Part C: The Construction of Tests for the SACMEQ II Project

The Main Steps in Test Construction

The following discussion provides information about the construction of the SACMEQ II reading and mathematics tests for pupils and teachers, and the scaling procedures that were used to calibrate test items and to permit pupil and teacher performance to be described in terms of hierarchies of competencies. The procedures used to construct the SACMEQ I reading test for pupils have already been presented in the national policy reports prepared for the seven countries that completed this project (Kulpoo, 1998; Machingaidze et al, 1998; Milner et al, 2001; Nassor and Ali Mohammed, 1998; Nkamba and Kanyika, 1998; Nzomo et al, 2001; Voigts, 1998). The testing undertaken for the SACMEQ II Project was far more extensive than for the SACMEQ I Project – with both Grade 6 pupils and their teachers being given both reading and mathematics tests.

The test construction for both projects was undertaken carefully so as to ensure that the structure of the pupil tests was congruent with the content (domains) and behaviours (skills) derived from detailed analyses of the curricula, syllabi, exams, and textbooks used in the SACMEQ countries.

The SACMEQ II tests for pupils and teachers included "overlapping" test items selected from five earlier studies: the Zimbabwe Indicators of the Quality of Education Study (Ross, 1995), the SACMEQ I and SACMEQ II Projects, the IEA's Third International Mathematics and Science Study (TIMSS) (Mullis et al, 2001), and the IEA's International Study of Reading Literacy (IRL) (Elley, 1992). These "overlaps", when combined with Rasch item analysis and test scoring techniques, made it possible to make valid comparisons among the following groups of respondents: pupils with teachers in the SACMEQ II Project, pupils in the SACMEQ I Project with pupils in the SACMEQ II Project, and pupils in both SACMEQ Projects with pupils in the IEA's TIMSS and IRL studies. In Appendix F and Appendix G the overlaps of test items across all of these studies have been presented in tabular form. For example, the 66th reading test item listed in Appendix F was located in the SACMEQ I pupil test ("ptembo05"), the SACMEQ II pupil test ("pread17"), the SACMEQ II teacher test ("tread04"), and the pupil test used in the Zimbabwe Indicators of the Quality of Education Study ("tembo05").

In Figure 2.3 the key steps involved in constructing the SACMEQ II tests have been presented in diagrammatic form. The main aim of this process was to ensure high levels of face validity and construct validity by achieving congruence between the test blueprint (prepared as a framework for test construction) and the descriptions of increasing levels of competence generated from a Rasch analysis of the item difficulty levels in combination with a skills audit of test items.

The selection of teacher test items had to cover the full range of pupil item difficulties – but did not contain too many easy pupil test items. In addition, in order not to antagonize teachers with an extended testing session, the teacher tests had a much smaller number of test items than the pupil tests.

