statistics test the assumption that the statistical properties of any one part of an overall dataset are the same as any other part. In other words, it indicates how much of the overall variation in measurement of a trait (such as mean scores) can be attributed to variation among different clusters. In SEACMEQ studies, homogeneity is estimated by calculating the coefficient of intraclass correlation (roh) which varies between 0 and 1. The design effect (DEFF) is a function of roh given by: $DEFF = 1 + (b-1)roh$, where b is the cluster size. This means that the expression for ESS given above becomes: $ESS = n / (1 + (b-1)roh)$, where n is the sample size. Rearranging this formula gives us the formula for calculating the sample size as:

$$n = ESS \times (1 + (b-1)roh); \text{ if the effective sample size is known.}$$

The values of roh for the tests per country are given in **Table 22**. Taking Botswana pupil reading scores as an example, **Table 22** shows that **37%** (.37) of the variation is due to differences among schools; hence **63%** is accounted for by within school differences.

Table 22: Values of the coefficient of intraclass correlation for the tests used in SACMEQ IV project

School System	Intra-class Correlation Coefficient (Roh)		
	Reading	Mathematics	HAKT
Botswana	.37	.31	.33
Kenya	.46	.43	.39
Lesotho	.42	.30	.30
Malawi	.39	.20	.37
Mauritius	.27	.26	.32
Mozambique	.45	.36	.34
Namibia	.44	.34	.37

Seychelles	.18	.11	.20
South Africa	.54	.51	.43
Swaziland	.33	.26	.34
Tanzania	-	-	-
Uganda	.50	.46	.43
Zambia	.37	.26	.36
Zanzibar	.22	.17	.21
Zimbabwe	.47	.43	.33
SACMEQ IV	.23	.22	.09

7.0 DATA COLLECTION PROCEDURE FOR SACMEQ IV PROJECT

In this report “Data Collection” includes preparations before the field work, the actual field work and activities that followed field work. Preparations for data collection focused on instrument review, communication to schools, printing and distribution of instruments and training of data collectors.

7.1 Instrument review

As soon as the 2011 *SACMEQ* Assembly of Ministers took a decision to conduct *SACMEQ IV* project in 2013, the National Research Teams (NRTs) under the auspices of the *SACMEQ* Coordinating Centre in Paris, set out to prepare and update the instruments (tests and questionnaires). Between 2012 and 2013 the *SACMEQ* Coordinating Centre hosted at least three working sessions for the NRTs in Nairobi (Kenya), Lusaka (Zambia), and Pretoria (South Africa) that were focused on reviewing existing test items and ensuring that, where there had been curriculum changes, the items were still relevant. Invariably, there were no significant changes on the Reading, Mathematics and Health Knowledge test items. *SACMEQ IV* test items were piloted, first, in a few primary schools in South Africa, and then in individual member countries. The pilot study was intended to ensure that the language in *SACMEQ IV* tests was accessible to learners, that there were no cultural biases in the items and learners comprehended how to write their responses.

In some countries the tests were subsequently translated into respective language(s) of instruction (Kiswahili, Portuguese). Care was taken to ensure that the English and other languages used for the tests were equivalent to avoid unfair advantage in any of the language(s).

The final statistical and content validity and reliability checks of the instruments were carried out by NRTs and specialists at the *SACMEQ* Coordinating Centre who then declared the instruments ready to print and take to the field.

7.2 Communication to schools

Officials in the respective Ministries of Education informed the sampled schools through the Regional offices during mid-2013. The National Research Teams were responsible for distributing the data collection schedules, intensifying and monitoring communication to schools and among data collectors.

7.3 Printing and distribution of data collection instruments

Data collection instruments included a) School Head Booklets, b) School Information Booklets, c) Teacher Booklets, d) Pupil Booklets e) Pupil Name Forms and f) School Forms. Each participating country received print-ready copies from the Coordinating Centre and was responsible for printing correct numbers of copies for their respective schools.

When all instruments were printed, the NRTs conducted a “hand check” of all materials so as to verify that there were no missing/extra pages, misprints or omissions. All work related to the printing and packaging of the data collection instruments was undertaken under strict security arrangements – so that there was no possibility of “leakage” of information about the content of the learner and teacher Reading, Mathematics, and Health Knowledge tests.

