

Module

T. Neville Postlethwaite
Institute of Comparative Education
University of Hamburg

Educational research: some basic concepts and terminology



Quantitative research methods in educational planning

These modules were prepared by IIEP staff and consultants to be used in training workshops presented for the National Research Coordinators who are responsible for the educational policy research programme conducted by the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ).

The publication is available from the following two Internet Websites:

<http://www.sacmeq.org> and <http://www.unesco.org/iiep>.



International Institute for Educational Planning/UNESCO

7-9 rue Eugène-Delacroix, 75116 Paris, France

Tel: (33 1) 45 03 77 00

Fax: (33 1) 40 72 83 66

e-mail: information@iiep.unesco.org

IIEP web site: <http://www.unesco.org/iiep>



September 2005 © UNESCO

The designations employed and the presentation of material throughout the publication do not imply the expression of any opinion whatsoever on the part of UNESCO concerning the legal status of any country, territory, city or area or of its authorities, or concerning its frontiers or boundaries.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means: electronic, magnetic tape, mechanical, photocopying, recording or otherwise, without permission in writing from UNESCO (International Institute for Educational Planning).

Graphic design: Sabine Lebeau

Typesetting: Sabine Lebeau

Printed in IIEP's printshop



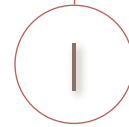
Content

1. Introduction	1
2. Types of educational research	2
3. Three types of research questions in educational planning	6
Descriptive questions	6
Correlational questions	8
Causal questions	9
4. Identifying research issues for educational planning	11
5. Sequential stages in the research process	16
General and specific research questions	16
Literature review	16
Research design	17
Instrumentation	18
Pilot testing	20
Data collection	22
Data analysis	24
Research report	28

6. Conclusion	29
Appendix A	
Terminology used in educational research	30
Formative and summative evaluation	31
Assessment, evaluation, and research	33
Measurement	33
Surveys and experiments	34
Tests	36
1. Test items	36
2. Sub-scores/Domain scores	37
Variable	37
1. Types of variables	38
Validity and reliability	39
1. Validity	39
2. Reliability	41
Indicator	42
Attitude scales	43
Appendix B	
Further reading suggestions	46
Introductory texts	46
Examples of educational research studies that aimed to have an impact on educational planning	47
Encyclopedias and handbooks	48
Journals	48
Appendix C	
Exercises	49

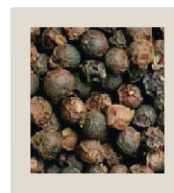


Introduction



Research is the orderly investigation of a subject matter for the purpose of adding to knowledge. Research can mean ‘re-search’ implying that the subject matter is already known but, for one reason or another, needs to be studied again. Alternatively, the expression can be used without a hyphen and in this case it typically means investigating a new problem or phenomenon.

Within the realm of educational planning, many things are always changing: the structure of the education system, curriculum and textbooks, modes of teaching, methods of teacher training, the amount and type of provisions to schools such as science laboratories, textbooks, furniture, classroom supplies, and so on. These changes may lead to an improvement, or a worsening, in the quality of an educational system. Sometimes they may result in no impact upon quality – in which case major government expenditures on such changes have been wasted. The educational planner working within this kind of environment must be able to undertake assessments of the effects of major changes and then provide policy advice that will consolidate and extend the most productive courses of action, and also intercept and terminate existing practices that are shown to be damaging and wasteful.



2

Types of educational research

There are many types of educational research studies and there are also a number of ways in which they may be classified. Studies may be classified according to topic whereby the particular phenomena being investigated are used to group the studies. Some examples of educational research topics are: teaching methods, school administration, classroom environment, school finance, etc. Studies may also be classified according to whether they are exploratory or confirmatory.

An exploratory study is undertaken in situations where there is a lack of theoretical understanding about the phenomena being investigated so that key variables, their relationships, and their (potential) causal linkages, are the subject of conjecture. In contrast a confirmatory study is employed when the researcher has generated a theoretical model (based on theory, previous research findings, or detailed observation) that needs to be tested through the gathering and analysis of field data.

A more widely applied way of classifying educational research studies is to define the various types of research according to the kinds of information that they provide. Accordingly, educational research studies may be classified as follows:

1. **Historical research** generates descriptions, and sometimes attempted explanations, of conditions, situations, and events that have occurred in the past. For example, a study that documents the evolution of teacher training programs since the turn of the century, with the aim of explaining the historical origins of the content and processes of current programs.

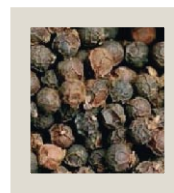
2. **Descriptive research** provides information about conditions, situations, and events that occur in the present. For example, a survey of the physical condition of school buildings in order to establish a descriptive profile of the facilities that exist in a typical school.
3. **Correlational research** involves the search for relationships between variables through the use of various measures of statistical association. For example, an investigation of the relationship between teachers' satisfaction with their job and various factors describing the provision and quality of teacher housing, salaries, leave entitlements, and the availability of classroom supplies.
4. **Causal research** aims to suggest causal linkages between variables by observing existing phenomena and then searching back through available data in order to try to identify plausible causal relationships. For example, a study of factors related to student 'drop out' from secondary school using data obtained from school records over the past decade.
5. **Experimental research** is used in settings where variables defining one or more 'causes' can be manipulated in a systematic fashion in order to discern 'effects' on other variables. For example, an investigation of the effectiveness of two new textbooks using random assignment of teachers and students to three groups – two groups for each of the new textbooks, and one group as a 'control' group to use the existing textbook.
6. **Case study research** generally refers to two distinct research approaches. The first consists of an in-depth study of a particular student, classroom, or school with the aim of producing a nuanced description of the pervading cultural setting that affects education, and an account of the interactions

that take place between students and other relevant persons. For example, an in-depth exploration of the patterns of friendship between students in a single class. The second approach to Case Study Research involves the application of quantitative research methods to non-probability samples – which provide results that are not necessarily designed to be generalizable to wider populations. For example, a survey of the reading achievements of the students in one rural region of a particular country.

- 7. Ethnographic research** usually consists of a description of events that occur within the life of a group – with particular reference to the interaction of individuals in the context of the sociocultural norms, rituals, and beliefs shared by the group. The researcher generally participates in some part of the normal life of the group and uses what he or she learns from this participation to understand the interactions between group members.

For example, a detailed account of the daily tasks and interactions encountered by a school principal using observations gathered by a researcher who is placed in the position of ‘Principal’s Assistant’ in order to become fully involved in the daily life of the school.
- 8. Research and development research** differs from the above types of research in that, rather than bringing new information to light, it focuses on the interaction between research and the production and evaluation of a new product. This type of research can be ‘formative’ (by collecting evaluative information about the product while it is being developed with the aim of using such information to modify and improve the development process). For example, an investigation of teachers’ reactions to the various drafts and redrafts of a new mathematics teaching kit, with the information gathered at each stage being used to improve each stage of the drafting

process. Alternatively, it can be 'summative' (by evaluating the worth of the final product, especially in comparison to some other competing product). For example, a comparison of the mathematics achievement of students exposed to a new mathematics teaching kit in comparison with students exposed to the established mathematics curriculum.



3

Three types of research questions in educational planning

In research on issues concerned with educational planning, the main educational research questions can be subsumed under three categories: descriptive, correlational, and causal.