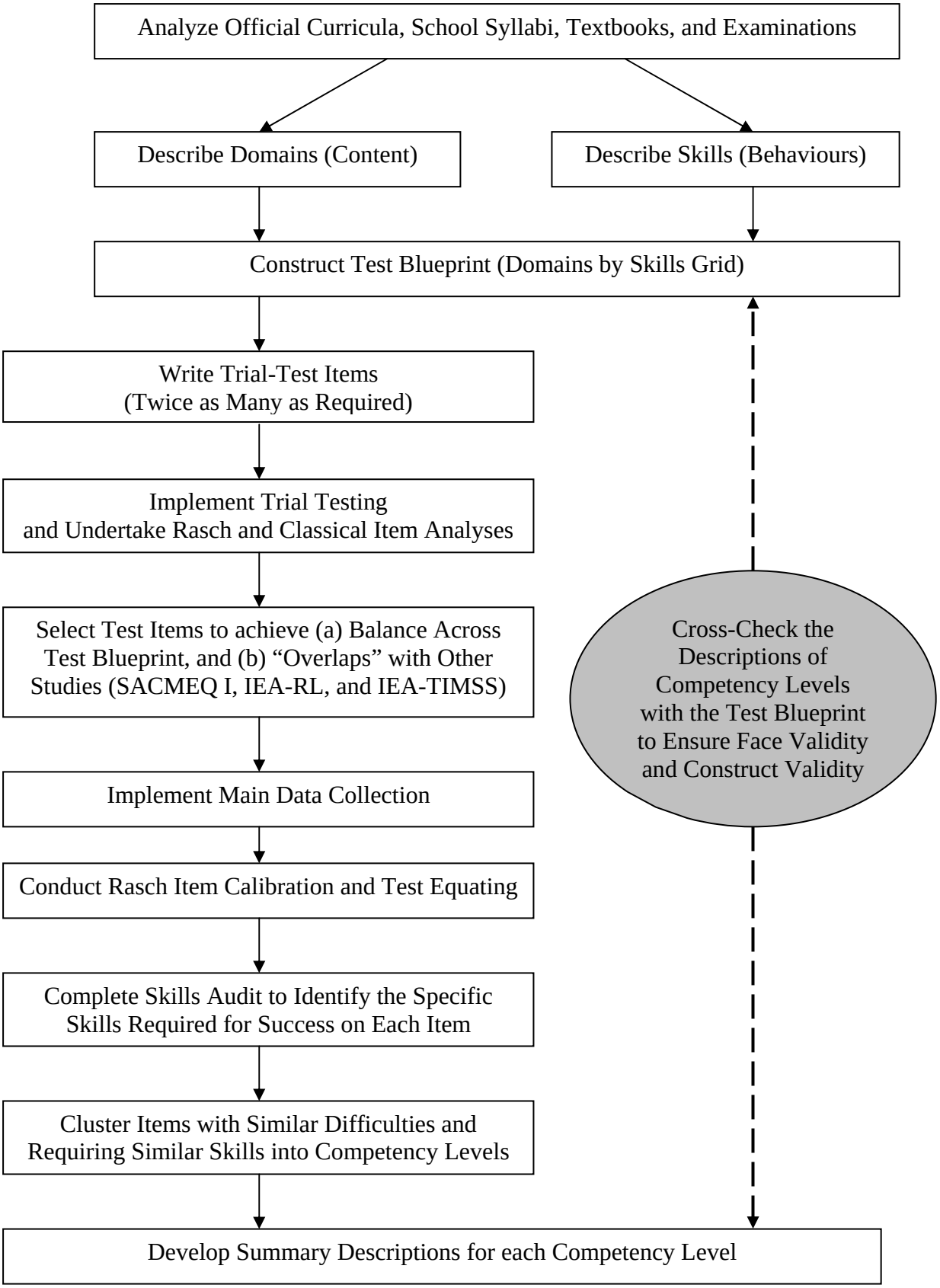

Figure 2.3: Main Steps Involved in Test Construction for the SACMEQ II Project

The Structure of the SACMEQ II Reading Tests

(a) The Definition of "Reading Literacy"

In the SACMEQ II Project "reading literacy" was defined as " the ability to understand and use those written language forms required by society and/or valued by the individual."

This was the agreed definition that was used by the 35 countries that participated in the International Reading Literacy Study that was conducted by the International Association for the Evaluation of Educational Achievement (Elley, 1992). It was also the general definition accepted by the SACMEQ National Research Coordinators (NRCs) for the SACMEQ I Project. The NRCs found this definition to be general enough to accommodate the diversity of traditions and languages represented in the SACMEQ countries, and yet still sufficiently specific to provide guidance for test construction.

(b) The Three Reading Domains

In both SACMEQ Projects there was an initial detailed curriculum analysis undertaken across all countries in order to define – after exhaustive discussion of the most important skills contained within the reading curricula at Grade 6 level - the reading skills that were considered by all countries to be the most important. The NRCs invested a great deal of time in this process because they wanted to enhance the validity of the tests by ensuring that they provided a balanced coverage of the main reading domains and the required reading skills. The NRCs decided to accept the three broad content domains for reading literacy (presented in Figure 2.4) that had been adopted for the International Reading Literacy Study, and also previously applied by the NRCs in the SACMEQ I Project.

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.

Figure 2.4: The Three Domains for the SACMEQ II Reading Test

(c) A "Proposed" Hierarchy of Reading Skills

It was decided that the construction of the SACMEQ II test should draw upon advanced psychometric procedures that would enable the establishment of a meaningful dimension of increasing competence that could be applied to both the SACMEQ I and II reading tests. This outcome was highly desirable because it permitted valid comparisons to be made of the reading performance of countries across the two projects.

Level 1: Pupils at this level should be able to link words and pictures where the pictures depict common objects of a "concrete" nature.

Level 2: Pupils at this level should be able to link words to more abstract concepts such as propositions of place and direction, and, perhaps, ideas and concepts such as comparatives and superlatives (happiest, biggest, below, etc.)

Level 3: Pupils at this level should be able to link words (such as a phrase or short sentence) from one setting to words in another setting where there is a word match between the two settings.

Level 4: Pupils at this level should be able to deal with longer passages of text that contain a sequence of ideas and content, and that require understanding derived from an accumulation of information gathered by reading forward.

Level 5: Pupils at this level should be able to read forwards or backwards through a text in order to: confirm understanding, or link new information with a piece of information encountered previously, or link ideas from separate parts of a text, or demonstrate the capacity to infer an author's intention.

Figure 2.5: The "Proposed" Skill Levels for the SACMEQ II Reading Test

The first step in achieving this result was to undertake an intensive examination of curricula in order identify descriptive skill levels that would define a recognizable and meaningful dimension. This dimension, taken in combination with the three domains of reading, formed a framework (or blueprint) for the construction of suitable test items. Five reading skill levels were identified as shown in Figure 2.5. This step may be described as building a "proposed" hierarchy of reading skills.

(d) Constructing the Test Blueprint by Combining Domains with Skill Levels

The NRCs recognized that each of the skill levels specified in Figure 2.5 needed to be carefully interpreted within the context of the kind of text (or reading domain) that was being encountered by pupils. That is, for any single level, the description of the skills had to be refined in order to more closely reflect whether the reader was dealing with, for example, a fictional story (narrative), a factual account (expository), or a graph, chart, or diagram (document). The NRCs deliberated on this matter for some time and then proceeded to examine the intersections of the three reading domains (Figure 2.4) with the five skill levels (Figure 2.5) in order to form the test blueprint (Figure 2.6).

The skill descriptions in each cell of the blueprint showed how reading behaviour increased in complexity through each of the five skill levels for each of the three domains. The numbers of items in the cells in Figure 2.6 were approximately in proportion to the time spent on parts of the reading curriculum in the SACMEQ countries, and they reflected the advice received from national curriculum experts.

Skill Level	Reading Domain			
	Narrative	Expository	Documents	
Level 1	Word/picture association involving positional or directional prepositions requiring the linkage of a picture to a position or a direction in order to answer the question	Word/picture association involving positional or directional prepositions requiring the linkage of a picture to a position or a direction in order to answer the question	Word/picture association involving positional or directional prepositions requiring the linkage of a picture to a position or a direction in order to answer the question	
Items	2	2	2	6
Level 2	Recognising the meaning of a single word and being able to express it as a synonym in order to answer the question	Recognising the meaning of a single word and being able to express it as a synonym in order to answer the question	Linking simple piece of information to item or instruction	
Items	7	6	9	22
Level 3	Linking information portrayed in sequences of ideas and content, when reading forward	Linking information portrayed in sequences of ideas and content, when reading forward	Systematic search for information when reading forward	
Items	8	10	8	26
Level 4	Seeking and confirming information when reading backwards through text	Seeking and confirming information when reading backwards through text	Linking more than one piece of information in different parts of a document	
Items	9	5	4	18
Level 5	Linking ideas from different parts of text. Making inferences from text or beyond text, to infer author's values and beliefs	Linking ideas from different parts of text. Making inferences from text or beyond text.	Use of embedded lists and even subtle advertisements where the message is not explicitly stated	
Items	6	3	2	11
Total Items	32	26	25	83

Figure 2.6: The Test Blueprint for the SACMEQ II Pupil Reading Test

In the final version of the SACMEQ II reading test there was a total of 83 test items, with (a) 32, 26, and 25 items allocated to the narrative, expository, and documents domains, respectively; and (b) 6, 22, 26, 18, and 11 items set at skill levels 1 to 5, respectively.