The printed materials were distributed to leaders of research teams that were assigned to collect data in each school. The Team Leaders were responsible for checking the accuracy of the instruments in terms of correctness of numbers and languages before carrying the instruments to the schools. The first level of checking was done during data collection training sessions. The data collectors were charged with further and final checks a day before the data collection.

7.4 Training of data collectors

On the first day of training the NRT presented a “simulated” data collection exercise in which they acted as a data collector and the trainees took the roles of learners, teachers, and School Heads. The second day involved an intensive study of the Manual for Data Collectors. This document sets out, in sequential order, all of the actions to be taken by the data collector from the time of receiving packages of data collection instruments from the Ministry of Education to the time when the data collector had completed the data collection and was preparing all materials for return. The third day involved a second “simulated” data collection whereby the trainees supervised a full-fledged data collection in several schools that were not involved in the main data collection. The experiences gathered during these exercises were shared and discussed during a meeting so that all data collectors understood the procedures to be completed within schools.

7.5 The main data collection

“Main Data Collection” in this report refers to the actual field work. Three data collectors were assigned three sampled schools to carry out the data collection exercise. Special effort was made to ensure that data collection was conducted according to explicit and fully-scripted steps so that the same verbal instructions were used (for learners, teachers, and School Heads) by the data collectors in all sample schools, in all countries, and for each aspect of the data collection. This was a very important feature of the study because the validity of cross-national comparisons

arising from the data analyses depended, in large part, on achieving carefully structured and standardized data collection environments.

The main *SACMEQ IV* data collection occurred for most *SACMEQ* Ministries of Education in the period September to December 2013.

Two days of data collection were required for each sampled school. On the first day the data collectors had to sample learners from all the Grade 6 classes in the sampled schools, using a list of provided random numbers. The sampled learners were then given the Pupil Questionnaire, the HAKT and the Reading tests. Part of the Pupil Questionnaire required learners to get confirmation of the accuracy of the information from their parents; therefore the questionnaires were taken home by the pupils and returned the following day. On the second day the pupils were given the Mathematics test.

In addition to completing a questionnaire, a teacher who taught the majority of the sampled learners for each of Reading, Mathematics and Life skills\Health also took the respective test(s).

The data collectors were provided with a 40-point checklist in order to ensure that they completed all important tasks that were required before, during, and after their visits to schools. Each task was cross-referenced to specific pages of instructions in the data collectors' manual. The data collectors also checked all completed questionnaires (Pupil, Teacher, and School Head) and, if necessary, obtained any missing or incomplete information on the second day before they left the school. The materials were then handed over to the Regional Coordinator for safekeeping, "hand editing" and dispatching to the National Research Coordinator (NRC) at the Ministry of Education as soon as all data collection was completed.

8.0 DATA ENTRY PROCEDURE FOR *SACMEQ IV* PROJECT

8.1 Installation of DME for *SACMEQ IV* data entry

It is a standard practice that SEACMEQ study data collected in each participating countries are uploaded into several computers set up in respective countries. A special data capturing software called DME was developed for *SACMEQ IV* study and sent to all participating countries to be installed in each computer. This was accompanied by a detailed instruction on how to install and use the software.

Given in the following part of this report are some important features and procedures for DME.

8.2 DME codebooks

The following is an explanation of how the DME codebook files are named, using Botswana as an example.

Figure 1: DME coding for SACMEQ IV project

Instrument number	Instrument name	Country	Study	Month of study	Year of study
01	PBK	BOT	S4	Sept	2013
02	TBK	BOT	S4	Sept	2013

For the Botswana example, the files are named as follows:

Pupil Booklet:

- 01_PbK_BOT_S4_Sept2013.dme

Teacher Booklet:

- 02_TbK_BOT_S4_Sept2013.dme

School Head Booklet:

- 03_Shbk_BOT_S4_Sept2013.dme

School Form:

- 04_Sf_BOT_S4_Sept2013.dme

Pupil Name Form:

