Planning the quality of education

The collection and use of data for informed decision-making

Edited by
Kenneth N. Ross
and
Lars Mählck

Unesco: International Institute for Educational Planning
Pergamon Press

Chapter 6

Improving data collection, preparation, and analysis procedures: a review of technical issues[1]

Introduction

Improving the collection, preparation, and analysis of data that are required to guide decisions aimed at improving the quality of education requires, in addition to a concern about the scope of the data collection, careful design and management of the data collection and preparation procedures (especially in the areas of sampling, instrumentation, field work, data entry and data preparation), and appropriate data analysis and reporting.

These concerns are related through the costs involved in each, with a tradeoff between scope (broad targets for, and the frequency of, data collection) and depth (the complexity of the data collection instruments and related data collection process). That is, a decision to collect data must be informed by prior decisions regarding the units of observation (how many), the questions to be answered (how many), and the resources available for data collection. In general, for the same costs, more questions may be asked of fewer respondents or fewer questions may be asked of more respondents. Common errors in data collection often arise from a desire to collect more

1. This chapter was prepared by Kenneth Ross, T. Neville Postlethwaite, Marlaine Lockheed, Aletta Grisay and Gabriel Carceles Breis.

information from more respondents than resources realistically permit

This chapter is focused mainly on the technical issues that often arise after the initial decisions have been taken with respect to what data should be collected. However, it is important to note that decisions taken concerning what data should be collected need to be supported by an explanation of why these data are be collected and how these data are to be collected, prepared, analysed, and then used by decision-makers operating at various planning levels of an education system.

In essence, improving the collection, preparation and analysis of data requires attention to detail. That is, there are no "shortcuts" to achieving high standards in these areas. However, there are many easy ways to ensure that standards are low and that the results of the data analyses are meaningless. In the following sections of this chapter a discussion of some of the technical issues associated with data collection, preparation and analysis have been presented within a framework that puts the "*basic requirements*" for success alongside the authors' observations, in both developed and developing countries, of "*what often happens*" to prevent success. The chapter concludes with an exploration of "*what might be done in future*" -- especially in terms of training programmes designed to improve the capacity of educational planners to undertake productive projects in this area. While the paper concentrates on the features of survey sample approaches, many of the issues raised are pertinent to most data collection activities -- including censuses.

The scope of the data collection

The first issue to be decided before any data are collected, is the scope of the data collection. That is, an initial decision must be taken with respect to the question: "Information about what?". For example, are data to be collected for a whole population (of say schools, teachers, or students) or is a probability sample to be selected from a well-defined target population? Data collection efforts in many countries concentrate on enumerating the population of students, teachers and schools. This is a decision that, because of the breadth/depth tradeoff, yields little information about many units. Some data on the entire population (that is, a census) need to be collected regularly in order to inform managerial decisions, such as the allocation of resources. These types of data include the total number of students, teachers, and schools. In some countries, data on student populations are required in order to monitor compliance with compulsory education laws.

For most purposes, sample surveys with units sampled from sampling frames developed for a well-defined target population are sufficient. Sample surveys, when designed and executed appropriately, can provide as much information as complete censuses, at considerably less cost. For example, sample surveys are often adequate for providing accurate estimates of enrolment rates, and are virtually mandatory for estimating national achievement levels, particularly for students in grades not regularly examined for selection purposes.

Whether a decision is made to use a census or a sample, it is important to ensure that data that are collected can be reported as quickly as possible. The collection of too many data for too many units may be counter-productive because it may result in delays in the final reporting of results. This is clearly unsatisfactory because decision-makers need to obtain a clear picture of the current state of an education system before considering the particular policy measures required to improve the quality of education. In situations where time-series data is being used, this is extremely important because decision-makers will be analyzing trends in the data and they need to know quickly if some overall trend downwards, or upwards, in the quality of education has occurred.

The sample design

Due to practical constraints on research resources, data collections that include an assessment of educational outcomes for students are usually restricted to the study of a sample rather than a complete coverage of the population for which these generalizations are required. Provided that scientific sampling procedures are used, the use of a sample often provides a number of advantages compared with a census. For example, reduced costs associated with all aspects of data collection and analysis, reduced requirements for specialized personnel to conduct the field work, greater speed in most aspects of data manipulation and summarization, and greater accuracy due to the possibility of closer supervision of the fieldwork. Samples are perfectly adequate for describing most characteristics of an education system. In fact most analytic work depends upon samples, even when census type data are available because the computational requirements for analyzing complete population data are often very large.

Good sample designs for studies of educational outcomes do not occur by chance -- they are constructed by using established sampling procedures in association with a practical knowledge of the ways in which populations of schools, teachers, and students are administratively and geographically arranged. While the optimal sample design for a particular data collection in a particular country will always contain many unique features, the basic requirements listed in the following section are common to most well-designed samples.

Sample design: basic requirements

(a) *Target population definitions.* Descriptions should be prepared for the desired target population (the population for which results are ideally required), the defined target population (the population which is actually studied and whose elements have a known and non-zero chance of being selected into the sample), and the excluded population (the population comprised of the elements excluded from the desired target population in order to form the defined target population).

(b) *Specification of domains and strata.* The domains for the data collection (the sub-populations for which separate estimates are specifically planned) should be nominated. The stratification variables should then be selected and justified in terms of gains in sampling precision.

(c) *Sampling error requirements.* The required level of sampling precision (the permissible boundaries of sampling error associated with sample estimates of important population parameters) should be established and this should be checked against prevailing administrative, financial, and political constraints.

(d) *Size of sample.* The size of the sample should be calculated by using information concerning the proposed sample design (the number of stages of selection in the sample design, the stratification procedures, the nature of the primary sampling units, the magnitude of the coefficient of intraclass correlation and/or the design effect), the required level of sampling precision (see above), and the proposed data collection environment (for example, the numbers of students per school that, under standardized conditions, can be administered the instruments associated with the data collection).

(e) *Sampling frame*. The sampling frame should be constructed for the defined target population in separate parts representing the strata. Appropriate measure of size figures (Kish, 1965: 222) should be assigned, and then a check should be made to ensure that the total and stratum subtotal numbers of students in the frame are in accord with the numerical description of the defined target population.

(f) *Mechanical selection procedure*. A suitable mechanical selection procedure (Kish, 1965: 26) should be applied in order to select the sample members from the sampling frame with known probabilities of selection.

(g) *Sampling weights and sampling errors*. Appropriate sampling weight calculations and sampling error estimation techniques should be selected in order to cope with any complexities (stratification, multiple stages, clustering) that have been introduced into the sample design.

Sample design: what often happens

(a) *The defined target population and the excluded population are never clearly defined*. This may arise because the researcher either does not bother to specify the size and nature of these populations or, due mainly to confusion, is unable to provide precise definitions. Unfortunately, this problem often goes hand-in-hand with the researcher making generalizations about a desired target population that, upon careful scrutiny, consists mostly of the excluded population.