Descriptive questions

In the field of educational planning, the research carried out on descriptive questions is often focused on comparing the existing conditions of schooling with: (i) legislated benchmark standards, (ii) conditions operating in several other school systems, or (iii) conditions operating in several sectors of a single school system.

Some examples are:

- What is the physical state of school buildings in the country? Do some districts or regions have better or worse school buildings than others? (Behind these two questions are the implications that the Ministry of Education wishes to ensure that all schools have a minimum standard of school building while at the same time ensuring that there are not large differences among schools with respect to the state of their buildings.)

- Do the supplies and equipment in classrooms in the schools match the legislated standards set by the Ministry? (The supplies and equipment might be textbooks, exercise books, pencils, erasers, seats and desks. The Ministry may have norms that each student in a particular grade must have one mother tongue textbook, one math textbook, one science textbook and one social studies textbook, four exercise books, three pencils and one eraser in a year, and that each student must have one seat and one writing place. The research required in this situation then consists of undertaking a count of the supplies and equipment in all schools, or in a scientific sample of schools that can be used to estimate the situation in all schools, and then matching every school or classroom against the Ministry's norms. The main aim of this kind of research study would be to examine whether there are particular districts or regions which are under-supplied or over-supplied.)
- Where systems of education have teacher housing, how adequate is the housing? (Some Ministries take the view that if the personal needs of teachers such as teacher housing do not satisfy the teachers then they will not be committed teachers. So, again, the research will consist of fact-finding about teacher housing and about teachers' satisfaction with these housing conditions.)
- What is the level of achievement in the core subject areas at a particular level of schooling? Does such achievement accord with the Ministry's view of what should have been learned by all students or particular sub-groups of students? A further question is often raised about student achievement – is it better, worse, or the same as last year's achievement for the particular grade group? And, again, are there differences between regions, or urban and rural children, and so on? (By the early 1990s there was a thrust from some systems of education – especially that of the United States of America – to have comparisons

of achievement (at the same age level) of different national systems of education for countries at a similar level of economic development.)

Correlational questions

Behind these kinds of questions, there is often an assumption that if an association is found between variables then it provides evidence of causation. However, care must be exercised when moving between the notions of association and causation. For example, an 'association' may be discovered between the incidence of classroom libraries and average class reading scores. However, the real 'cause' of higher reading scores may be that students from high socio-economic backgrounds, while they tend to be in classes with classroom libraries, read better than other students because their home environments (in terms of physical, emotional, and intellectual resources) facilitate the acquisition of reading skills.

Some examples are:

- Do students in poorer school buildings have lower achievement scores than those in better buildings?
- Do students in better equipped classrooms have better achievement scores than those in less well-equipped classrooms?
- Do students in schools where the teachers have better teacher housing have higher achievement than students in schools where teachers have poorer teacher housing?
- Do male students do better than female students in the scientific/technical subject areas?

Causal questions

Causal questions are usually the most important to educational planners. For example, in some schools it is considered normal for children to have a desk at which to sit. In other schools the children sit on the ground and write on their laps. It is important to know if schools (with a particular socio-economic background of children) with a shortage of desks and seats achieve less well than schools (with a similar socio-economic background of children) with an adequate supply of desks and chairs. Or, to put the question in a different way, is it the desks and chairs, or something else, which really cause the better achievement? It may be a better supply of books or better qualified teachers or, or, or.... It is, therefore, important to disentangle the relative influence of each of the many input and process factors in schools on achievement.

As will be seen from another module in this series on 'Research Design' both survey and experimental designs can be used to assess the relative influence of many factors on educational achievement. It is unusual in education to find only one factor influencing student educational achievement. It is rather the case that several, or even many, factors from outside and inside the school influence how well or poorly students achieve in school.

Thus, causal questions take one of two forms. Some examples are:

- **All other factors being equal** do students with Textbook A achieve better than students with Textbook B?
- **What is the relative effect** on school achievement of the following factors:
 - the socio-economic level of students in the school;
 - the general parental help given to the children with their homework;

- peer group pressure;
- the condition of the school buildings;
- the supplies and equipment in the classroom;
- the curriculum;
- the quality of teaching, etc.

Given that many factors will affect student achievement, then it is those factors that have a large influence which must be of concern to educational planners. Once the important factors have been identified then the planners can decide on the action they wish to take. Let us assume that it is found that the existence of sufficient desks and chairs does have a major influence on achievement, then it is up to the Ministry of Education to ensure that sufficient desks and chairs are made available. But, once this has been accomplished it will be necessary to undertake further research to discover what now is the most important factor influencing achievement.



Identifying research issues for educational planning

The reason why all educational planners should be prepared to undertake research is that it is important to be sure of the facts before making suggestions for changes in educational policies and practices. The maxim must always be “when in doubt, find out”.

The questions listed above are only examples and are very general in nature. It is up to each Ministry and educational planning office to pose its own questions in order to remove doubt. The formulation of research questions is, however, not an easy matter.

A wise educational researcher, educational planner, and ex-Minister of Education, Dr. C. E. Beeby from New Zealand, once wrote:

“I have suggested areas of research that seem to me to be of special importance. But not once have I asked a specific question to which I want an answer... I know enough about research to be aware that the formulation of the proper question may take as much skill and professional insight as the finding of an answer to it, and it may be a skill in which the administrator is not adept. So, the research workers must be involved in the asking of the questions, and must be prepared, in turn, to play a necessary, but secondary, part in devising the policies that may follow from the research, where their expertise is limited”. (ACER Radford Memorial Lecture, Melbourne, Australia 1987).

The examples of questions given earlier are general questions. There is a lot of work involved in turning such general questions into research questions. How this is done is dealt with in another module in this series that is entitled: "Specification of Research Aims".

Before proceeding to a discussion of the sequential steps in the research process, some examples are given of issues on which research was undertaken by educational planners in the 1980s.

1. Indonesia national evaluation of grade nine achievement levels

(Jiyono and Suryadi, 1992)

- A description of the achievement of the Grade nine students in Mathematics, Science, Social Studies, Moral Education, Mother Tongue, and English.
- The identification of the relative importance of the in – and out-of-school factors associated with achievement in each subject matter.
- A comparison of the achievement in 1981 with the achievement of the students in the same grade in 1976 (Jiyono and Suryadi, 1982).

2. Indonesia non formal learning behaviour

(Mappa, 1982)

- To compare the 'booklet only' learning groups with 'booklet plus radio' learning groups in a distance education project on the achievement of learners in: literacy, general knowledge, and numeracy.
- To measure the attitudes of learners toward innovation, health and nutrition, marriage, family planning and agriculture.

3. Thailand adult education project

(Thongchua et al, 1982)

- To measure the skills and knowledge gained by participants in typing and sewing courses of different duration.
- To identify variables having an effect on the achievement of participants at the end of the course.
- To investigate whether the graduates took up employment in typing/sewing within six months of the end of the course.
- To assess how participants utilized the skills six months after completing the course.

4. Thailand community secondary schools project

(Sawadisevee et al, 1982)

- To estimate the projected target of the number of teachers and students involved in the Community Schools Project.
- To evaluate whether the vocational teaching and learning program had responded to student needs and community needs.
- To assess whether the schools had been able to provide community development activities and services.