- 05_Pnf_BOT_S4_Sept2013.dme

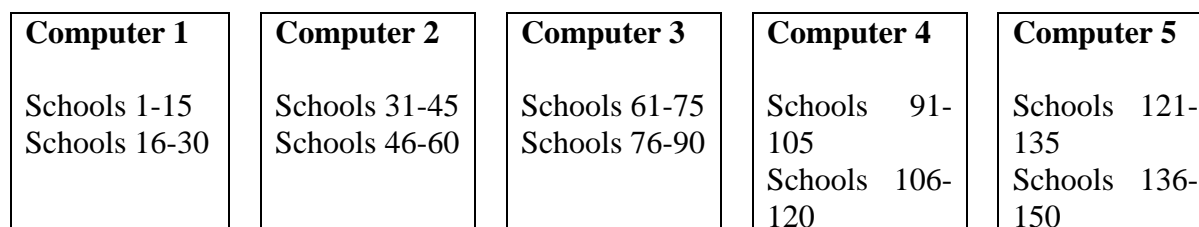
School Information Booklet:

- 06_Sibk_BOT_S4_13Sept2013.sdb

8.3 Data entry arrangement

It is proven to be very helpful to dedicate a computer to enter data from a specified group or sub-population. For example, one may use a computer for each region. This would reduce the confusion when compiling files from different computers. The arrangement is illustrated diagrammatically in **Figure 2**.

Figure 2: Diagram of use of computers 1 to 5



8.4 General instructions for data punchers

Clear instructions were prepared for data punchers to guide them in the following areas:

- Separating the different parts (parts A, B, C, D, E) of the paper version of Pupil Booklet and Teacher Booklet in the electronic version by entering a specified code for each part. It is important to enter not only the Instrument codes but also the codes associated with the parts;
- The User Name and Password to be used by data punchers (which were different from those of the supervisor/NRC);
- How to progress from one field to the next when entering data;
- How to save the entered data.

8.5 Instructions for missing responses

The National Research Coordinators are required to verify data before data entry commences, hence there should be no missing information in the instruments. However, if some missing information is encountered, the rules below were followed unless otherwise indicated specially:

- Questions which have been administered but contain no information should be coded as 9 (if the response requires 1 digit), 99 (if the response requires 2 digits), 999 (if the response requires 3 digits), etc.;
- Questions which have not been administered should be coded as 8 (if the response requires 1 digit), 98 (if the response requires 2 digits), 998 (if the response requires 3 digits), etc. If the whole instrument was missing, use this code.

8.6 Coding rules for Pupil Booklet

- *Multiple response code for Pupil Reading, Mathematics, and Health Tests* – The code 7 is entered when more than one response is given as an answer for any test item. This does not apply to the questions in the questionnaires;
- *Check digits in Pupil Reading, Mathematics, and Health Tests* – A “check digit” was inserted at several stages in the tests. These check digits were used to make sure that data enterers were ‘awake’. Only the specified letter should be entered for each check digit;
- Special attention was paid to filter questions, such as PB#35. For example, if the response to PB#35 is “1”, then a code “8” was entered for PB#35.1 to PB#35.6.

8.7 Coding rules for Teacher Booklet

- For the Teacher ID section, if any of the grade 6 English or Math or Health classes is left blank, then “1” (for no) is entered for the classes, and “98” is entered for the class size;
- *Multiple response code for Teacher Reading, Mathematics, and Health Knowledge Tests* – The code 7 is entered when more than one response is given as an answer for any test item;
- *Check digits in Teacher Reading, Mathematics, and Health Knowledge Tests* – A “check digit” was inserted at several stages in the tests. These check digits were used to make sure that data enterers were ‘awake’. Only the specified letter should be entered for each check digit.

8.8 Coding rules for School Information Booklet

- For the lines not used for the session information (SIB#12), the not-administered codes were entered. For example, if there was only one session which existed in a school, the lines for 2nd through 6th sessions were coded 9998 for the number of pupils per session and 98 for the number of classes per session.

8.9 Coding rules for School Head Booklet

- If option 1 was selected for SHB#10 and SHB #11, then the boxes at the end of option 2 was coded “98”. If option 2 was selected, but the boxes are left blank, then it was coded “99”;
- Extra care was exercised with questions SHB#49.01 to SHB#49.15. These questions consisted of five data entry fields and not six (see **Figure 3**). Field A is for the response to “This happened at my school this year”. ‘1’ is entered if the response is “No” or ‘2’ if the response is “Yes”. For fields B, C, D and E the coding is different. The code ‘1’ was entered if the box was not ticked, or code ‘2’ if the box was ticked.