(b) *The participants in the study are nominated rather than sampled*. This approach is often justified in terms of cost or accessibility considerations, however both of these "constraints" can usually be addressed by adjusting the defined target population definition and then applying appropriate stratification procedures. These non-probability samples, sometimes referred to as "nominated samples", are generally described in scientifically meaningless terms such as "quota", "representative", "purposive", "expert choice", or "judgemental" samples. Kish (1965) characterized data collections based on this approach as "investigations" and pointed out that they should not be confused with appropriately designed experiments or surveys. The main problems associated with the use of nominated samples are that it is not possible to estimate the sampling errors or to have any idea of the magnitude of the bias associated with the selection procedures (Brickell,1974). Consequently, nominated samples should be used only for the trial-testing of instrumentation or new curriculum materials because in these activities it is sometimes desirable to employ a "distorted" sample that has, for example, a disproportionately large number of students at the extremes of a spectrum of ability, ethnicity, socio-economic status, etc.

(c) *The sampling frame is faulty because it is out of date and/or is incomplete and/or has duplicate entries.* The construction and maintenance of a comprehensive sampling frame for schools, teachers, and students may be neglected because it is considered to be too expensive or because the systematic collection of official statistics in a country is error-prone. This is sometimes the situation in countries where population growth rates are high and where large and uncontrolled movements of population from rural to urban settings are commonplace. However, there are also a number of countries that are unable to provide accurate information in this area because the management and financing of schooling is undertaken by local communities, or because there is an independently managed non-government school sector. The researcher faced with these difficulties often proceeds to use a faulty sampling frame based on poor quality official statistics in the mistaken belief that there are no other alternatives. In fact there are well-established solutions to these problems that employ "area sampling" (Ross, 1986) and, provided that a trained team of "enumerators" is available to list schools within selected areas, it is possible to prepare a high quality sample design without having access to an accurate sampling frame based on a listing of individual schools.

(d) *Confusion surrounding the terms "total sample size" and "effective sample size" results in the total sample size for a complex cluster sample being set at the wrong level either by the use of simple random sampling assumptions or, quite frequently, by guesswork.* In school systems that are highly "streamed", either explicitly on the basis of test scores or implicitly through the residential segregation of socio-economic groups, the use of complex cluster sampling can have dramatic effects on the total sample size that is required to reach a specific level of sampling precision. Researchers with a limited knowledge of this situation often employ simple random sampling assumptions for the estimation of the required total sample size. In order to illustrate the dangers associated with a lack of experience in these matters, consider the following two examples based on schools in a country where the intraclass correlation for achievement scores at the Grade 6 level is around 0.6 for intact classes. A sample of 40 classes with 25 students selected per class would provide a total sample size of 1000 students -- however this sample would only provide similar sampling errors as a simple random sample of 65 students when estimating the average population achievement level! Further, a sample of 50 classes with 4 students selected per class would provide a total sample size of "only 200 students" but would nevertheless provide estimates that were more precise than the above sample of 1000 students!

(e) *The wrong formulae are used for the calculation of sampling errors and/or for the application of tests of significance.* This usually occurs when the researcher employs a complex cluster sample (for example, by selecting intact classes within schools) and then uses the sampling error formulae appropriate for simple random sampling to calculate the sampling errors (Ross, 1985). The most extreme form of this mistake occurs when differences in means and/or percentages are described as being "important" or "significant" without providing any sampling error estimates at all -- not even the incorrect ones! These kinds of mistakes are quite common -- especially where "treatment versus control" comparisons are being made in order to compare, for example, current practices with new curriculum content or new teaching materials (Ross, 1987).

(f) *The researcher has undertaken a few hours of training in sampling as part of a "research methods" course and this provides* *just enough knowledge and personal confidence to precipitate the occurrence of major errors.* The five errors that have been observed most often by the authors have been listed below.

- *Population size and sample size.* A researcher uses the once-fashionable sampling approach of basing sample size on a fixed percentage of the population size because he believes that a large population requires a large sample, and vice versa. In fact, for most surveys in educational research, the finite population correction factor in sampling error formulae is very close to one and therefore the population size need not be considered when planning the size of the sample.
- *Bias adjustment for non-response.* A researcher designs a sample so that it is 25 percent larger than required in order to cope with an expected response rate of around 70 to 80 percent. This approach may deliver a final sample with the required total sample size, but it will not guarantee freedom from bias because non-response is often associated with a sub-population that has unique characteristics. A failure to understand this point is due to confusion between two independent sources of "total error": sampling error and bias.
- *Units of sampling and units of analysis.* A researcher employs a commonly-used sample design based on a probability proportional to size sample of schools followed by a simple random sample of a fixed number of students in each sample school. This sample design is self-weighting for students and therefore the between-student data analyses are reasonably straightforward. However at the between-school level of analysis caution needs to shown with respect to interpreting univariates based on weighted and unweighted school means (Kish, 1965: 186).
- *Bias in selection.* A researcher has an accurate list of schools to use for a sampling frame and proceeds to select a simple random sample of schools followed by the selection of a simple random sample of a fixed number of students within each selected school. The researcher then proceeds to conduct unweighted data analyses without realizing that the sampling procedures have given students in smaller schools a much higher chance of selection than students in larger schools.
- *Sampling errors output by the standard statistical software packages.* A researcher conducts a survey using a sample of

students selected as intact class groups and then uses a statistical software package (for example, SAS, SPSS, BMDP, or MINITAB) to conduct significance tests on the differences in two sample means. The test statistics that are provided in the computer output are likely to be incorrect because they will be based on formulae that assume simple random sampling.

The design of data collection instruments

Data collection requires the use of some medium for "collecting" data: notebooks, questionnaires, optical scanning forms, and in some cases microcomputers used in the field. The type of data collection instrument will determine the type of data preparation required before the data analyses. Open-ended questionnaires or interviews require more preparation effort than do pre-coded questionnaires, and pre-coded questionnaires require more than questionnaires that are designed for optical scanning, which, in turn, require more attention to data preparation than do data that are directly entered into a microcomputer.

Data collection instruments: basic requirements

The data collection instruments should be clear in terms of the information they seek, retain data disaggregated at an appropriate level, and permit the matching of data within hierarchically designed samples or across time. Furthermore, they must be designed to permit subsequent statistical analysis of data for reliability and (if possible) validity. In this chapter, it is not appropriate to describe in great detail the procedures for good instrument development. However, the basic requirements are that the questions posed do not present problems of interpretation to the respondent, and that, when forced choice options are provided, the choices are mutually exclusive and are likely to discriminate among respondents. Since in most countries, mainframe computers permit the storage of data at considerable levels of disaggregation, data collection instruments need to allow for this level of disaggregation. If hierarchically designed samples are developed (containing, for example, data for students, their teachers, their schools, their parents, etc.) and the merging of data from different sources is required for data analysis, then the instruments must include appropriate identifying numbers for linking the different sources of data.