The Structure of the SACMEQ II Mathematics Tests

(a) The Definition of "Mathematics Literacy"

In the SACMEQ II Project "mathematics literacy" was defined as "the capacity to understand and apply mathematical procedures and make related judgements as an individual and as a member of the wider society."

This broad interpretation – with an emphasis on both understanding and decision-making – was prepared to ensure that the mathematics tests were not overly concentrated on mechanical rules and calculations. It was derived by the NRCs following an analysis of the mathematics content domains specified by the International Association for the Evaluation of Educational Achievement (IEA) (Mullis et al., 2001) and the Organization for Economic Cooperation and Development (OECD, 2000). These two frameworks were constructed with widespread participation and reviews by educators around the world – and took into consideration the intended (school system) curriculum, the implemented (school level) curriculum, and the attained curriculum. The IEA identified five mathematics domains: number, algebra (which at the primary school level was called "patterns, equations, and relationships"), measurement, geometry, and data. In contrast the OECD focussed on what their research teams referred to as the "big ideas" in primary school: "change and growth" and "space and shape".

(b) The Three Mathematics Domains

Unlike the OECD study - which focussed on more generic skills, the SACMEQ II Project was concerned with skills related to the school curriculum. The SACMEQ NRCs therefore used the slightly more detailed IEA domains as a beginning point for an extensive investigation of curricula, textbooks, and examinations for Grade 6 pupils within SACMEQ school systems. On the basis of this work the IEA framework was modified in order to bring it into alignment with what was actually being taught in SACMEQ classrooms in Southern and Eastern Africa. The first IEA domain, "number", was retained. The second, "algebra", was not seen as being relevant at the Grade 6 level in African schools, and was therefore removed. The third, "measurement", was retained. The fourth, "geometry", was re-expressed by the NRCs as "space" and then combined with the IEA's fifth domain of "data" to form a domain of "space-data".

The final domains selected by the NRCs for the SACMEQ II mathematics tests were focussed on the three areas listed in Figure 2.7.

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.

Figure 2.7: The Three Domains for the SACMEQ II Mathematics Test

(c) A "Proposed" Hierarchy of Mathematics Skills

A detailed investigation of the tasks given to pupils (problems, exercises, test questions, exam questions, etc.) across curricula for the 15 SACMEQ school systems enabled the NRCs to specify a set of descriptive skill levels that defined a recognizable and meaningful dimension of mathematics performance. A total of five mathematics skill levels were identified as shown in Figure 2.8. This step may be described as building a "proposed" hierarchy of mathematics skills.

Level 1: Pupils at this level should be able to identify simple shapes and link simple patterns and shapes to simple digits, to recognize units of measurement, to name basic shapes, and to undertake simple single operations using up to two-digit numbers.

Level 2: Pupils at this level should be able to recognize simple fractions in both numerical and graphical forms, to identify data presented in tables, to make basic calculations using simple measurement units, and to understand numeration with simple computations.

Level 3: Pupils at this level should be able to extend and complete number patterns, to translate shapes and patterns, and to convert measurement units when making simple single-step calculations.

Level 4: Pupils at this level should be able to combine operations in order to link information from tables and charts in performing calculations, to apply two or three-step number operations applied to measurement and conversion problems, and to identify and use appropriate information in the subsequent steps of a calculation.

Level 5: Pupils at this level should be able to make calculations and interpretations linking data from tables and graphs, and to make computations involving several steps and a mixture of operations using fractions, decimals, and whole numbers.

Figure 2.8: The "Proposed" Skill Levels for the SACMEQ II Mathematics Test

(d) Constructing a Mathematics Test Blueprint by Combining Domains with Skill Levels

The NRCs followed the approach used for the construction of the reading tests by combining the mathematics skill levels with mathematics domains to develop a test blueprint. For each skill level this provided a tighter definition of competencies by linking mathematical content to mathematical skills. The level of complexity and skill required also increased within each domain from level to level.

Following extensive discussions and consultations with mathematics education specialists, the NRCs produced Figure 2.9 – which linked domains with skills and gave guidance for the numbers of test items that were required. There were fewer test questions for the mathematics test because each item represented an individual separate task – whereas the reading test was actually based on single passages of text – each of which was attached to sets of items. In the final version of the SACMEQ II pupil mathematics test there was a total of 63 test items, with 27, 18, and 18 items allocated to the number, measurement, and space-data domains, respectively, and 6, 20, 17, 12, and 8 items set at skill levels 1 to 5, respectively.

Skill Level	Mathematics Domain			
	Number	Measurement	Space-Data	
Level 1	Recognize numbers. Link patterns to numbers.			
Items	6	0	0	6
Level 2	Apply single operations to two digit numbers or simple fractions.	Recognize units of measurement. Apply basic calculations using simple measurement units.	Link patterns and graphs to single digits. Recognize and name basic shapes.	
Items	8	8	4	20
Level 3	Extend and complete number patterns.	Convert measurement units when undertaking one-step operations.	Translate shapes and patterns. Identify data in tabular form.	
Items	6	4	7	17
Level 4	Combine arithmetic operations in order to link information from tables and charts when performing calculations.	Apply two and three-step arithmetic operations to numbers. Use and convert measurement units.	Combine arithmetic operations in order to link information from tables and charts.	
Items	4	4	4	12
Level 5	Combine operations in order to make calculations involving several steps and a mixture of operations using combinations of fractions, decimals, and whole numbers.	Combine operations in order to make calculations involving several steps and a mixture of operations using a translation of units.	Link data from tables and graphs in order to make calculations involving several steps and a mixture of operations.	
Items	3	2	3	8
Total Items	27	18	18	63