Figure 3: Question 49 in School Head Booklet

		This happened at my school this year		With the assistance of			
		No	Yes	Ministry (Dep't) of Education	Ministry (Dep't) of Health	Other Govern't Org'ns	Other Groups ¹
Support to pupils		A		B	C	D	E
49.01	Guidance / counselling for orphans and vulnerable pupils	(1)	(2)	(2)	(2)	(2)	(2)
49.02	Guidance / counselling for pupils with HIV and AIDS	(1)	(2)	(2)	(2)	(2)	(2)

The special codes have been summarized in **Table 23**.

Table 23: Special codes for SACMEQ IV project

Code	When to use
7	<u>Multiple response code</u> : Use this when more than one option is ticked in test items.
9, 99, 999, etc.	<u>Missing response code</u> : Use this when no response is given from the administered respondents
8, 98, 998, etc.	<u>Not-administered code</u> : Use this when the question was not encountered by the respondent.
A, B, C, etc.	<u>Check letter code</u> : Use this when check letter needs to be inserted during any test.

8.10 Special instructions on data verification after data entry

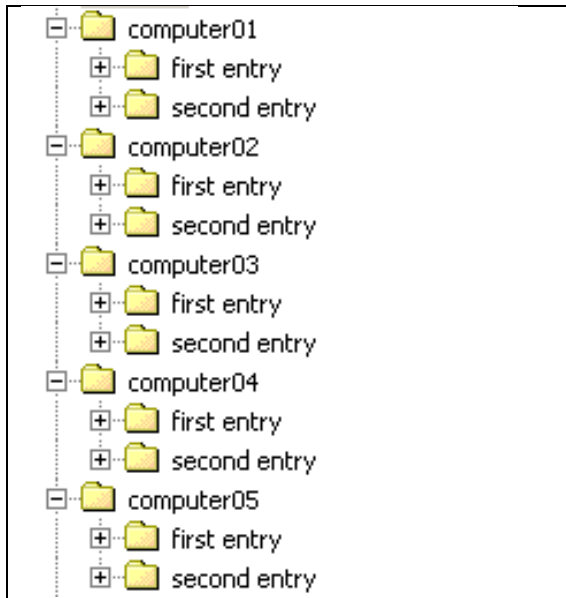
The NRCs/supervisors were required to independently verify the entered data in all computers after data. They started this by logging out of DME and re-opening it by using their own special User Name and Password. The original diagram illustrating which computers were used for which data entry was referred to during verification. The verification steps are given below.

Step 1: Collect files from different computers

After data were entered using several computers, they were merged into a single file for each instrument; for each of the first and the second entry.

First, several sub-directories were made on a diskette or USB; representing different computers. Since the double-entry of data was used, two sub-directories within each sub-directory were made: (1) first entry; and (2) second entry. The structure of sub-directories is shown in **Figure 4**.

Figure 4: Sub-directories to be created to copy different files



DME files from all computers were copied into the corresponding sub-directories.

Step 2: Create a New Folder on one computer

One computer is assigned as the parent computer, say, computer01. On this computer, subfolders for each computer that was used for data entry were created first. A USB was then used to copy and paste data files (.dme) from each computer on the corresponding subfolder.

Step 3: Merging data files

After completing the copying of data files, the DME software was opened. The following instructions were followed.

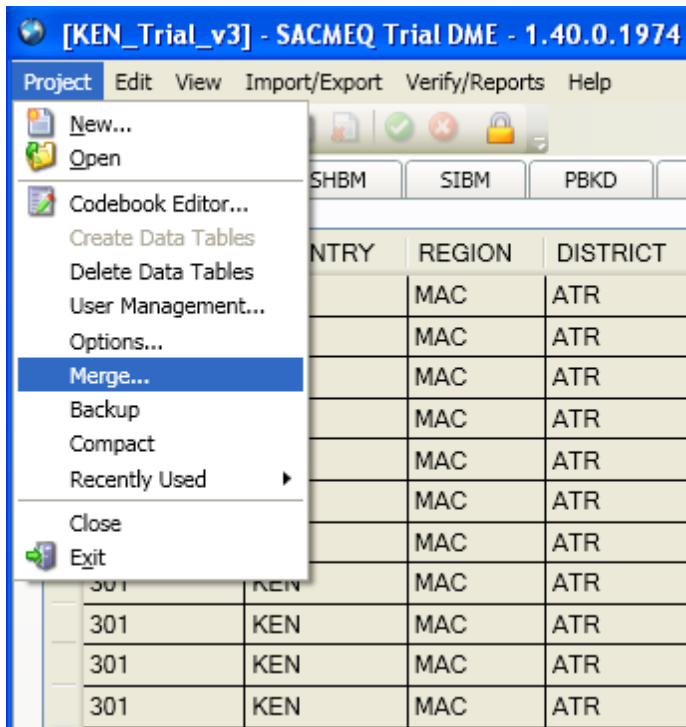
- (i) Locate and open DME code book in subfolder '**computer1**'. **Your screen will look like this...**