5. Malaysia remedial reading project

(Norisah bt Atan et al, 1982)

- To assess the impact of parent-teacher involvement on student's reading performance.
- To establish the reading objectives attained by the students.
- To describe the levels of parent participation in the reading activities of their children.
- To describe the levels of teacher participation in the reading activities of students.

6. Malaysia moral education project

(Asmah bt Mohd Taib et al, 1982)

- To compare attitudinal outcomes according to three methods of moral education.
- To assess differential attitudinal outcomes of urban and rural students.
- To measure to what extent teachers are able to apply the three methods of moral education and to use the materials supplied by the Ministry.

These are only a few selected examples of research conducted by Ministries of Education in the 1980s . To take a final example from the 1990s, a sixth grade survey in Zimbabwe had the following aims.

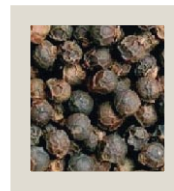
7. Indicators of the quality of education: a summary of a national study of primary schools in Zimbabwe

(Ross and Postlethwaite, 1992)

- What are the baseline data for the selected inputs to Zimbabwe primary schools?
- What percentage of schools in Zimbabwe fall below the norms for equipment and supplies?
- How equitably are these resource inputs distributed across primary schools in Zimbabwe?
- What is the level of achievement in the schools, and to what extent does achievement vary across the major administrative regions of Zimbabwe?
- What are the linkages between selected inputs to Zimbabwe schools and the learning outcomes of pupils? Which of these inputs can be identified as the most likely to have a beneficial input on pupil achievement through the Ministry of Education reallocation of or increase in input resources?

Indeed in the mid-1990s there were another seven southern African countries that undertook sixth grade surveys with either identical or very similar aims to the above.

Most studies were concerned with inputs to schools and the relationship of inputs to achievement outcomes. In research studies emanating from universities it often occurs that specific aspects of education and different forms of outcomes are researched. For example the self-esteem of students, the different kinds of motivation of students, grade-repeating, and different modes of teaching are favourite topics.



5

Sequential stages in the research process

General and specific research questions

The types of general research questions asked have been shown above. In order for the research to proceed in a focused and systematic manner, these questions must be refined to form more specific research questions that indicate exactly which target populations and which variables or factors should be included in the research study.

Literature review

The review of literature aims to describe the 'state of play' in the area selected for study. That is, it should describe the point reached by the discipline of which the particular research study will form a part. An effective literature review is not merely a summary of research studies and their findings. Rather, it represents a 'distillation' of the essential issues and inter-relationships associated with the knowledge, arguments, and themes that have been explored in the area. Such literature reviews describe what has been written about the area, how this material has been received by other scholars, the major research findings across studies, and the major debates in terms of substantive and methodological issues.

Research design

Given the specific research questions that have been posed, a decision must be taken on whether to adopt an experimental design for the study or a survey design. Further, if a survey design is to be used, a decision must be taken on whether to use a longitudinal design, in which data are collected on a sample at different points of time, or a cross-sectional design, in which data are collected at a single point of time.

Once the variables on which data are to be collected are known, the next questions are: Which data collection 'units' are to be employed? and Which techniques should be used to collect these data? That is, should the units be students, the teachers, the school principals, or the district education officers. And should data be collected by using observations, interviews, or questionnaires? Should data be collected from just a few hand-picked schools (case study), or a probability sample of schools and students (thus allowing inferences from the sample to the population), or a census in which all schools are included? For a case study, the sample is known as a 'sample of convenience' and only limited inferences can be made from such a sample.

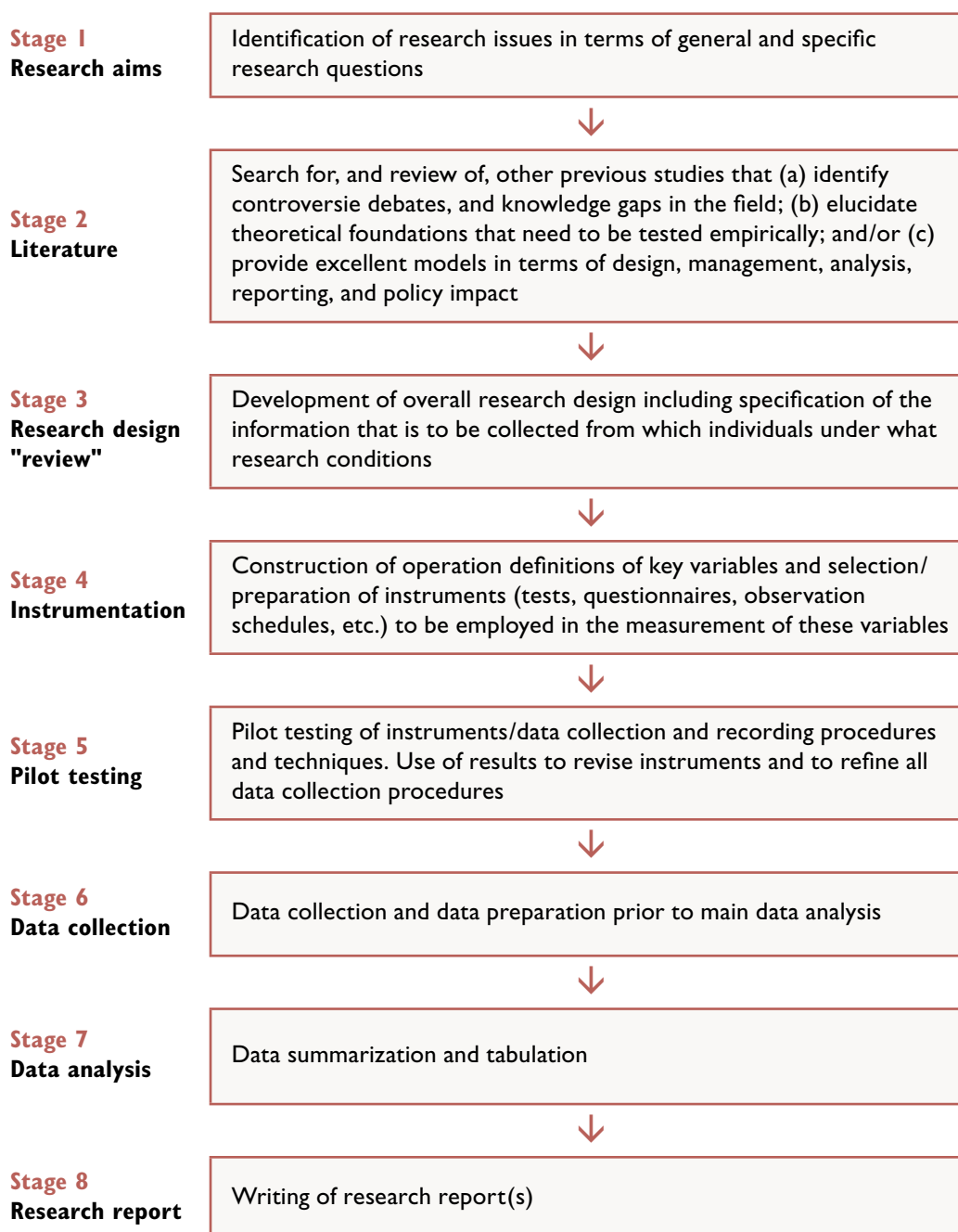
For research that aims to generalize its research findings, a more systematic approach to sample selection is required. Detailed information on the drawing of probability samples for both experimental and survey design is given in another module in this series entitled 'Sample Design'.