Data collection instruments: what often happens

(a) *Poor physical layout.* Many data collection instruments reflect a desire to reduce costs at the outset through limiting the amount of paper used. The result is that questions are shortened to the degree that their meaning is unclear, response alternatives in forced choice questions are not elaborated to the degree that their meaning is clear, information is crammed into sheets of paper with little concern for subsequent data preparation and coding. The result of a poor layout is that considerable errors are introduced into the data. For example, a school survey in Africa listed all household members on a single sheet of paper, but data about each member were entered, in some cases, vertically on the paper and in other cases, horizontally. To enter the data into the computer from this form would have taken more time than to have recopied the data into another instrument.

(b) *Lack of question pretesting.* Pretesting questions is a necessary step in instrument development, and one that is frequently overlooked. The result of a failure to pretest is that respondents (particularly in a large sample survey, where an individual "enumerator" is not available to clarify matters) can be confused by a question, and answer inappropriately. When obtaining measures of student achievement, pretesting is absolutely necessary so that items may be checked for their difficulty and discrimination levels. If items are either too hard or too easy, there will be little discrimination in the resulting test score. Where open-ended responses are to be subsequently coded into categories, pretesting can assist in the development of the categories, or can even lead to eliminating the need for a separate coding step. For example, in one international survey, students were asked to indicate the total number of brothers and sisters in their family. The question was asked in the form of a forced choice response with the maximum value being "five or more" and in several countries more than 80 percent of the children indicated this category as their choice. Pretesting would have indicated the need to extend the number of categories to allow for the very large family sizes in these countries, or to leave the question open-ended.

(c) *Failure to use technology.* The use of some types of technology can reduce errors associated with data preparation. For example, optical scanning sheets when used in appropriate situations can be read by an optical scanner much more quickly than printed instruments can be hand coded, the use of this type of technology can reduce the length of time between data collection and analysis. Other types of technology, such as using microcomputers in the field for entering data, can also improve the quality of data collections.

The management of the data collection

The management of the data collection requires the researcher to arrange for a standardized administration of the research instruments (tests, questionnaires, etc.) to the persons selected into the sample. In practice, this means that the researcher, or a suitably trained field-worker, will go to the school and administer the relevant instruments in the manner prescribed by the research design to the appropriate students, and, in some cases, to the appropriate teachers, school principals, parents, and community leaders. There may also be a requirement for the recording of observational data concerning classroom teaching, school management, environmental conditions, etc. In large studies of educational outcomes, for example the "Indonesian Quality of Education Study" (Postlethwaite and Ross, 1987), all of the data sources listed above will be involved and will be intimately interconnected in the sense that it will be important to be able to link each student's data with the data describing his/her own classmates, teacher, school principal, parent, and community leader.

In the following discussion, some of the basic requirements associated with conducting an effective data collection for a study of educational outcomes have been outlined. These requirements are centred around the need for the principal researcher to maintain control of field operations through an effective management plan and the use of high quality field manuals.

Data collection: basic requirements

(a) *Maintenance of control over the execution of the sample design in the field.* All persons involved in the field work should understand that the researcher is the only person who has the authority to nominate the various sampling units for inclusion or exclusion. This requires that the researcher should supervise the preparation of an unambiguous list of the *names* of the schools to be involved in the data collection and then, following a preliminary visit to the nominated schools by an trained enumerator, the researcher should supervise the preparation of an unambiguous list of the *names* of the students to be involved. Detailed instructions should also be prepared so that there is no doubt as to identity of any other persons (teachers, school principals, etc.) that are to be included in the data collection.

(b) *The role of the researcher in the data collection.* Sustained staff enthusiasm is essential in order to ensure the success of the data collection and this can only be achieved if the researcher gets involved with the work at all levels. "Getting involved" means doing much more than giving speeches, writing papers, circulating memos, or talking on the telephone -- it means taking personal responsibility for day-to-day practical operations and actually doing some of the less glamorous research tasks. For example: participating in the training of the field workers, helping with some of the field work, making personal visits to sample schools that may be reluctant to take part in the data collection, being accessible to staff and showing appreciation for their good work, meeting regularly with supervisory staff in order to monitor progress.

(c) *The preparation of field manuals.* The data collection at the school level should be planned in great detail and these plans should be outlined in two easily-understood field manuals: the "School Co-ordinator's Manual" and the "Test Administrator's Manual". The School Co-ordinator's Manual should describe every step to be taken by the person responsible for the data collection in each school -- from the time the instruments arrive in the school to the time they are packaged and returned to the central research office. In some situations this will be the School Principal, or a teacher in the school appointed by the school principal. In other situations it will be a field worker appointed by the researcher to go to the school in order to arrange the data collection. In both situations the manual needs to be written so that the School Co-ordinator is absolutely clear about what is required to be achieved and this should be reinforced by providing a training programme during which all of the materials to used in the data collection are presented and explained clearly. The Test

Administrator's Manual should be a separate document from the School Co-ordinator's Manual in order to cover the situation where someone other than the School Co-ordinator is to be responsible for the administration of the data collection instruments. In order to standardize the administration conditions across all of the data collection points, the Test Administrators Manual should be written in the form of a scripted play (with prompts) so that there is little, or no, opportunity for the Test Administrator to confuse the respondents. If observation schedules are to be used in the data collection, these two manuals will need to be supplemented by further training which includes inter-rater reliability investigations.

Data collection: what often happens

(a) *The researcher is unable to manage the project after it has been launched and therefore cannot monitor daily data collection operations in a manner that will permit timely and effective responses to be made to major crises.* The main reasons for these difficulties are usually associated with the following four areas. Each area has been illustrated with examples that have been observed by the authors.

- *Inadequate training and/or experience.* The researcher has either never been trained in the techniques of data collection or has had some theoretical training but no practical experience. For example: A government engineer working on projects that were mainly concerned with road construction was suddenly told, because an international aid agency had requested an evaluation of several of its projects, to conduct a series of large-scale educational evaluation studies. A second example: an education ministry official returned to work, after completing a Ph.D. based on a study of the linguistic development of eight children, and was directed to undertake a national evaluation programme for all of the core subjects in each grade level of primary schools.
- *Professional status.* Professional status sometimes prevents the researcher from getting involved with the mundane but difficult aspects of the data collection and consequently when "real" problems arise in the field he/she is not able to respond with a workable solution because of an incapacity to understand the difficulties in practical terms. For example: A researcher was

required to supervise the data preparation for a national testing programme involving ten thousand students but would not acknowledge that he needed to do some of the data preparation himself in order to know how to train and supervise the team selected to do the work. A second example: A researcher chose to participate in the field work for a few first-class schools in the wealthy neighbourhoods of a province that was known internationally for its excellent tourist attractions, fine cuisine, and first class hotels. The data collection for these schools was found to be running smoothly, and the researcher returned home after having a relaxing and comfortable "holiday" -- but without getting any first-hand knowledge of the kinds of difficulties that were being experienced by his own field workers in other provinces.