Figure 2.9: The Test Blueprint for the SACMEQ II Pupil Mathematics Test

Constructing "Overlapped Tests" to Use in Scaling

When the NRCs had completed the reading and mathematics test blueprints, they worked in teams to either select or write all of the required test items for the SACMEQ II tests. As items were prepared they were classified according to the cells in the test blueprints. For each cell twice as many items as required were prepared so that the rejection of poor items after the trial testing did not result in a shortage of items in some cells. Most test items were in multiple-choice format with four options per item. The item pools were then sent to all countries for review by panels of curriculum specialists. This resulted in editorial changes to the items and recommendations for additional items by the panel members who made sure that the items met the requirements of the respective national curricula.

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

In the case of the measurement of reading performance, there were three groups of respondents: the SACMEQ I pupils, the SACMEQ II pupils, and the SACMEQ II teachers. Each group completed a reading test that was "different but overlapped". That is, each group completed a reading test that contained some unique test items and some items that also appeared on one or both of the other two tests. In the case of numeracy measurement, the tests were also "different but overlapped", however there were only two groups of respondents: the SACMEQ II pupils and SACMEQ II teachers. The various overlaps of test items have been presented in diagrammatic form in Figures 2.10 and 2.11.

Although data were gathered at different time points for the SACMEQ I (1995-1997) and SACMEQ II (2000-2002) projects, Figures 2.10 and 2.11 suggest that it is possible to think of the reading and mathematics tests used in the projects as two "artificial" or "composite" tests of 148 different reading items and 91 mathematics items, respectively. This conceptualisation of the tests implies that the three sets of reading test respondents and the two sets of mathematics test respondents can each represent a single group of respondents for the purposes of undertaking "concurrent" scaling of the tests using the Rasch Model.

For the 148-item "composite" reading test described in Figure 2.10 there were 36 items that came only from the SACMEQ I pupil reading test (part g of the diagram), 52 test items that came only from the SACMEQ II pupil reading test (part a of the diagram), and 26 items that came only from the SACMEQ II teacher reading test (part e of the diagram). An additional 34 items were located in more than one test (parts b, c, d, f of the diagram), with 9 of these items being located in all three tests (part c of the diagram), and 3 sets of items (parts b, d, and f of the diagram) associated with pairs of tests. For the 91-item "composite" mathematics test described in Figure 2.11 there were 50 items that came only from the SACMEQ II pupil mathematics test (part a of the diagram), and 28 items that came only from the SACMEQ II teacher mathematics test (part c of the diagram). An additional 13 items were located in both tests (part b of the diagram).

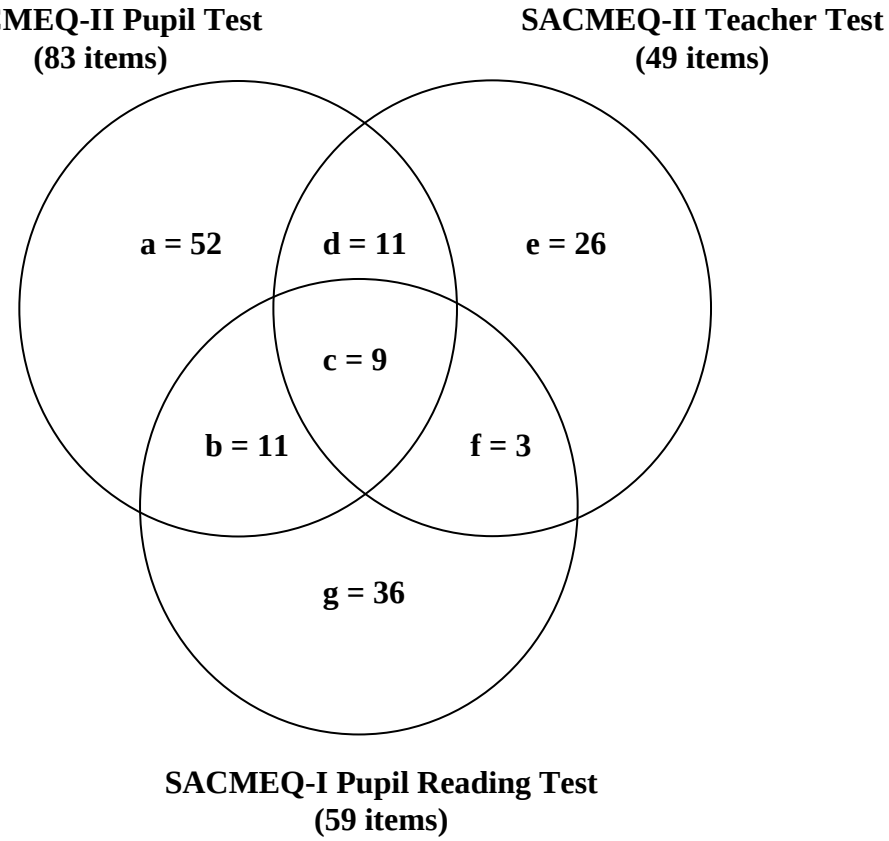

Figure 2.10: The 148 Reading Items for Three Groups of Respondents Taking "Different but Overlapped" SACMEQ Reading Tests