Instrumentation

Occasionally, data that are required to undertake a research study already exist in Ministry files, or in the data archives of research studies already undertaken, but this is rarely the case. Where data already exist, the analysis of them is known as “secondary data analysis”. But, usually, primary data have to be collected. From the specific research questions established in the first step of a research study it is possible to determine the indicators and variables required in the research, and also the general nature of questionnaire and/or test items, etc. that are required to form these. Decisions must then be taken on the medium by which data owe to be collected (questionnaires, tests, scales, observations, and/or interviews).

Once these decisions have been taken, the instrument construction can begin. This usually consists of the writing (or borrowing) of test items, attitude items, and questionnaire items. The items should be reviewed by experienced practitioners in order to ensure that they are unambiguous, and that they will elicit the required information. The broad issue of ‘Instrumentation’ (via both tests and questionnaires) has been taken up in more detail in several other modules in this series.

Figure 1 Stages in the research process



Pilot testing

At the pilot testing stage the instruments (tests, questionnaires, observation schedules, etc.) are administered to a sample of the kinds of individuals that will be required to respond in the final data collection. For example, school principals and/or teachers and/or students in a small number of schools in the target population. If the target population has been specified as, for example, Grade 5 in primary school, knowledge should exist in the Ministry, or in the inspectorate, about which schools are good, average, and poor schools in terms of educational achievement levels or in the general conditions of school buildings and facilities. A 'judgement sample' of five to eight schools can then be drawn in order to represent a range of achievement levels and school conditions. It is in these schools that the pilot testing should be undertaken.

The two main purposes of most pilot studies are:

- a. To assess whether a questionnaire has been designed in a manner that will elicit the required information from the respondents. This process allows weaknesses in the questionnaire to be detected so that they can be removed before the final form is prepared.

Typical weaknesses that are found in questionnaires include:

- Ambiguities in the phrasing of questions.
 - Excessive complexity in the language that has been used.
 - Inappropriate response categories for some questions.
 - Some questions are redundant.
- b. To assess whether test items can be understood by the students, that the items are pitched at the appropriate level of complexity (assessed by the 'Difficulty Index'), provide a stable measure

of student ability (assessed by the 'Reliability Index'), and lead to the construction of total test scores that are meaningful in terms of the student ability being examined (assessed by the 'Validity Index').

Typical weaknesses that are found in tests include:

- Some items have either no correct answer or more than one correct answer.
- Some distractors in multiple choice items are not functioning.
- Some items measure abilities different from the ability measured by other items (assessed by the 'Discrimination Index').
- Some items contain internal 'tricks' that result in high ability students performing worse than low ability students.

At the same time that the instruments are subjected to pilot testing, it is desirable to assess the effectiveness of the data collection procedures being used. These procedures include the steps to be followed for ensuring that the correct number of instruments with appropriate identification numbers on them for district, school, and student arrive at the schools punctually. Furthermore, there are procedures for selecting and then administering the questionnaires to the school principal, the teachers (all selected teachers) and the students (all students, one class of students, or a random sub-sample of students within a selected school). These activities address the following important questions: Are any problems evident in the procedures? How can the procedures be improved?

The same can be said about the procedures for entering data, cleaning data, and merging files. This work is usually undertaken by the planning office data processing unit, but again the results of the pilot testing experience can help to 'de-bug' the procedures. Once the instruments and procedures have been finalized, the main data collection can begin.

Data collection

When a probability sample of schools for the whole of the target population under consideration has been selected, and the instruments have been finalized, the next task is to arrange the logistics of the data collection. If a survey is being undertaken in a large country, this can require the mobilization of substantial resources and many people.

The management of this research stage will depend on the existing infrastructure for data collection. In many countries there are regional planning officers and within each region there are district education officers. These people are often used for collecting data. However, the problem of transport can loom large, especially in situations where there is a shortage of transportation and spare parts, and where office vehicles are booked weeks in advance. In such countries, two to three months of careful prior planning will be required.

In countries where schools are inundated with requests for data collection and where schools have the rights to refuse to participate in a data collection exercise, permission for the data collection must be sought well in advance. In some cases, the sample design must allow for replacement schools. This is a tricky matter as can be seen in the module on 'Sample Design'. The replacement of sample schools is often incorrectly carried out, which can then cast doubt on the results of the whole of the data collection.

In yet other countries, there is a practice of mailing the questionnaires to the schools and hoping that they will be returned. This often results in only a 40 to 60 percent response rate. This is disastrous and should be avoided. All efforts must be made to have the data collected by well-trained data collectors who visit the schools.

It is also important to stress that it is the educational planners who must control the selection of the students to be tested. From previous experience, it can be shown that if the district education officers are allowed to select the schools a bias will creep into the results. If the school head selects classes within the school there will be more bias, and if the teacher selects the students the bias will be greatest.

The instructions for the completion of tests and questionnaires must be clear. When the testing of students is involved, it is important to have a special "Manual for Test Administration". This manual explains how to arrange the testing room, and provides standard instructions that are given to the students about how to complete the test, questionnaire, or attitude statements, when to start and when to finish.

Finally, instructions must be clear on how the data collection instruments are to be returned from the field. When data collection instruments are returned to the National Planning Office, checks must be undertaken to ensure that all instruments have the correct identification numbers.

It is becoming popular in many countries to use optical scanning systems. Many examination centres use these systems, and many planning offices are beginning to use this mode of data collection. However, success in this approach does require detailed prior planning of the lay-out of pages, and that the appropriate type of paper be used. In most data collections conducted by educational planning offices, the data are entered onto computer diskettes directly from a keyboard. In many cases, a standard data entry program, such as a word-processing program or text-editing program, can be used. However, more specialized data entry programs are available which ensure more valid and verified forms of data entry. These specialized computer programs can be adapted to suit the needs of a variety of data collections. For example, for

open-ended questions, the expected minimum and maximum values are entered and if a value outside of this range is entered, the computer will give an error message. At this point the person entering the data must check against the data collection instrument. Usually he or she has made an error and can then correct it.

The net result of using such software is to optimize the accuracy of the final data set. This prevents many problems from occurring at the data analysis stage of a study.

After the data have been entered, cleaned, and merged – which often requires the student, teacher, and school files for a particular school to be joined into one record – the data analysis can begin.

Data analysis

If there are unequal probabilities of selection for members of the sample, or if there is a small amount of (random) non-response, then the calculation of sampling weights has to be undertaken. For teacher and school data there are choices which can be made about weighting. For example, if one is conducting a survey and each student in the target population had the same probability of entering the sample, then the school weights can either be designed to reflect the probability of selecting a school, or school weights can be made proportional to the weighted number of students in the sample in the school. In this latter case, the result for a school variable means the school value given is what the 'average student' experiences. This matter has been discussed in more detail in the module on 'Sample Design'.

a. Descriptive

Typically, the first step in the data analyses is to produce descriptive statistics separately for each variable. These statistics are often called univariates. Some variables are continuous – for example ‘size of school’ which can run from, say, 50 to 2,000. In this case the univariate statistics consist of a mean value for all schools, the standard deviation of the values, and a frequency distribution showing the number of schools of different sizes. Other variables are proportions or percentages. Such a variable could be the percentage of teachers with different types of teacher training.