Too many jobs. The capacity of well-qualified and experienced people in developing countries to do research is often diminished for two main reasons. First, people with this high level of training are in short supply and therefore they are always being asked by their governments to take responsibility for a range of administrative tasks and a large number of research studies. Second, these people, mainly through education and travel, have often acquired a taste for a life-style that exceeds the buying power of their official government salaries with the result that they become involved in "outside work" for both the government and the private sector. For example: A researcher, having completed a Ph.D. that included training in most aspects of survey research methods, was given responsibility for supervising the data collection for one of the largest studies of elementary schooling ever conducted. At the same time he was the principal researcher for another large project, a part-time lecturer at two universities and a community college, a paid member of a range of advisory committees for other large research projects, an active partner in a private consultancy company working on various large-scale contract research tasks, a member of committees responsible for drafting government policy papers, a "counterpart" for a constant stream of overseas consultants, and was also required to prepare answers to Ministerial requests for information on education.

Communication difficulties. In some countries the standard forms of communication (mail, telephone, telex, etc.) are often inaccessible or not operating effectively. Consequently, it is not always possible for problems in the field to be conveyed to the researcher in good time to attend to difficulties before they begin to disrupt the data collection. Some of these problems cannot be anticipated in advance and therefore they cannot be included in either the School Co-ordinator's Manual or the Test Administrator's Manual. For example: A field worker travelled to a sample school in a remote part of a country and found that the data collection could not proceed for two sample schools because one was closed when the teacher left without informing the Ministry of Education, and the other did not exist because the buildings had recently been destroyed by a landslide. Should the field worker conduct the administration in nearby schools? A second example: A field worker in the same country found that a sample school in a mountainous region could not be reached because it was on the other side of a flooded river. Should the field worker wait for the flood to subside?

(b) *The field staff make on-site alterations to the sample design without seeking approval.* There are many temptations for the field worker, especially a poorly-paid one working in an isolated region of a country, to engage in "substitutions" whereby students, or even whole schools, are substituted for the students and schools in the "official sample design". These pressures are sometimes financial -- because the field workers are paid fees and per diems on the basis of an agreed number of schools, they are sometimes cultural -- because local custom makes it difficult to say "no" when a school principal or a regional official "suggests" that only the best students should be tested, and they are sometimes due to the field worker replacing the sample school with another school -- because the "replacement" school was more accessible.

It is extremely difficult to obtain complete protection against these kinds of actions. However, all data collections should attempt to include some form of "external validation" through either conducting a "post-enumeration survey" (Casley and Lury, 1981), or through including a few already-known pieces of information in the data collection (for example, class/school size, age of school principal, number of teachers, sex-composition of school) and then later comparing these returns with official records. Unfortunately, these issues are often ignored or,at best, a half-hearted "internal

(e) *The data collection date is set at an inappropriate time.* The researcher needs to be extremely careful about setting the date for the data collection because a bad choice of dates can sometimes lead to poor response rates or unreliable data. For example, the "school climate" at certain times of the school year may not be appropriate for a data collection to proceed (due to impending vacations, approaching examinations, etc.), or there may be certain times of the year when religious festivals, especially those that require fasting, are likely to affect the collection of data from students.

(f) *The researcher lacks local knowledge.* There are many different situations where a lack of local knowledge can interfere with the collection of accurate data. For example, a researcher, who was responsible for estimating school participation rates gathered data in three provinces of a country where Islam was the predominant religion without realizing that many students would be attending two schools (government in the morning and Islamic in the afternoon). The resulting "double counting" of students attending school provided estimated participation rates in some villages of 120 percent! All of the field work for this part of the project needed to be repeated using appropriate counting methods.

The management of the data preparation

The data preparation phase is concerned with the transformation of raw data obtained during the data collection phase into a form that is suitable for later analysis. There are two main steps involved in this: data coding and data entry. *Data coding* requires the allocation of numerical codes to each piece of information gathered during the data collection. Sometimes this allocation will be self-evident -- as in allocating a 1, 2, 3, or 4 to the first, second, etc. responses to a test item. At other times it will require a careful use of special tables of information -- as in allocating a score on a scale of "socio-economic status" for an open-ended response describing the occupation of a student's father. *Data entry* requires the transformation of all of these codes into a form that can be "read" by a computer.

The data preparation phase of the work can spoil efforts that have been made to ensure a well-executed data collection. This often occurs because many researchers tend to see data preparation as being unworthy of their full attention since it involves neither the practical challenges of data collection nor the intellectual stimulation that goes along with data analysis. The authors have participated in the "recovery" of several data collections that had initially failed because the researchers responsible for them held this kind of attitude. In each of these cases it was found that the recovery was very expensive in financial and manpower terms.

Data preparation: basic requirements

(a) *The coding team and their working environment.* The data coding should be conducted by a trained team under the careful supervision of an experienced person who understands the content and purposes of the data collection. This team should be given a quiet and comfortable room in which to work so that there is little likelihood that the "normal distractions" of office life will occur. (For example, telephones should be removed or silenced, excessive social chatter should be discouraged, and office traffic should be kept to a minimum.) The room should be furnished in a manner that permits the coders to work quietly without interrupting each other. (For example, each coder should have a table that is sufficiently large to facilitate the management of questionnaires, tests, and related coding documents, and there should be adequate shelving and storage areas for the completed instruments.)

(b) *The "Codebook"*. The data coding should be carried out according to the instructions set out in a Codebook that has been prepared by the researcher for the coding team. The Codebook should include an accurate reproduction of the each of the questions and test items, a list of the answers that are possible for these, a list of the codes that are to be assigned to the possible answers, an explanation of the missing data codes that are to be used, information describing the scoring and re-coding that will take place on the computer (for example, when test scores are produced, or highly detailed classifications are "collapsed" into a smaller number of categories), and the location of each coded value after it has been entered into the computer data file. The Codebook should contain sufficient information to permit a person who has had no prior knowledge of the data collection to understand the meaning and origins of every numerical value stored in the computer data file.

(c) *The coding procedures*. The coding should commence with a rapid general edit of the questionnaires in order to check for any errors associated with obvious omissions or inconsistencies and, if possible, to correct these errors. For example, at this point it will be possible to identify whether there are any missing questionnaires and to check the reasons for this with the field staff. The next task is to select a sample of around 25 questionnaires that will be coded by all members of the coding team and also by the researcher. If there are any differences in codes allocated for particular questions these should be noted and made the subject for a rentable discussion so that the researcher and all members of the coding team reach agreement about the requirements of the task and also about any areas of the coding that may occasionally require a "second opinion". At this stage, the coding of the all of the questionnaires should commence and the supervisor of the coding team should be available to participate in the work, to answer questions, and to conduct quality control checks by inspecting samples of completed work. The supervisor should keep an accurate record of exactly which questionnaires have been coded by each member of the coding team so that it will be possible to find and correct questionnaires when quality control procedures reveal that a particular person is "error prone". The researcher should meet with the supervisor at the end of each day in order to discuss any problems and, from time to time, should participate in the coding in a enthusiastic manner that demonstrates to the coding team that their work is important and that the researcher cares a great deal about this aspect of the data preparation.