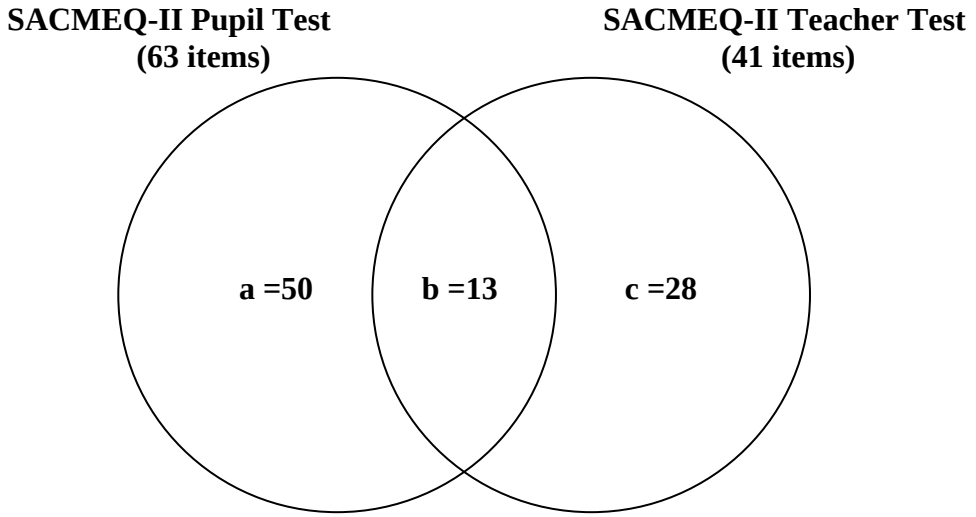

Figure 2.11: The 91 Mathematics Items for Two Groups of Respondents Taking "Different but Overlapped" SACMEQ Maths Tests

SACMEQ-II Pupils	a 52 items	b 11 items	c 9 items	d 11 items	e --	f --	g --
SACMEQ-II Teachers	a --	b --	c 9 items	d 11 items	e 26 items	f 3 items	g --
SACMEQ-I Pupils	a --	b 11 items	c 9 items	d --	e --	f 3 items	g 36 items

Figure 2.12: The Data Matrix Developed for Scaling the 148 Reading Items Contained within the SACMEQ Reading Tests

SACMEQ-II Pupils	a 50 items	b 13 items	c --
SACMEQ-II Teachers	a --	b 13 items	c 28 items

Figure 2.13: The Data Matrix Developed for Scaling the 91 Items Contained within the SACMEQ Mathematics Tests

The data matrices used in the Rasch analyses have been presented in diagrammatic form in Figures 2.12 and 2.13. The blank areas of the diagram (denoted by "—") refer to items not given to respondents. For example, the SACMEQ II pupils received blocks of reading test items that have been labelled in Figure 2.12 as a, b, c, and d. These pupils did not receive the blocks of reading test items in Figure 2.12 that have been denoted by "—" and labelled e, f, and g.

The data matrix used in the Rasch analyses to scale the 148 reading items was constructed by combining data from the three groups of respondents – with valid response codes in the matrix columns referring to each group's own test, and "missing data" codes in the matrix columns referring to items only found in either or both of the other two tests.

Similarly, the data matrix used to scale the 91 mathematics items was constructed by combining data from two groups of respondents – with valid response codes in the matrix columns referring to each group's own test, and "missing data" codes in the matrix columns referring to items only found in the other test. While the computer software treated these columns as "missing data", in fact these items were not actually given to these respondents.

The two data matrices were analysed using computer software that applied the Rasch Model of measurement (Andrich and Luo, 2000). The first step was to calibrate the test items by calculating the Rasch difficulty values for each item within the 148-item reading test and the 91-item mathematics test. This step was conducted by using an input data file constructed from a simple random sample of pupils and teachers from each school system. The results of the calibration were then used to calculate reading and mathematics scores for all pupils.

Construction of Test Items for the Teacher Tests

The main challenge in the construction of the reading and mathematics tests for teachers was to "fine-tune" the difficulty range of test items so that it would suit the higher levels of competence that were expected of teachers. At the same time it was necessary to ensure that there was sufficient "item overlap" with the pupil tests to permit the performance of teachers and pupils to be measured on the same scale.

In the reading test for teachers, several passages were selected because of the more subtle nature of the messages that they conveyed, and the less-visible underlying assumptions of the writers. For example, one passage on the topic of "smoking" required the teachers to identify the unstated values and beliefs of the writer. Another passage on the topic of "effective thinking" required the teachers to identify assumptions made by the writer about the readers and their knowledge of the topic. These kinds of skills were far beyond the competencies that had been identified from the analyses of Grade 6 curricula.

The "extra" reading and mathematics items for teachers were expected to assess the higher competence levels of teachers – but not to be so difficult that the teachers would be daunted by the challenge. In addition, the selection of easier test items that "overlapped" with the pupil tests had to be made with extreme care because the teachers may have felt insulted if these items were ridiculously easy or if they were concerned with issues that would only interest young children.

In the teacher reading test the extended levels of competence mainly focussed on expository texts – rather than on documents or narratives. It was felt that the use of narratives and documents at this level would have required very complex and long texts that would have generally extended the time required to complete the test.

In the teacher mathematics test the extended levels of competence mainly emphasized problem solving strategies that required the extraction of information from verbal, graphic, or tabular presentations. For these items, the teachers were expected follow three steps: to identify the nature of the problem, to transform the problem into mathematical language, and to solve the problem. In some cases this required the rearrangement of information, and in others it meant translating the problem into one or more equations and then solving the equations.

Using a "Skills Audit" to Identify "Derived" Competence Levels

The SACMEQ tests had been prepared according to systematically-generated test blueprints that described "proposed" levels of competence in reading and mathematics. The results of the Rasch analyses provided a means of assessing whether the levels proposed in the test blueprints in Figures 2.6 and 2.9 were congruent with a detailed examination of the actual test items located at different difficulty levels along the dimensions that had been generated. The descriptions that were obtained after the NRCs had conducted the skills audit were called "derived" levels of competence. The skills audit analyses focussed on the matter of whether the NRCs had actually been able to write test items that were aligned along the five increasing skill levels proposed in the test blueprints.