These descriptive statistics describe the characteristics of the students, teachers, and schools in the sample. If a good probability sample has been drawn, then generalizations (within narrow limits) can be made about the target population.

Often comparisons between the Ministry norms and sample averages are made. For example, if the Ministry has stated that each student should have 1.25 square meters of space in the classroom, then this norm can be examined for each school by dividing the total number of square meters of classroom floor space in the school by the total enrolment of students in the school. This statistic may then be used to give direct feedback to the educational planners in charge of buildings about the extent to which their norms are being met.

A further use of univariates is to examine the means or percentages for particular groups in the sample. This may be urban vs. rural schools, or the schools in different regions in the country, or for boys’ schools vs. girls’ schools vs. co-educational schools. This procedure is known as cross-tabulation or break-downs. In other words, the data are cross-tabulated (or cross-classified) or broken down into segments. A simple example is shown below.

Table X Educational facilities provided for primary schools in Country X

Educational provision	All schools		Rural Schools		Urban schools	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Desks per classroom	X	X	X	X	X	X
Chairs per classroom	X	X	X	X	X	X
Floor space per student	X	X	X	X	X	X
Pens/Pencils per student	X	X	X	X	X	X

The first pair of columns presents the mean value and standard deviation for all schools in the sample for desks per classroom, chairs per classroom, floor space per student, etc. However, the total sample is 'broken down' in the second and third pairs of columns into rural and urban schools.

b. Correlational

In this case product moment correlations or cross tabulations can be calculated. There are statistical tests which can be applied to determine whether the association is more than would occur by chance. When the association between two variables is examined, this is known as 'bivariate' analysis.

c. Causal

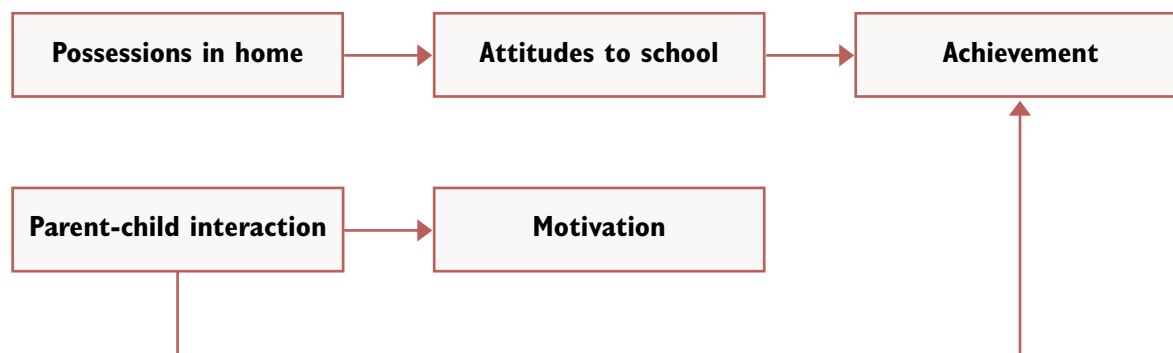
If the research design used is an experimental one, then tests can be applied to see if the performance of the experimental group (that is, the group subjected to the new treatment) is better than the control group.

There are statistical techniques for determining this. However, the use of this approach depends on the application of randomization in order to ensure that the two groups are 'equivalent' in all other respects.

If the research design is based on a survey, then it is possible to calculate the influence of one variable on another with other variables being "held statistically constant". Where calculations are made of the relationships among more than two variables at the same time, this is known as 'multivariate analysis'. It is possible to build causal models using the technique of path analysis. This technique requires the development of a causal model which describes not only the variables (or indicators) in the model, but also the pattern of causation among them. Analyses can be conducted to estimate the 'fit' of the data to the model. (See *Figure 2*)

An example of a path model could be:

Figure 2. Example of path model



In this example, it is posited, that the wealth of the home (represented by “Possessions in the Home”) influences ‘Attitudes to School’ which in turn influences both ‘Motivation’ and ‘Achievement’. ‘Parent-child Interaction’ influences ‘Motivation’ which also influences ‘Achievement’ but ‘Parent-child Interaction’ also has a direct effect on ‘Achievement’.

In survey type designs – even when causal models and path diagrams are used – association does not prove causality. However, if there is no association, then this casts doubt upon the original assumption of causality. If there is a strong association, and if a strong association is found repeatedly in several studies, then there is reasonable ground for assuming causality.

Research report

There are three major types of research reports. The first is the Technical Report written in great detail and showing all of the research details. This is typically read by other researchers. It is this report that provides evidence that the research was conducted soundly. This is usually the report which is written first.

The second report is for the senior policy makers in the Ministry of Education. It is in the form of an Executive Summary of about 5 or 6 pages. It reports the major findings succinctly and explains, in simple terms, the implications of the findings for future action and/or policy.

The third General Report is usually in the form of a 50 to 100 page booklet and is written for interested members of the public, teachers, and university people. This report presents the results in an easily understood and digestible form.



6

Conclusion

Each system of education has its political goals, its general educational goals, and its specific educational objectives. For example, some political goals stress equality of opportunity, others stress quality of education, and many stress both.

In every system of education changes are made by educational planners with the aim of improving the quality of education. These changes can include a revised curriculum, new methods of teacher training, increasing the amount of provisions to schools, changing the structure from a selective to a comprehensive system, reducing class size and many other changes. In some cases, innovations need to be tried out to identify their likely shortcomings, effects, and side effects before they are implemented. In other cases, student achievement over time in one or more subjects needs to be monitored, or where there are optional subjects the percentage of a grade or age group selecting such subjects needs to be known. Or, the attitudes and perceptions of students need to be assessed.

Educational planners are responsible for the planning of the various component parts of a system of education. Decisions must be taken on what to do to improve equality, quality, or both of each component part. As much information as possible is required to help decision-makers to operate successfully within the temporal, financial and political constraints within which they work.

This module has provided an introductory overview of the meaning of research, different types of research questions that are posed in many countries, and the sequential stages in the research process. This overview has also illustrated the main steps involved in this process and has provided examples of the kind of work that is required at each stage.

Appendix A

Terminology used in educational research

Educational research, like many other fields of research, has its own jargon. This appendix attempts to present a brief explanation of some of the terms commonly used in educational research studies undertaken by educational planners. Some have already been explained. The list provided below is not exhaustive, however, it aims to help educational planners to begin to read the research literature that impacts upon their work.

Operational research aims

Operational research aims are used to provide clear guidance concerning the conduct of educational research studies. These aims emerge from the general and specific research questions that are formulated for the study. In particular, operational research aims make it clear exactly what aspects of the educational environment should be measured.

For example, we may have a **general research question** that asks: "What are the linkages between inputs to schools and student achievement outcomes?" This question can be refined to form several **specific research questions** of which the following is one illustration: "What are the effects of teacher housing, school facilities, and classroom supplies on student achievement in reading?"

The **operational research** aims emerging from these general and specific research questions could include the following:

“To determine the likely effects of the following variables on students’ reading achievement in Grade 6: ‘Teacher Housing’ (as measured by the existence of water, existence of electricity, number of persons per square meter, extent of repairs required, and distance in travel time to school); ‘Quality of School Building’ (as measured by the number of complete classrooms, number of incomplete classrooms, number of open air classrooms, and amount of repair required); and ‘Classroom Equipment’ (as measured by the number of desks and chairs per student, existence of a blackboard, number of textbooks per student, and number of exercise books per student)”.