(d) *Data entry*. The data entry component of data preparation refers to the transformation of the information compiled at the coding stage into a form that can be "read" by a computer. This is normally achieved by entering the codes into a computer via a terminal or a personal computer. The persons carrying out the data entry usually work either from sheets of figures produced during the data coding or from the questionnaires and tests themselves in the case where these contain mostly "self-coded" responses. Generally there will be less errors if the data are entered directly from the questionnaires and tests because this reduces the chances of error related solely to transcribing, however this approach may increase the time required for the data entry because of the need to read from various page locations and different pages. In some cases it may be possible to employ optical scanning whereby the respondent enters responses onto a form that can be read directly into the computer. This more advanced approach needs to be used with care, especially when applying it with young children, because a certain amount of maturity and dexterity is required on the part of the respondent in order to handle these forms in a manner that does not diminish the validity of the responses.

Several guidelines should be observed during the data entry: all codes should preferably be numeric, each respondent should have a unique identification code that includes both the respondent's location (for example, country, state, province, district, school, class) and the associated sampling frame information (for example, domain, stratum, substratum, cluster), a clear distinction should be made concerning the various forms of non-response (for example, omitted, not present, not reached, etc.), the value zero should not be used to indicate non-response, a "check digit" should be inserted every 10 or 20 columns in order to permit a visual check of data alignment to be made rapidly from the print-out, and the data should be "double punched" or validated in some similar manner.

In ideal circumstances a specialist data editing programme should be used to monitor the data entry in "real time" by conducting pre-programmed logic checks on the data as they are entered. These checks are conducted by the computer and may be as simple as a basic range check or as sophisticated as a complex check for unlikely combinations of many codes.

(e) *Data cleaning*. The final part of the data entry is called data cleaning and it consists of running a series of preliminary analyses on the data in order to look for errors, omissions, etc., and then employing the results of these analyses to edit the original data file. These analyses should expose some or all of the following problems: differences between the number of cases on the computer file and the number of questionnaires, non-numeric codes, out-of-range errors, logical consistency irregularities, mismatches between data collected at different levels (for example, data may be available for a particular teacher but not for that teacher's students), and errors in the preparation of composite variables on the computer. As a "rule of thumb" no more than three sets of these preliminary analyses should be undertaken because the authors' experience shows that, provided a careful clerical examination is undertaken on the results of the analyses, further runs do not always justify the time and computer resources involved.

After three sets of computer runs, the data should be run through a "conditioning" programme that either sets the values of imperfect composite variables to missing data according to pre-specified rules (for example, assigns a value of "missing" to a total test score when more than ten percent of the items are missing) or creates an imputed value for an imperfect composite variable (for example, assigns the class mean score for a student who did not provide information about one of the variables that is to be included in a construct describing the socio-economic circumstances of the student's home background). There is no single correct way with which to deal with non-response and therefore the researcher's task is to "choose the method with the least disadvantages for a specific situation" (Kish, 1965: 558). It should be remembered in the treatment of missing values that "doing nothing about it" (for example, by excluding the missing responses from all calculations of sample means) usually requires the simple, but usually incorrect, assumption that the non-respondents are sufficiently similar to the respondents to justify ignoring them in the calculations.

The final task of the data cleaning work is to produce two copies of the data files in addition to the working file. One of these copies should be stored on-site for backup purposes or for use in transferring data between computers, the other copy should be stored in appropriate long-term, secure, off-site storage.

Data preparation: what often happens

(a) *The coding and data entry team is untrained, poorly supervised, and works in an inadequate environment.* The implications of this kind of situation are best described by reference to the authors' experience in trying to recover a dataset that had been prepared under very poor conditions for a national survey of Grade 9 students. The coding team hired for the survey consisted of young university undergraduates and they worked on the coding in very cramped conditions on bench tables that had been placed in the corridors of the Ministry of Education building. The team was given a minimal amount of training and little, or no, supervision. Their main task was to transcribe student responses to questionnaires and tests onto coding sheets and then to enter the data into a computer.

The coding team did not complain about their working conditions -- indeed the young males seemed to enjoy being confined in small and unsupervised working area along with their attractive female colleagues. The roar of conversation, occasional squeals of laughter, incessant corridor traffic, smiling faces, etc. showed that the team was having a very good time as they spent several months working their way through the tests and questionnaires associated with some five thousand students. When the coding and data entry was completed, the tests and questionnaires were bundled together in a rather haphazard fashion and sent off to storage in a government warehouse located ninety minutes drive away.

The researchers conducted no preliminary analyses in order to check the quality of the data preparation. Instead, with an extraordinary demonstration of courage they launched into the main analyses using huge amounts of expensive computer and programmer time in scoring, merging, and analysing. However, it wasn't very long before some of the findings set the danger bells ringing: the mean scores on the multiple choice tests were close to values that would be expected if all the students had guessed the answers, a number of variables that had four possible responses had many values in the range 5 to 9, strange student and school identification codes were appearing in large numbers and seemed to resist any attempts made by the researchers to use them for file merging purposes.

A decision was made to check the first 100 cases on the computer files by referring back to the questionnaires and tests. This took a little while to get started because the researchers had to first go back to the warehouse and then to find the correct questionnaires among the thousands of unsorted returns from students, teachers, and school principals that had been stored in over a hundred cardboard boxes. When the checking was finally completed it was discovered that the miscodes on some of the coding sheets were as high as 75 percent. In addition, a large number of errors were discovered to be associated with the data entry work. There were so many errors that the only solution was to start the whole coding and data entry operation again from the beginning.

(b) *The coding team scores the tests and scales by hand.* It is still quite common to see researchers directing their coding teams to add the item scores to obtain a total test score and then enter the total scores, not the item scores, into the computer file. This inevitably results in the unnecessary inclusion of an extra source of error into the data preparation. The authors have noted a number of instances of this approach in projects that were conducted under the direction of external "expert" consultants. In addition to introducing errors into the total test, this procedure prevents the researcher from being able to conduct item analyses so as to remove poor items, to check the suitability of item membership of subtests, and to use item characteristics to check the validity of the answer key for the test. Many people would find the last item mentioned here a little unusual. However, the authors' experience in both developed and developing countries has been that errors may be be found in answer keys through the use of item analyses in around one out of every four projects.