To address this issue the NRCs examined the 148 items in the "hypothetical" reading test in Figure 2.10 and the set of 91 items in the "hypothetical" mathematics test in Figure 2.11. The two sets of items were first arranged in order of difficulty, and then examined item-by-item in order to describe the specific skills required to provide correct responses. When items had been linked to specific skills they were placed into groups of test items such that the items in each group had similar difficulty values and shared a common "theme" with respect to the underpinning competencies required to provide correct responses.

The three tasks of defining specific skills for each test item, identifying groups of items with similar difficulties, and then naming the "theme" (or competency level) linked to each group were extremely difficult because it required the NRCs to first reach agreement on how the respondents arrived at correct solutions, and to then name the competency required. This required the NRCs to use their practical knowledge of the ways in which pupils solve problems, and then to portray this with a meaningful description of the thought processes that had been applied. The next step was to compare the "proposed" levels of competence to the "derived" levels of competence in order to check the accuracy of the item writers' skills and the validity of the test.

Reading and Mathematics Competencies Generated from the Skills Audit

The skills audit for the reading and mathematics tests resulted in the identification of eight levels of competence for each test. This was more than had been proposed in the test blueprints.

For both tests there was a strong correspondence between the descriptions of the five blueprint levels and most of the derived levels arising from the skills audit – which suggested that the three "extra" levels were defining more detail on the same reading and mathematics scales. That is, the overall dimensions remained substantially the same, but the skills audit meant that the empirically-generated (or "derived") dimensions of reading and mathematics were, as expected, somewhat more detailed than the subjectively described (or "proposed") dimensions used to stimulate test and item development.

Level 1: Pre Reading (Linked with Level 1 in the Test Blueprint)
(a) Skills: Matches words and pictures involving concrete concepts and everyday objects. Follows short simple written instructions.
(b) Example Test Items

- locate familiar words in a short (one line) text
- match words to pictures
- follow short and familiar instructions

Level 2: Emergent Reading (Linked with Level 2 in the Test Blueprint)
(a) Skills: 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 on.
(b) Example Test Items

- read familiar words and identify some new words
- use simple and familiar prepositions and verbs to interpret new words
- match words and very simple phrases

Level 3: Basic Reading (Linked with Level 3 in the Test Blueprint)
(a) Skills: Interprets meaning (by matching words and phrases, completing a sentence, or matching adjacent words) in a short and simple text by reading on or reading back.
(b) Example Test Items

- 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

Level 4: Reading for Meaning (Linked with Level 4 in the Test Blueprint)
(a) Skills: Reads on or reads back in order to link and interpret information located in various parts of the text.
(b) Example Test Items

- interpret sentence and paragraph level texts
- match phrases across sentences
- read forwards and backwards in order to locate information in longer texts

Figure 2.14: Levels of Reading Competency Generated from Skills Audit

Level 5: Interpretive Reading (Linked with Level 5 in the Test Blueprint)
(a) Skills: Reads on and reads back 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.
(b) Example Test Items

- 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

Level 6: Inferential Reading (Linked with Level 5 in the Test Blueprint)
(a) Skills: Reads on and reads back through longer texts (narrative, document or expository) in order to combine information from various parts of the text so as to infer the writer's purpose.
(b) Example Test Items

- **i**nterpret, and make inferences from, different types of texts by reading backwards and forwards to confirm links between widely separated information pieces
- extract information from a non-traditional (left to right) document
- make judgments about an author's intentions or purpose beyond the text content

Level 7: Analytical Reading (Linked with Level 5 in the Test Blueprint)
(a) Skills: Locates information in longer texts (narrative, document or expository) by reading on and reading back 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).
(b) Example Test Items

- combine several pieces of information from a range of locations in complex and lexically dense text or documents
- analyse detailed text or extended documents for an underlying message
- identify meaning from different styles of writing

Level 8: Critical Reading (A New Level Generated from the Skills Audit)
(a) Skills: Locates information in a longer texts (narrative, document or expository) by reading on and reading back 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).
(b) Example Test Items

- use text structure and organisation 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

Figure 2.14 (Ctd.): Levels of Reading Competency Generated from Skills Audit

Level 1: Pre Numeracy (Linked with Level 1 in the Test Blueprint)
(a) Skills: Applies single step addition or subtraction operations. Recognizes simple shapes. Matches numbers and pictures. Counts in whole numbers.
(b) Example Test Items

- count illustrated objects
- recognise basic numbers and shapes
- carry out simple single operations of addition and subtraction

Level 2: Emergent Numeracy (Linked with Level 1 in the Test Blueprint)
(a) Skills: 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.
(b) Example Test Items

- link simple verbal, graphic, and number forms with single arithmetic operations on whole numbers up to four digits
- recognise common shapes or figures in two dimensions
- estimate accurately lengths of simple shapes

Level 3: Basic Numeracy (Linked with Level 2 in the Test Blueprint)
(a) Skills: 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.
(b) Example Test Items

- recognise three-dimensional shapes and number units
- use a single arithmetic operation in two or more steps
- convert in single step units using division

Level 4: Beginning Numeracy (Linked with Level 3 in the Test Blueprint)
(a) Skills: 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.
(b) Example Test Items

- convert units in two steps and count tabulated data
- analyse a visual prompt and interpret triangular shapes
- translate verbal to arithmetic form using two operations on fractions

Figure 2.15: Levels of Mathematics Competency Generated from Skills Audit

Level 5: Competent Numeracy (Linked with Level 3 in the Test Blueprint)
(a) Skills: 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 measurement and/or whole and mixed numbers. Converts basic measurement units from one level of measurement to another (for example, metres to centimetres).
(b) Example Test Items

- convert 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

Level 6: Mathematically Skilled (Linked with Level 4 in the Test Blueprint)
(a) Skills: 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).
(b) Example Test Items

- 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 graphic prompt (involving an analysis of steps) to identify the correct sequence of calculations
- convert, and operate on, units of measurement (time, distance, and weight)