Formative and summative evaluation

Formative evaluation is conducted during the development or improvement of an educational program or product. It is usually conducted, and used, by personnel who are part of the team working on the program or product. However, it may sometimes be undertaken by an internal or external evaluator, or a combination or both.

Summative evaluation is conducted after the program or product has been completed. It is usually conducted by a researcher who was not part of the team working on the program or product, and is often undertaken for the benefit of an external audience or decision-maker. It, too, may be undertaken by internal or external evaluators or a combination of both.

Both types of evaluation require research rigor. An example can be taken from the field of curriculum. If new textbooks or teaching/learning units are to be developed, it is important that the units are written according to certain agreed specifications, that teachers

are trained to teach the new units, and that the units are tried out in a range of schools. Examples of the kinds of questions asked in formative evaluation would be: Do the specific objectives which have been developed cover the general objectives to be learned which are in the curriculum? Can the teachers cope with the new units? Are there any 'gaps' in the curriculum units which result in a poor coverage of some of the specific objectives? Can the layout of the curriculum units be changed so as to make the material more interesting for students?

In the curriculum development area, it is often specified that each objective should be mastered by 80 percent of students, and for a whole curriculum unit 80 percent of students should master 80 percent of the objectives. These performance levels may be used as benchmarks which can guide the identification of weaknesses in new curriculum units. If weaknesses are apparent then action can be taken to revise the procedures, the content, and/or the way in which the content is presented.

Summative evaluation, in this example, could focus on two areas of student performance. The first area would involve an evaluation of the achievement of students with respect to the material contained in the textbook as a whole or on all of the curriculum units together. This would involve the testing of all students on either all objectives, or on a selection of objectives from all of the curriculum units. Again, criteria of acceptability must be set in order to judge whether the totality of units are 'working'. The second kind of summative evaluation could be to compare the learning from 'Textbook A' with that from 'Textbook B'. In this case, it is assumed that the objectives to be learned from both textbooks are the same.

In a situation where funds for evaluation are limited, it makes more sense to place the emphasis on formative evaluation. If a new curriculum is not developed by using systematic formative

evaluation then the application of a summative evaluation becomes meaningless.

Assessment, evaluation, and research

Assessment usually refers to persons. It covers activities such as grading (formal and informal), examining, certifying, and so on. Most educational systems record student achievement in some way: With a number, letter code, or comment such as ‘good’, ‘satisfactory’, or ‘needs improvement’. One exception to this generalization is the use of the word ‘assessment’ as in National Assessment of Education Progress (NAEP) in the United States. Such assessments are based on the testing of probability samples within the nation. In this case, the objective is not to know how any individual student is achieving, but to discover more about the achievement of groups of students in different regions of the country.

Evaluation involves the general weighing of the value or worth of something in terms of the objective sought, or in comparison with other programs, curricula, or organizational schemes.

Research is the orderly and systematic investigation of a phenomenon for the purpose of adding to knowledge.

Measurement

Measurement is a process that assigns a numerical description to some attribute of an object, person, or event. Just as rulers and stop-watches can be used to measure, for example, height and speed, so can other quantities of educational interest be measured indirectly through the use of achievement tests, questionnaires and the like.

Surveys and experiments

Surveys involve the collection of information at one or several points in time from scientifically designed probability samples of students, teachers or schools. The information is usually collected by means of questionnaires and tests, but also sometimes by means of observation schedules and interviews. The information collected is usually from a probability sample selected from a tightly defined population, but can also be from a full count of schools in the form of census data. The data are generally for descriptive purposes, but may also be analyzed for relationships among variables. In some cases, causal path models are developed and tested.

Although surveys can never prove causality, it is assumed that if sufficient replications of a study or set of relationships are made, all of which show a particular relationship which is generally deemed to be causal, then it is reasonable to infer causality.

There are two main groups of surveys. The most common is a “cross-sectional survey”. This involves the collection of data at a specific point in time and is rather like taking a photograph on one day. The data collected can involve what the situation is now (for example, the number of books in the school library or students’ achievement in one or more subject areas) or retrospective information about what level of education a student’s parents had or prospective information about what type of further education a student wishes to pursue.

‘Longitudinal Surveys’ involve following a particular group of students or schools over a period of time. Some longitudinal studies follow persons from birth to death, others over a certain number of years, others over the period of one school year and others over a period of 3 or 6 months. If students leave school and are followed up for a year or two into their employment, this kind of longitudinal study is often known as a ‘Tracer Study’.

Longitudinal studies tend to cost more than cross-sectional surveys and, although in theory they are superior to cross-sectional surveys in terms of determining causal patterns, the costs and problems of knowing the students' names and addresses (in countries with strict data protection laws) often debar researchers and planners from undertaking this kind of work.

Experiments usually involve data collections where schools and students have been randomly assigned to different experimental treatments. Examples of treatments could be "Having a Classroom Library vs. Not Having a Classroom Library" or "Textbook A vs. Textbook B" or "Teaching Method X vs. Teaching Method Y". In the first case the experiment would consist of randomly assigning students and teachers to classrooms with libraries and classrooms without libraries. If the aim of the study is to see to what extent classroom libraries cause 'better' reading comprehension, then an appropriate test of reading comprehension is used and if, after a period of time, those with libraries achieve more on a reading comprehension test, then this is said to be due to the existence of classroom libraries.

If there is no difference in reading comprehension achievement between the two groups, then the existence of classroom libraries (and how they were used) is said not to make a difference.

In conducting an experiment the researcher must be careful about the factors that may limit the validity of the experiment. In particular, the following four questions need to be posed: (a) Does an empirical relationship between the operationalized variables exist in the population (statistical conclusion validity)?, (b) If the relationship exists, is it causal (internal validity)?, (c) Can the relationship between the operationalized variables be generalized to the treatment and outcome constructs of interest (construct validity, and cause and effect)?, and (d) Can the sample relationship be generalized to or across the populations of units, settings, and times (external validity)?

Much has been written about experiments (see the module on 'Research Design') but, in practice, it is often difficult to conduct experiments in educational settings, because Ministries of Education and School Principals are loath to permit the researcher to use 'randomization' as fully as is required for a valid experiment.

Tests

A test is an instrument or procedure that proposes a sequence of tasks to which a student is to respond. The results are then used to form measures to define the relative value of the trait to which the test refers.

I. Test items

A test may be an achievement, intelligence, aptitude, or practical test. A test consists of questions, known as items.

An **item** is divided into two parts: the stem and the answer. Stems pose the question. For example, a stem could be:

- What is the sum of 40 and 8?
- or in Reading Comprehension it could be a reading passage followed by specific questions.

The **answer** could be an 'open-ended', a 'closed', or a 'fill-in' answer.

For example, in the first stem given above an open or fill in answer could require the student to write the answer in a box. Or, it could be put into multiple choice format as follows:

- What is the sum of 40 and 8?
a. 84 b. 50 c. 48 d. 408

In this closed format the student is requested to tick the correct answer.

2. Sub-scores/Domain scores

The score of a student on the whole test is known as the “total test score”. A sub-score refers to the achievement of the students on a sub-set of items in the overall test. Thus, for example, in a Science test it may be considered desirable to classify the items into Biology items, Chemistry items, and Physics items. Each of these constitutes a domain of the test and the scores on each are known as ‘sub-scores’ or ‘domain scores’.