(c) *The coding team is given inadequate documentation.* It is important that the documentation given to the coding team covers all possible situations. That is, the members of the coding team should never be left to "work it out amongst themselves". An example of this occurred in a country in Europe where, instead of being given a standard table, the coding team was left to transform dates of birth into a variable referring to age in months. This variable was a key marker variable for the project and the many "out of range" values cast doubt not only upon the sampling procedures, but also prevented the use of the variable as an important control for student achievement scores.

(d) *The researcher fails to "fine-tune" the early stages of the coding operation.* During the first stages of the coding operation it is important to check the validity of the the Codebook with respect to the open-ended and free response items. It is often the case that the respondents will provide some answers that do not completely fit all of the possibilities that are listed in the Codebook. In some circumstances this will require further categories to be added, while in other cases it may require a complete rewrite of a section of the Codebook. These changes need to be made very quickly and this is yet another reason why the researcher needs to participate with the coding team in the early parts of the data preparation. An example from a project in Australia illustrates this with respect to the coding of occupations on a scale of occupational status. The coding manual for the project provided the coders with a table to be used to look up a "socio-economic status score" based on the question "What is the name of your fathers occupation?". The coders dutifully followed these instructions without taking advantage of the much more informative responses that the students had provided to a later question which asked "What does your father do when he is at work?". In a large number of cases, students received a "missing data" code because they had not responded to the first question or had written "I don't know". Nearly all of these students could have been assigned valid scores if responses to the later question had been taken into consideration by the coding team.

(e) *The data are "lost"*. The data collected for the project are often lost shortly after the researcher completes the project, and this prevents important secondary analyses from being undertaken. Some of the main reasons for losing the data are: the tapes and disks containing these data are stored carelessly and then simply "lost"; the principal researcher moves to another job and the project data are cleaned off the main computer's disks in order to make way for new data; the Codebook is either inadequate or non-existent and consequently when the principal researcher leaves nobody can remember what is in the computer data files; there are so many versions of the data files that it is impossible to know which data represent the "clean" files. The authors know of one country (that has a strong tradition for conducting well-designed, large-scale, and expensive national evaluation studies) in which only one of the datasets for the past four national testing programmes conducted for Grade 6 and 9 can be located.

Data analysis

The data analysis stage is mainly concerned with the preparation of usable summaries of the data that have been collected and prepared for analysis. These are usually, at least, in the form of descriptive statistics (for example, means, standard deviations) and cross tabulated frequency counts. In some cases there will be a need for significance tests (for example, "treatment versus control" comparisons), and/or tests of the fit of proposed models (for example, causal models based on path analyses). The data analyses for sample surveys must be preceded by the researcher deciding upon the sampling weights that are required, and then followed by the researcher calculating appropriate measures of sampling error for each estimated population parameter.

Throughout all of this work the temptation to allow the computer to take control of the analyses must be resisted. To achieve this, the researcher must "get his/her hands dirty with the data" by selecting a small sub-sample of cases and using these to replicate some of the computer analyses by hand. For example, using paper, pencil, and calculator, some simple descriptive statistics and frequency distributions should be prepared, "outliers" should be examined in detail, and "unusual" combinations of scores should be noted for further consideration.

The following discussion sets out the steps that should be followed in order to set the stage for a successful analysis of the data. No attempt has been made to advise upon the selection of particular data analysis techniques because these must be selected to fit in with the aims of the data collection and must also match the researcher's capacity to manage and interpret results produced by these techniques.

Data analysis: basic requirements

(a) *The reward structure.* The reward structure for data processors is a key factor in obtaining, and keeping, a qualified and experienced research team for one or more projects which may extend over several years. Talented data processors are few in number and they take a very long time to gain the kind of experience that is necessary to manage complex data collections in the field of education. The best way to keep these people interested in, and enthusiastic about, a long-term project is to offer an attractive "package" of working conditions and benefits which should include appropriate remuneration, access to computing facilities that are suited to the task at hand, and, in some cases, acknowledgement as a co-author of the project report.

(b) *The computer output should be designed to fit the requirements of the decision-makers before the data have been analysed.* The general format of the computer output should be prepared in draft form during the design of the data collection. That is, discussions should be held at a very early stage with the planners at the appropriate decision-making levels of the education system in order to establish sets of dummy tables. These discussions should be held mainly with those persons who will *use the data -- for decision-making purposes.* Care should be taken to limit the participation of "arm-chair sociologists" at this time because, in the authors' experience, such persons tend, often in a well-meaning way, to confuse and side-track the discussion. An attempt should be made, before the data preparation stage has been completed, to use some hypothetical data to fill out the dummy tables by hand. The data processing staff should then use these data to reproduce the hand-made tables on the computer for inspection and approval by the educational planners.

(c) *The construction of test and sub-test scores.* The preparation of test scores and sub-test scores should be accompanied by appropriate reliability and validity information. At the most minimal level for norm-referenced tests, a traditional item analysis should be undertaken in order to check that the items are "behaving" in an acceptable manner with respect to discrimination, difficulty level, and distractor performance. The reliability of each test and sub-test should be calculated and an attempt should be made to establish the validity of these where this has not been carried out in other projects. If the reliability of a test or sub-test fails to meet an acceptable level then consideration should be given to removing the test or sub-test concerned from the analyses.

(d) *Sampling errors and sampling weights.* The sampling errors and sampling weights should be constructed with the assistance of a sampling statistician in situations where the sample design deviates from simple random sampling by including complexities such as stratification, multiple stages of selection, or clustering (Ross, 1985). The importance of using the correct procedures to calculate the sampling errors has been mentioned previously in this paper. It is important to remember that sampling weights are usually always required, even for so-called "self-weighting" sample designs, because of imperfections in the sampling frame and/or the need to correct for non-response.

(e) *The computing equipment.* The advent of relatively inexpensive and very high-powered modern micro-computers now means that, except perhaps for some extremely large data collections such as student-level censuses using many variables, it is possible for the researcher to do most of the data processing on micro-computers. This situation has a number of advantages over the use of a centralized mainframe computer: it reduces "turn around" time for the writing and testing of computer runs, it allows the researcher to conduct the data processing operations without the need to fit in with the needs of other computer users, it encourages the researcher to stay in "stay in touch with the data", it provides the flexibility to work at times and on days when central mainframe machines might not be operating, it facilitates the sharing of data in electronic form on diskettes or via modem using telephone lines.

While the use of a personal computer has many advantages, there is one important disadvantage: the researcher now becomes responsible for making security backups of important data files. The best way to manage these backup procedures is to employ a "streaming tape" system that can be used in association with software that automatically, at pre-specified times and dates, transfers important datasets from hard disk storage to tape.