Level 7: Concrete Problem Solving (Linked with Level 5 in the Test Blueprint)
(a) Skills: Extracts and converts (for example, with respect to measurement units) information from tables, charts, visual and symbolic presentations in order to identify, and then solves multi-step problems.
(b) Example Test Items

- 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 fractions to decimals

Level 8: Abstract Problem Solving (A New Level Generated from the Skills Audit)
(a) Skills: Identifies the nature of an unstated mathematical problem embedded within verbal or graphic information, and then translate this into symbolic, algebraic, or equation form in order to solve the problem.
(b) Example Test Items

- 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

Figure 2.15 (Ctd.): Levels of Mathematics Competency Generated from Skills Audit

The results of the skills audit have been presented in Figures 2.14 and 2.15. The NRCs decided to add a name to each of the levels – in order to summarize the competencies associated with each group. The first three competency levels in reading and mathematics employed the same prefixes ("Pre", "Emergent", and "Basic") in order to reflect the mostly mechanical nature of the most elementary competencies. From the fourth level upwards the prefixes of the summary names were different and tended to reflect deeper levels of understanding of subject specific competencies.

The NRCs considered that the use of a skills audit to generate the eight levels presented in Figures 2.14 and 2.15 was important because the competencies 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. Such descriptions are of great assistance for the construction of textbooks, the design of teacher in-service training programmes, and the development of general classroom teaching strategies - because all of these activities require a sound knowledge of the skills already acquired and the higher order skills that should be aimed at in order to transfer to the next stage of learning.

New levels were identified and derived through the skills audit, but as can be seen in Figures 2.14 and 2.15, the match between the "proposed" and "derived" levels in the dimensions of reading and mathematics competency were strikingly similar. This indicated that the NRCs had been quite successful in designing tests according to specifications as set out in the original test blueprints. It is also provided clear evidence of the content and construct validity of the reading and mathematics tests.

Some examples of test items for each of the eight competency levels in the reading and mathematics tests have been presented in Appendix H and Appendix I, respectively.

The Score Ranges for the Competency Levels

The software used to generate the Rasch reading and mathematics scores automatically adjusted the scores to a scale with an arbitrary zero point and a standard deviation of one. This meant that many pupils were assigned negative scores. Most educationalists are not comfortable with score patterns of this kind. Therefore it was decided to undertake a linear transformation of the reading and mathematics scores that would result in the mean and standard deviation of pupil scores for the SACMEQ II tests being 500 and 100, respectively (for the pooled data with equal weight given to each country). As a result a score of 500 was equal to the average of all SACMEQ II country mean scores. The transformed scores have been referred to below as "500 Scores".

The Rasch analysis made it possible for the ability of the pupils to be matched to the difficulty of the test items – which allowed pupils and items to be mapped onto the same scale. This meant that the pupils could also be grouped in the same "ability" or "difficulty" range as the items that had similar difficulty values. In Tables 2.10 and 2.11 the ranges of the "500 Scores" that define the eight reading and mathematics competency levels, respectively, have been presented. The two tables also contain the percentages of pupils and teachers that were located at each competency level.

Table 2.10: Reading Competency Levels Cut-off Points and Frequency Distributions

Reading Competency	Rasch Score Range	500 Score Range	Percentage at Competency Level (SE)		
			Pupils		Teachers
			SACMEQ I	SACMEQ II	SACMEQ II
1 : Pre Reading	Lte -1.765	Lte 373	3.2	6.7	0.1
2 : Emergent Reading	Gt-1.765- -1.332	Gt 73- 414	7.1	14.9	0.2
3: Basic Reading	Gt-1.332- -0.881	Gt414- 457	22.2	18.4	0.1
4: Reading for Meaning	Gt-0.881- -0.334	Gt457- 509	28.7	20.2	1.4
5: Interpretive Reading	Gt-0.334- 0.232	Gt509- 563	19.1	16.8	1.0
6: Inferential Reading	Gt 0.232- 0.807	Gt563- 618	9.3	10.7	4.1
7: Analytical Reading	Gt 0.807- 1.692	Gt618- 703	7.2	8.4	28.1
8: Critical Reading	Gt 1.692	Gt703-	3.3	3.8	65.0

Table 2.11 : Mathematics Competency Levels Cut-off Points and Frequency Distributions

Mathematics Competency	Rasch Score Range	500 Score Range	Percentage at Competency Level (SE)	
			Pupils	Teachers
			SACMEQ II	SACMEQ II
1 : Pre Numeracy	Lte –2.199	Lte364	6.2	0.0
2 : Emergent Numeracy	Gt-2.199- -1.325	Gt364- 462	34.3	0.0
3: Basic Numeracy	Gt-1.325- -0.709	Gt462- 532	29.8	0.9
4: Beginning Numeracy	Gt-0.709- -0.213	Gt532- 587	14.6	2.0
5: Competent Numeracy	Gt-0.213- 0.293	Gt587- 644	7.5	6.0
6: Mathematically Skilled	Gt 0.293- 0.962	Gt644- 720	4.6	16.7
7: Concrete Problem Solving	Gt 0.962- 1.728	Gt720- 806	2.2	36.0
8: Abstract Problem Solving	Gt 1.728	Gt806	0.9	38.5

Conclusion

The aim of this chapter was to describe the research procedures that were applied for the execution of the SACMEQ II Project. The chapter was prepared in three parts that covered the fourteen main phases of the research, the sample design procedures, and the construction of the reading and mathematics tests for pupils and their teachers.

The first part of the chapter described how the SACMEQ II Project commenced with an innovative "pre-planning" phase that underpinned the whole research design. During this phase key decision-makers in Ministries of Education were consulted concerning their "General Policy Concerns" – which were then collated across countries, grouped into five themes, and used as a foundation for the design of the whole data collection and the research reporting procedures.

One of the important messages that emerged from this part of the chapter was that the speed at which a cross-national research project proceeds is strongly influenced by the speed with which the slowest country can complete all aspects of its data collection and data preparation.