SACMEQ IV Project Technical Report

PUNCHER	COUNTRY	REGION	DISTRICT	SCHOOL	SCORER	MARKER	PUPIL	INSTR	PART_A	ChkD_RA1	ChkD_MA1	ChkD_RA2	ChkD_MA2	ChkD_RB1
301	KEN	MAC	ATR	12330122	201	.	10	1	A	RB1
301	KEN	MAC	ATR	12330102	207	.	10	1	A	RB1
301	KEN	MAC	ATR	12330102	205	.	18	1	A	RB1
301	KEN	MAC	ATR	12330102	207	.	22	1	A	RB1
301	KEN	MAC	ATR	12330102	208	.	2	1	A	RB1
301	KEN	MAC	ATR	12330102	208	.	14	1	A	RB1
301	KEN	MAC	ATR	12330102	208	.	26	1	A	RB1
301	KEN	MAC	ATR	12330102	208	.	6	1	A	RB1
301	KEN	MAC	ATR	12330102	201	.	3	1	A	RB1
301	KEN	MAC	ATR	12330122	201	.	1	1	A	RB1
301	KEN	MAC	ATR	12330134	207	.	2	1	A	RB1
301	KEN	MAC	ATR	12330134	207	.	22	1	A	RB1
301	KEN	MAC	ATR	12330134	207	.	10	1	A	RB1
301	KEN	MAC	ATR	12330134	207	.	6	1	A	RB1
301	KEN	MAC	ATR	12330134	208	.	14	1	A	RB1
301	KEN	MAC	ATR	12330134	208	.	26	1	A	RB1
301	KEN	MAC	ATR	12330134	208	.	18	1	A	RB1
301	KEN	MAC	ATR	12330113	207	.	6	1	A	RB1
301	KEN	MAC	ATR	12330113	207	.	10	1	A	RB1
301	KEN	MAC	ATR	12330113	207	.	14	1	A	RB1
301	KEN	MAC	ATR	12330113	208	.	2	1	A	RB1
301	KEN	MAC	ATR	12330113	208	.	18	1	A	RB1

Number of records in this instrument (PBKM)

(ii) Click project, then merge ...



- (iii) Repeat (i) and (ii) until you have merged all data files.
- (iv) At the end of the merging exercise, each country would have twelve (6 for first entry, and 6 for second entry) combined data files. Each entry should have:
- 1 combined file for Pupil Booklets (# of records = # of pupils surveyed),
 - 1 combined file for Teacher Booklets (# of records = # of teachers surveyed),
 - 1 combined file for School Information Booklets (# of records = # of schools surveyed),
 - 1 combined file for School Head Booklets (# of records = # of schools surveyed),
 - 1 combined file for School Form (# of records = # of existing classes in all the surveyed schools), and
 - 1 combined file for Pupil Name Form (# of records = # of pupils surveyed)

These data files were used to do the rest of the cleaning procedures within DME.

References

Andrich. D., Sheridan. B. S., & Luo. G. (2003). *RUMM2020: Rasch Unidimensional Models for Measurement*. RUMM Laboratory: Perth. Western Australia.

International Bureau of Education (2006). *Manual for Integrating HIV and AIDS Education in School Curricula*. Paris: UNESCO

Ross N., Saito M., Dolata S., Ikeda M., Zuze L., Murimba S., Postlethwaite N.T., Griffin P. (2006). *The Conduct Of The SACMEQ II Project*. Paris: SACMEQ

Ross N., Saito M., Dolata S., Ikeda M., Zuze L., Murimba S., Postlethwaite N.T., Griffin P. (2006). *The Construction of Reading Tests for the SACMEQ II Research Programme Conducted by SACMEQ*. Paper presented at the 2006 World Bank Human Development Forum. Washington DC, October 31.

UNAIDS (2007). *Core Indicators for National AIDS Programmes*. Geneva: Author