(f) *The software.* As a minimum requirement, the researcher should have access to a copy of one of the main-stream statistical packages that are now available for both personal computers and mainframe computers. Two of the more widely-used examples of these are the SPSS and SAS packages. These packages provide almost all of the data management and data analysis procedures required to analyse most small or large data collections, they are very well-documented and, because of the continual testing of them by a wide community of users, they are comparatively "bug-free". These packages often also include data entry software that permits the researcher to design project-specific data entry routines.

If possible, one of these packages should be supplemented by a general purpose item analysis programme to be used for the investigation of test and item characteristics, a spreadsheet programme to be used for the construction and manipulation of tables of figures, a word processing programme to be used for report preparation, and a full-screen editing programme that will permit the perusal and editing of data sent to the researcher on diskette.

Data analysis: what often happens

(a) *The data processing is never completed satisfactorily.* There have been many data collections carried out in the field of education that have contributed very little because of a failure to complete the data processing to a satisfactory standard. In the worst cases, failure takes the form of the analysis terminating at a very preliminary stage with the researcher caught in a tangled, and expensive, mess that has lost contact with the data analysis needs of the educational planners who were originally perceived as the clients for the whole exercise.

The authors' observations suggest that this kind of failure is generally associated with one of the following three causes: a lack of basic training and experience with respect to managing computer-based data analyses, an over reliance the "authority" of computer print-out at the expense of some personal familiarity with the data, or a fascination with computing technology that leads the researcher to go on-and-on playing around with "computing gymnastics" without finding the time to produce the analyses needed to write the kind of report that is meaningful and accessible to educational planners.

In other situations, failure occurs because of the the reward structures that are are associated with the component of the project concerned with data analysis and report writing. The authors have observed a number of talented research workers who, when the field work stage of a project is completed turn their attention to lobbying for the next opportunity to obtain control of a project - preferably an "externally funded" project that is likely to have high per diem and allowance rates. The data processing is then left in the hands of poorly trained and inexperienced junior staff, or in some cases in the hands of external contractors, who have little real interest in the purposes or importance of the project. This results in the data analysis being conducted in a superficial manner that rarely meets the requirements of the original project objectives. There are a number of countries in the world where this kind of behaviour has become institutionalized to such an extent that a whole generation of researchers now consider it to be "a way of life".

(b) *The researcher lacks fundamental knowledge and experience concerning the application of basic statistical and psychometric* *procedures.* There are many researchers who have received some formal training in basic statistical and psychometric theory. However, this training is often presented in a way that fails to establish links between the theory and the kinds of "real" data analysis questions that they are confronted with in their own countries. This kind of training offers a broad, but mostly superficial, knowledge structure that bypasses the need for a solid grounding in *applied* scaling, estimation, and model building. The problem is best illustrated by listing some examples of the types of errors that occur constantly in published research reports.

- *Measurement Errors.* The researcher reported the mean, standard deviation, and number of items for the total scores on a test but claimed that the reliability could not be calculated because only total test scores, and not individual item-level information, were entered onto the computer file. In fact, the most commonly presented estimate of test reliability, Kuder-Richardson formula 21 (KR-21) may be calculated using a little arithmetic in association with the three statistics mentioned above. The authors calculated the values of KR-21, by hand, for two key criterion test scores used by the researcher and found them to be so low as to suggest that the validities of the tests, and the sweeping generalizations made about them, were questionable.
- *The interpretation of test scores.* The researcher reported the results of a testing programme in which tests in school subject areas were administered to a sample of respondents. For each test, the average of the percentage of correct items per respondent was calculated. The researcher then discussed differences in these values across the tests as if they had some clear linkage to hypothetical knowledge domains. In fact, they were merely an artefact of differences in test difficulty. This mistake is often observed in analysis of national examination scores, where "pass rates" are set arbitrarily to match available places at the next level of education. In one country, national statistics on primary level completion scores were reported at the same level for several years, and interpreted as an indicator of the stability of the education system's quality. In fact, the pass rates were arbitrarily set, so that the same fixed percentage of students passed the test each year.

- *Reporting "differences" in mean scores.* The researcher reported the mean scores for subgroups of a sample on a multiple choice test of Reading. Statements were made about "differences" in performance levels between the groups. These statements were presented without any information being provided concerning the standard deviation of the test scores, the sampling errors of the test scores, or the reliability of the test scores.
- *Opportunities for creating composite scores.* Te researcher gathered many variables describing the socio-economic circumstances of students' home backgrounds and then reported some of these as univariate frequency distributions or in cross tabulations. There was no attempt made at constructing and employing one or more composite scores that might have summarized what were obviously replicated measurements of an underlying indicator of "socio-economic status".

(c) *Poor quality computing equipment is purchased and then operated in an inappropriate environment.* There is always a temptation, when research budgets are limited, to purchase the cheapest possible copies or "clones" of mainstream computing equipment, and then to cut corners in terms of the infrastructure required to house these computers. These kinds of "savings" are always illusory -- especially when a research unit needs to work effectively with large bodies of data and to be able to complete jobs according to tight deadlines. In many countries these problems are exacerbated by temperature and humidity extremes that do not suit the operation of fragile electronic equipment, by unpredictable electrical power supplies, and also by difficulties in obtaining spare parts and service when breakdowns do occur.

The authors' experienced difficulties in this area while working to an extremely tight schedule on a very large project concerned with the quality of primary schooling in Asia. The data processing for the project was conducted in a building that had no air-conditioning to control the large fluctuations of temperature and humidity. At the end of three weeks of heavy usage, one of the four personal computers allocated for the project had broken down, two others had developed intermittent and unpredictable processing errors, and after one more week the fourth machine was giving occasional read-write errors on its hard disk drive.

(d) *The researchers lack training in the use of some of the data manipulation procedures associated with the main statistical software packages.* Many educational researchers can operate the main statistical software packages, such as SPSS and SAS, at an elementary level. However, they have had no training in how to use this software to conduct a number of important exercises in data manipulation: sorting, merging, disaggregation, aggregation, etc.

The result of a lack of knowledge in this area is that student, teacher, and school datasets, are often analysed separately at a superficial level without merging datasets in order to explore the inter-relationships between all sources of data. The analysis of separate data files prevents the researcher from exploring research questions "across" the different sources of data. For example: What are the relationships between student achievement levels and teacher knowledge of subject matter after controlling for home, school, and community factors that might influence the educational environment?

(e) *Observational data or open-ended responses.* In any situation where observational data and/or open-ended test questions are used it is essential that, at least for a part of the data, some attempt is made at calculating the inter-rater reliability coefficient. This coefficient sometimes may be used to reveal that an excessive level of subjectivity has entered the scoring procedures, or that the raters have scored the responses along different dimensions by applying different criteria during the scoring. If the inter-rater reliability is very low it means that there is little, or no, agreement between raters with respect to the scores that they allocate for any single response and/or observation.