The second part of the chapter on sampling included an evaluation of the sampling procedures. The evaluation showed that nine countries satisfied the sampling accuracy requirements that had been set down for the SACMEQ II Project – by achieving equivalent sample sizes for the pupil tests that were in excess of 400 pupils. A further three countries (Kenya, Lesotho, and Swaziland) almost reached this standard by achieving equivalent sample sizes in the range of 350 to 390.

Unfortunately, the accuracy of the sampling in two countries (South Africa and Uganda) fell far below the 400 target – with South Africa and Uganda achieving equivalent sample sizes of only 230 and 176, respectively. These results indicated that care should be exercised in interpreting the reading and mathematics achievement levels that were obtained for these two countries, and also that even more care should be taken when examining within-country regional differences.

The third part of the chapter provided a detailed description of how the SACMEQ II Project moved away from traditional approaches to the calculation of test scores (based on numbers of correct responses to test items) towards the use of Modern Item Response Theory to generate descriptions of "levels of increasing pupil competence". This approach to describing pupil reading and mathematics achievement offered a mechanism for describing the performance of pupils in a manner that was more meaningful within a teaching and learning context.

References

Andrich, D., and Luo, G. (2003). *Getting started: RUMM 2010.* Perth: The RUMM Laboratory.

Brickell, J.L. (1974). Nominated samples from public schools and statistical bias. *American Educational Research Journal*, 11(4), 333-341.

Chimombo, J., Dlamini, E., Kulpoo, D., Moyo, G., Murimba, S., Nassor, S. M., and Nkamba, M. (1994). *A project plan for the Southern Africa Consortium for Monitoring Educational Quality. (Vols. I and II).* Paris: International Institute for Educational Planning.

Deming, W.E. (1960). *Sample design in business research*. New York: Wiley.

Elley, W. (1992). *How in the world do students read?* The Hague: International Association for the Evaluation of Educational Achievement.

Finifter, B.M. (1972). The generation of confidence: Evaluating research findings by random subsample replication. In H.L. Costner (Ed.), *Sociological Methodology*. San Francisco: Jossey-Bass.

Frankel, M.R. (1971). *Inference from survey samples*. Ann Arbor, Michigan: Institute for Social Research.

Kish, L. (1965). *Survey sampling*. New York: Wiley.

Kish L. (1978). On the future of survey sampling. In N.K. Namboordi (Ed.), Survey sampling and measurement. New York: Academic Press.

Kulpoo, D. (1998). *The quality of education: Some policy suggestions based on a survey of schools in Mauritius.* Paris: International Institute for Educational Planning.

Machingaidze, T., Pfukani, P., and Shumba, S. (1998). *The quality of education: Some policy suggestions based on a survey of schools in Zimbabwe*. Paris: International Institute for Educational Planning.

McCarthy, P.J. (1966). *Replication: An approach to the analysis of data from complex surveys*. Washington: United States National Center for Health Statistics.

Milner, G., Chimombo, J., Banda, T., and Mchikoma, C. (2001). *The quality of education: Some policy suggestions based on a survey of schools in Malawi.* Paris: International Institute for Educational Planning.

Moyo, G., Murimba, S., Nassor, S. M., Dlamini, E., Nkamba, M., and Chimombo, J. (1993). *SADC proposal for monitoring progress toward attaining the goals of the EFA Jomtien Conference concerning the quality of education.* Harare: Ministry of Education and Culture.

Mullis, I. V. S., Martin, M. O., Smiith, T. A., Garden, R. A., Gregory, K. D., Gonzalez, E. J., Chrostowski, S. J., and O'Connor, K. M. (2001). *TIMSS assessment frameworks and specifications 2003*. Chestnut Hill, MA: Boston College.

Nassor, S and Ali Mohammed, K. (1998). *The quality of education: Some policy suggestions based on a survey of schools in Zanzibar.* Paris: International Institute for Educational Planning.

Nkamba, M. and Kanyika, J. (1998). *The quality of education: Some policy suggestions based on a survey of schools in Zambia*. Paris: International Institute for Educational Planning.

Nzomo, J., Kariuki, M., and Guantai, L. (2001). *The quality of education: Some policy suggestions based on a survey of schools in Kenya*. Paris: International Institute for Educational Planning.

Organization for Economic Cooperation and Development (OECD) (2000). *Measuring student knowledge and skills: The PISA 2000 assessment of reading, mathematical, and scientific literacy*. Paris: OECD.

Ross, K.N. (1976). *Searching for uncertainty: Sampling errors in educational survey research*. Hawthorn, Victoria: Australian Council for Educational Research.

Ross, K.N. (1978). Sample design for educational survey research. *Evaluation in Education*, 2, 105-195.

Ross, K.N. (1985). Sampling . In T. Husen & T.N. Postlethwaite (Eds.), *The International Encyclopedia of Education* (pp. 4370-4381). New York: Pergamon.

Ross, K.N. (1987). Sample design. *International Journal of Educational Research*, 11(1), pp. 57-75.

Ross, K.N. (1991). *Sampling manual for the IEA International Study of Reading Literacy*. Hamburg: International Association for the Evaluation of Educational Achievement.

Ross, K. N. (1995). From educational research to educational policy: An example from Zimbabwe. *International Journal of Educational Research* 23(4), pp. 301-403.

Ross, K. N., Saito, M., Dolata, S., and Ikeda, M. (2004). *The SACMEQ Data Archive (Version 1.2).* Paris: International Institute for Educational Planning.

Sylla, K., Saito, M., Ross, K. (2003). *SAMDEM (Sample Design Manager Software).* Paris: International Institute for Educational Planning.

Tukey, J.W. (1958). Bias and confidence in not-quite large samples (Abstract). *Annals of Mathematical Statistics*, 29, 614.

Voigts, F. (1998). *The quality of education: Some policy suggestions based on a survey of schools in Namibia*. International Institute for Educational Planning (UNESCO): Paris.