Another approach to reclassifying items is into ‘information’ items, items measuring the ‘comprehension’ of a principle, and items where the ‘application’ of skills are required. These are the first three categories of Bloom’s Taxonomy of Educational Objectives (Cognitive Domain) and are widely used. Again, the scores on each of these sub-classifications are known as sub-scores.

Variable

The term variable refers to a property whereby the members of a group being studied differ from one another. Labels or numbers may be used to describe the way in which one member of a group is the same or different from another.

Some examples are:

- ‘Sex’ is a variable with two values: The individuals being measured may be either male or female. The values attached for computer processing purposes could be 1 or 2.
- ‘Occupational Category’ is a variable that may take a range of values depending on the occupational classification scheme that is being used.

With variables like 'height', 'age', and 'intelligence', the measurement is carried out by assigning descriptive numerical values. Thus an individual may be 1.5 meters tall, 50 years of age, and have an Intelligence Quotient of 105.

1. Types of variables

Variables may be classified according to the type of information which different classifications or measurements provide. There are four main types of variables: *nominal*, *ordinal*, *interval*, and *ratio*.

a. Nominal

This type of variable permits statements to be made only about equality or difference. Therefore we may say that one individual is the 'same as' or 'different from' another individual. For example, colour of hair, religion, country of birth.

b. Ordinal

This type of variable permits statements about the rank ordering of the members of a group. Therefore we may make statements about some characteristics of an individual being 'greater than' or 'less than' other members of a group. For example, physical beauty, agility, happiness.

c. Interval

This type of variable permits statements about the rank ordering of individuals. It also permits statements to be made about the 'size of the intervals' along the scale which is used to measure the individuals and to compare distances at points along the scale. It is important to note that interval variables do not have true zero points. The numbering of the years in describing dates is an interval scale because the distance between points on the scale is comparable at any point, but the choice of a zero point is a socio-cultural decision.

d. Ratio

This type of variable permits all the statements which can be made for the other three types of variables. In addition, a ratio variable has an absolute zero point. This means that a value for this type of variable may be spoken of as 'double' or 'one third of' another value. For example, physical height or weight.

Note: It is not legitimate to apply simple arithmetic operations to nominal or ordinal variables. Addition and subtraction are possible with interval variables. Where ratio variables are used it is permissible to multiply and divide as well as to add and subtract.

Validity and reliability

Whenever individuals are measured on variables in educational research, there are two main characteristics of the measurement which must be taken into consideration: Validity and Reliability.

1. Validity

Validity is the most important characteristic to consider when constructing or selecting a test or measurement technique. *A valid test or measure is one which measures what it is intended to measure.* Validity must always be examined with respect to the use which is to be made of the values obtained from the measurement procedure. For example, the results from an arithmetic test may have a high degree of validity for indicating skill in numerical calculation, a low degree of validity for indicating general reasoning ability, a moderate degree of validity for predicting success in future mathematics courses, and no validity at all for predicting success in art or music.

There are three important types of validity in educational research: content validity, criterion-related validity, and construct validity.

a. Content validity

This type of validity refers to the extent to which a test measures a representative sample of subject-matter content and behavioural content from the syllabus which is being measured. For example, consider a test which has been designed to measure “Competence in Using the English Language”. In order to examine the content validity of the test one must initially examine the subject-matter knowledge and the behavioural skills which were required to complete the test, and then after this examination compare these to the subject-matter knowledge and behavioural skills which are agreed to comprise correct and effective use of the English language. The test would have high content validity if there was a close match between these two areas.

b. Criterion-related validity

This type of validity refers to the capacity of the test scores to predict future performance or to estimate current performance on some valued measure other than the test itself. For example, ‘Reading Readiness’ scores might be used to predict a student’s future reading achievement, or a test of dictionary skills might be used to estimate a student’s skill in the use of the dictionary (as determined by observation).

In the first example, the interest is in prediction and thus in the relationship between the two measures over an extended period of time. In the second example the interest is in estimating present status and thus in the relationship between the two measures obtained concurrently.

c. Construct validity

This type of validity is concerned with the extent to which test performance can be interpreted in terms of certain psychological constructs. A construct is a psychological quality which is assumed to exist in order to explain some aspect of behaviour. For example, “Reasoning Ability” is a construct. When test scores are interpreted as measures of reasoning ability, the implication is that there is a quality associated with individuals that can be properly called reasoning ability and that it can account to some degree for performance on the test.

2. Reliability

Reliability refers to the degree to which a measuring procedure gives consistent results. That is, *a reliable test is a test which would provide a consistent set of scores for a group of individuals if it was administered independently on several occasions.*

Reliability is a necessary but not sufficient condition for validity. A test which provides totally inconsistent results cannot possibly provide accurate information about the behaviour being measured. Thus low reliability can be expected to restrict the degree of validity that is obtained, but high reliability provides no guarantee that a satisfactory degree of validity will be present.

Note that reliability refers to the nature of the test scores and not to the test itself. Any particular test may have a number of different reliabilities, depending on the group involved and the situation in which it is used. The assessment of reliability is measured by the ‘Reliability Coefficient’ (for groups of individuals) or the “Standard Error of Measurement” (for individuals).

Indicator

An indicator generally refers to one or more pieces of numerical information related to an entity that one wishes to measure. In some cases, it consists of information about only one variable and this information may be gathered by only one question on a questionnaire. For example, consider an indicator of classroom library availability. In this case the indicator may be assessed by a single variable (which has only two values) that is measured by one question on a questionnaire:

Do you have a classroom library?

- Yes
- No

In some cases, an indicator may be made up of several variables. For example, an indicator of the 'Material Possessions of the Home' may need the addition of several variables. But, these variables may come from one question in a questionnaire:

Possession	No	Yes
Car		
Refrigerator		
T.V.		
Video		
etc.		

Which of the following exist in your home?

(Check once each row)

In other cases, several variables coming from several questions will need to be summed. For example, in a study in Indonesia (Mappa, 1982) an indicator of the 'Material Conditions of the Home' was produced. Four variables were used: Quality of floor, wall, roof, and lighting.

- **Quality of floor** was scored as:
bamboo = 1, wood = 2, cement = 3, and tiled = 4;
- **Quality of walls** was scored as:
woven palm leaves = 1, bamboo = 2, wood = 3, and cement = 4;
- **Quality of roof** was scored as:
palm leaves = 1, tile = 2, zinc = 3, and shingle = 4;
- **Quality of lighting** was scored as:
candle = 1, kerosine lamp = 2, hurricane lamp = 3, electricity = 4.

In this case there were four variables measured by four questions on a questionnaire. An indicator for each home was produced by adding the values for each of the four questions.

Attitude scales

Probably the least debated definition of an attitude is: "a moderate to intense emotion that prepares or predisposes an individual to respond consistently in a favourable or unfavourable manner when confronted with a particular object" (Anderson, 1985). In education, attitudes which are often measured are 'Like School', 'Interest in Subject-matter', and 'Teacher Satisfaction with Classroom Conditions'. Each of these titles implies a high to low measure. Thus, 'Like School' implies a measure that ranks students from those who love school to those who hate school.

Several ways of measuring and scaling attitudes have been devised. These are Thurstone Scales, Likert Scales, Guttman Scales and the Semantic Differential. These have been covered in detail in another module entitled “Questionnaire and Attitude Scale Construction”.