Often a failure to carry out these important checks on the stability of ratings is due to complete ignorance on the part of the researcher rather than a reluctance to carry out the small amount of extra work that is required. An example of this occurred where the researcher went to the expense of having two persons assess each piece of writing before coming to a decision on the rating to be given. Unfortunately, the researcher never considered taking the extra step of using these pairs of ratings to obtain a reliability estimate.

What might be done in the future

This chapter has explored some of the issues and problems associated with the sampling, collection, preparation, and analysis of data designed to assist educational planners to make decisions aimed at improving the quality of education. For each of these four activities, a discussion of the "basic requirements" was placed alongside some examples of problems that the authors have observed -- in both developed and developing countries.

While these problems covered a broad spectrum, the root causes of most of them may be grouped under two broad headings:

> *A lack of training and experience in the application of basic research procedures to "real" data.*
>
> and,
>
> *A lack of training and experience in research management techniques.*

There is a need for international agencies to take the initiative by preparing training programmes in these two areas. It is important that these training programmes should be flexible (with respect to sequence and length of presentation), and also capable of being constantly updated (with respect to content). Therefore, the design of these programmes must include a continuous formative evaluation process based on various sources of information that might, for example, include interviews with graduates of the programme and/or critiques prepared by successful practitioners. The results of these investigations could be used to extend and improve the training materials and, in some cases, form part of the pedagogical development of the programme. For example, some of the project reports prepared by the graduating trainees might eventually be included as "case studies" within the materials provided for later groups of trainees.

In the following section, the authors have listed a set of "guidelines" that they have developed for themselves based on their own successful *and* disastrous attempts at presenting a range of training programmes. This is then followed by some proposals for a training programme that attempts to satisfy most of these guidelines.

Authors' guidelines for training programmes

(a) Selecting and working with trainees

- Establish that the trainees are familiar with the language in which the training programme is to be presented.
- Select the trainees as "country teams" and ensure that all members of a team are actually working on the same, or a related, project.
- Ask each country team to appoint a "team leader" who will take responsibility for his/her team's overall behaviour and performance.
- Meet with, or at least speak with by telephone, the team leader's Senior Officer in order to explain fully the nature of the training programme, and also to reach agreement on the rights and responsibilities of each trainee. Confirm all of these matters in writing with this person.
- Avoid cultural, intellectual, and official blunders both within and across teams by becoming familiar with the following matters before the training commences: age and provincial origin (both very important in some parts of Asia and Africa), educational qualifications, previous research training/experience, status in terms of the "official hierarchy" (government rank) and the "unofficial hierarchy" (team rank).

(b) Process and content of the training programme

- Teach theory only in a "learning by doing" mode that requires the trainees to apply the theory as part of a "real" project.
- Keep interest and motivation at high levels by having the trainees work on a project that is nominated as "high priority" by the government of their own country.
- At the design stage of the project ensure that hypotheses and propositions are prepared in a form that permits clear policy directives to be made after the hypotheses and propositions have been subjected to empirical test.
- Visit the trainees in their own countries during the field work stage in order to provide assistance and/or intercept any problems.

(c) Outcomes of the programme

- Obtain a clearance from the participating countries before the commencement of the programme that the project will result in "knowledge made public" -- probably in the form of, at least, a short published research report prepared by each country team.
- Hold at least one seminar in each participating country in order to draw out the policy implications of the project results.
- Help the trainees to prepare a computer-based data archive.

Some suggestions for a training programme

(a) Programme content

Basic research skills. Throughout this chapter it has been pointed out that the practical skills and knowledge required to deal effectively with the logistics associated with the sampling, collection, preparation, and analysis of data are essentially those required in order to conduct educational research. There are many textbooks on educational research methods - but these books generally offer very little assistance for an educational planner faced with a "real" project in education.

The setting up of the training programme should therefore commence with a preliminary review of the basic practical skills and knowledge that are required to carry out this kind of work. This review would establish a "blueprint" for the design of the training programme. Flexibility in adapting to the needs of trainees from different backgrounds could be achieved by preparing the programme in the form of "stand alone" modules that could be combined in a variety of sequences of varying lengths. In the first instance, modules should be prepared to cover the areas of sample design, indicator specification, test and scale construction, data preparation, data analysis and reporting.

Research management. Training materials in the research management area should be prepared using the same strategy and format described above. In the first instance, modules should cover the areas of research design (including operationalization of research questions and planning the utilization of project resources according to a project timetable), the management of fieldwork (including the construction of manuals for the Test Administrator and the School

Co-ordinator), the management of data coding teams and data entry teams, the development of data archives, the selection and management of computing equipment and software.

(b) The format of the training.

The training programme should be centred around a "hands on" teaching approach in which the trainees would be expected to be already involved in, or about to be involved in, a project that requires the type of skills and knowledge addressed by the training programme. The training modules described above should be integrated with doing "real" projects nominated as "high priority" by the trainees' own countries.

The depth and scope of this training approach could not be covered in a "short course" of, say, two to four weeks. Such a course would not allow time for the trainees to apply their skills in a "learning by doing" mode. The length of the course should be arranged so as to be congruent with the the length of a reasonably substantial piece of educational research, say at least one year, and could be composed of the following four segments:-

- Two months in the home country engaged on pre-training tasks such as literature review, collecting statistics for sample design, etc.
- Three months outside the home country (or in a place far removed from the usual working environment) in order to plan the project, including the sample design and the development of specifications/first drafts for instrument construction.
- Six to eighteen months in the home country for instrument construction, trial testing, data collection, and data preparation.
- Three months outside the home country (or in a place far removed from the usual working environment) for data analysis and report writing.

The six to eighteen months spent in the home country could be supported by a visit from a member of the training team in order to review progress and, if necessary, arrange for any supplementary assistance that might be needed before the second three month training session.

(c) The selection of trainees

If the training programme is to be centred around the conduct of a project, it would be preferable for the trainees to be selected from among those persons who expect to be, or who already are, working on a project. Since these kinds of projects are usually worked on by a research team, it would be desirable for the training to be focused on "country teams" rather than individuals. That is, at least two persons from a research team in each participating country should attend the training programme. There are a range of supplementary benefits associated with a team approach to training. For example, a team approach provides a form of "insurance" against the project failing at the data collection stage (due to illness, job transfer, etc.), and it also gives the trainees a greater opportunity to discuss the ideas presented during the training in their own time and, in some cases, in their own language.

Conclusion

This chapter has highlighted some of the basic requirements, and some of the more common problems, associated with the sampling, collection, preparation, and analysis of data required by decision makers for planning the quality of education. It was argued that the majority of the problems that were described could be addressed by appropriate training programmes and, accordingly, some broad proposals were advanced concerning the design of one such programme. These proposals attempted to satisfy most of the "guidelines" that the authors had developed for themselves based on their own successful, and disastrous, experiences. It should go without saying that the detailed design and preparation of training programmes in this area will demand the active involvement of several potential clients so as to ensure that both the content and pedagogy of the programmes matches the requirements and learning styles of future trainees.