	Strongly disagree	Disagree	Incertain	Agree	Strongly agree
1. School is not very enjoyable					
2. I enjoy everything about school					
3. I am bored most of the time at school					
4. There are many subjects I don't like					
5. The most enjoyable part of my life is the time I spend at school					
6. I generally dislike my schoolwork, etc.					

Since the Likert Scale is the one most frequently used in educational research, a short explanation is given here. For example, consider the development of a ‘Like School’ scale to be used for 14-year-students. The researcher must first of all listen carefully to how 14-year-old students describe their like or dislike of schools. Both positive and negative statements are used, after editing, to form a set of statements about ‘Like School’. An example is given below:

The respondent is asked to check Strongly Disagree, Disagree, Uncertain, Agree, or Strongly Agree for each statement. Positive statements (like the second item) would be given the values 1 for Strongly Disagree to 5 Strongly Agree. Negative statements (like the first item) are coded to have 5 for Strongly Disagree and 1 for

Strongly Agree in order to have all reactions to statements being in the same direction. Only six statements have been given in the example above and hence the lowest score for a student would be 6 (6 x 1) and the highest would be 30 (6 x 5).

Appendix B

Further reading suggestions

Introductory texts

- Borg, W.R. and Gall, M.D. (1989). *Educational research: An introduction*. New York: Longman..
- Hopkins, C. D. & Antes, R. L. (1990). *Classroom measurement and evaluation (3rd ed.)*. Itasca, Illinois: Peacock.
- Oppenheim, A.N. (1992). *Questionnaire design, interviewing attitude measurement*. London: Pinter.
- Keeves, J.P. (1988). *Educational research, methodology, and measurement: An international handbook*. Oxford: Pergamon.
- Thorndike, R. L. & Hagen, E. (1977). *Measurement and Evaluation in Psychology and Education (4th ed.)*. New York: John Wiley.
- Wolf, R.M. (1991). *Evaluation in education (4th Ed.)*. New York: Praeger.

Examples of educational research studies that aimed to have an impact on educational planning

- Asmah bt Mohd Taib, Ahmad, Siatan, Solehan bin Remot, & Nordin, Abu Bakar. (1982). Moral education in Malaysia. *Evaluation in Education*, 6 (1), 109-136.
- Jiyono & Suryadi, Ace. (1982). The planning, sampling, and some preliminary results of the Indonesian Repeat 9th Grade survey. *Evaluation in Education*, 6 (1), 5-30.
- Mappa, Syamsu. (1982). Radio broadcasting program for out of school education in South Sulawesi (Indonesia). *Evaluation in Education*, 6 (1) 31-51.
- Murimba, S., et al (Ed.) (1995). The analysis of educational research data for policy development: An example from Zimbabwe. *International Journal of Educational Research*, 23 (4).
- Norisah bt Atan, H. Naimah bt Haji Abdullah, Nordin, Abu Bakar, and Solehan bin Remot. (1982). Remedial reading support program for children in Grade 2 in Malaysia. *Evaluation in Education*, 6 (1), 137-160.
- Ross, K.N. and Postlethwaite, T.N. (1992). Indicators of the quality of education: A summary of a national study of primary schools in Zimbabwe. Paris: UNESCO (IIEP).
- Sawadisevee, Amara, Padungrat, Jitsai, and Sukapirom, Rungrueng. (1982). Community secondary schools project in Thailand. *Evaluation in Education*, 6 (1), 83-107.
- Viboonlak Thongchua, Phaholvech, Nonglak, and Jiratprasoot, Kanjani. (1982). Adult education project – Thailand. *Evaluation in Education*, 6 (1), 53-81.

Encyclopedias and handbooks

Alkin, M. (Ed.). (1992). *American Educational Research Association sixth encyclopedia of educational research* (6th ed.). New York: Macmillan.

Husen, T. and Postlethwaite, T.N. (1994). *International Encyclopedia of Education: Research and Studies* (2nd Edition), Volumes 1-12. Oxford, Elsevier Science.

Keeves, J.P. (Ed.). (1988). *Educational research, methodology and measurement: An international handbook*. Oxford: Pergamon.

Walberg, H.J. and Haertel, G. (1990). *International encyclopedia of educational evaluation*. Oxford: Pergamon.

Journals

American Educational Research Journal
Applied Measurement in Education
Assessment in Education
Comparative Education Review
Educational Assessment
Educational Evaluation and Policy Analysis
International Journal of Educational Research
International Journal of Educational Development
International Review of Education
Journal of Education Policy
Research Papers in Education: Policy and Practice
Review of Educational Research
Studies in Educational Evaluation

Appendix C

Exercises

The following exercises are concerned with examining the general aims of an education system, establishing specific and operationalized aims, and then proposing research activities that will assess to what extent the education system is achieving its stated aims.

Five general aims are taken from a small publication «Planning for Successful Schooling» which was prepared by the Ministry of Education in the State of Victoria in Australia during 1990:

1. To expand educational opportunities for all students.
2. To encourage excellence in all areas of learning and to assist all students to develop their full potential.
3. To strengthen community participation in and satisfaction with the state school system.
4. To develop and improve the skills, potential and performance of school principals, teachers, and administrative and support staff.
5. To manage and control financial and physical resources in ways which maximize educational benefits for all students.

EXERCISE 1 (INDIVIDUAL WORK)

Select one of the five general aims above that you believe would probably receive a high priority in your country. For that general aim write five specific research questions. For each of the five specific research questions, prepare several operationalized research aims that focus on the performance of the education system in meeting these aims. Then, write down a broad outline of the sequence of activities that would need to be undertaken in order to assess the system's performance with respect to these aims.

EXERCISE 2 (SMALL GROUP WORK)

Collate the work of the individual members separately for each group. Compare and refine (or change) the wording of the specific research questions and operationalized aims for each general aim.

When this has been completed, discuss and write down, in outline form only, the sequence of activities to be undertaken in the research study or studies.

EXERCISE 3 (PLENARY GROUP WORK)

Collate the work of the small groups. Again refine the wording. Then write down in detail the sequential activities to be undertaken in a research study for each general aim covered to provide valid, reliable, and useful information for decision-makers to assess to which extent the general aims have been addressed.

Since 1992 UNESCO's International Institute for Educational Planning (IIEP) has been working with Ministries of Education in Southern and Eastern Africa in order to undertake integrated research and training activities that will expand opportunities for educational planners to gain the technical skills required for monitoring and evaluating the quality of basic education, and to generate information that can be used by decision-makers to plan and improve the quality of education. These activities have been conducted under the auspices of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ).

Fifteen Ministries of Education are members of the SACMEQ Consortium: Botswana, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania (Mainland), Tanzania (Zanzibar), Uganda, Zambia, and Zimbabwe.

SACMEQ is officially registered as an Intergovernmental Organization and is governed by the SACMEQ Assembly of Ministers of Education.

In 2004 SACMEQ was awarded the prestigious Jan Amos Comenius Medal in recognition of its "outstanding achievements in the field of educational research, capacity building, and innovation".

These modules were prepared by IIEP staff and consultants to be used in training workshops presented for the National Research Coordinators who are responsible for SACMEQ's educational policy research programme. All modules may be found on two Internet Websites: <http://www.sacmeq.org> and <http://www.unesco.org/iiep>.
