

# National assessments of educational achievement

T. Neville Postlethwaite  
and Thomas Kellaghan

9

Education policy

series



International Academy of Education

International Institute for Educational Planning



# The International Institute for Educational Planning



The International Institute for Educational Planning (IIEP) was established in Paris in 1963 by UNESCO, with initial financial help from the World Bank and the Ford Foundation. The French Government provided resources for the IIEP's building and equipment. In recent years the IIEP has been supported by UNESCO and a wide range of governments and agencies.

The IIEP is an integral part of UNESCO and undertakes research and training activities that address the main priorities within UNESCO's overall education programme. It enjoys intellectual and administrative autonomy, and operates according to its own special statutes. The IIEP has its own Governing Board, which decides the general orientation of the Institute's activities and approves its annual budget.

The IIEP's mission is capacity building in educational planning and management. To this end, the IIEP uses several strategies: training of educational planners and administrators; providing support to national training and research institutions; encouraging a favourable and supportive environment for educational change; and co-operating with countries in the design of their own educational policies and plans.

The Paris headquarters of the IIEP is headed by a Director, who is assisted by around 100 professional and supporting staff. However, this is only the nucleus of the Institute. Over the years, the IIEP has developed successful partnerships with regional and international networks of individuals and institutions – both in developed and developing countries. These networks support the Institute in its different training activities, and also provide opportunities for extending the reach of its research programmes.

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



# The International Academy of Education



The International Academy of Education (IAE) is a not-for-profit scientific association that promotes educational research, its dissemination, and the implementation of its implications. Founded in 1986, the Academy is dedicated to strengthening the contributions of research, solving critical educational problems throughout the world, and providing better communication among policy makers, researchers, and practitioners. The seat of the Academy is at the Royal Academy of Science, Literature and Arts in Brussels, Belgium, and its co-ordinating centre is at Curtin University of Technology in Perth, Australia.

The general aim of the Academy is to foster scholarly excellence in all fields of education. Towards this end, the Academy provides timely syntheses of research-based evidence of international importance. The Academy also provides critiques of research, its evidentiary basis, and its application to policy.

The members of the Board of Directors of the Academy are:

- Monique Boekaerts, Leiden University, The Netherlands (President)
- Barry Fraser, Curtin University of Technology, Australia (Executive Director)
- Erik De Corte, University of Leuven, Belgium
- Jere Brophy, Michigan State University, USA
- Eric Hanushek, Hoover Institute, Stanford, USA
- Denis Phillips, Stanford University, USA
- Maria de Ibarrola, National Polytechnical Institute, Mexico.

The following individuals are the members of the Editorial Committee for the Education Policy Booklet Series:

- Lorin Anderson, University of South Carolina, USA
- Eric Hanushek, Hoover Institute, Stanford University, USA
- T. Neville Postlethwaite, University of Hamburg, Germany
- Kenneth N. Ross, International Institute for Educational Planning (UNESCO), France
- Mark Bray, International Institute for Educational Planning (UNESCO), France.

# Preface

## Education Policy Series

The International Academy of Education and the International Institute for Educational Planning are jointly publishing the Education Policy Series. The purpose of the series is to summarize what is known, based on research, about selected policy issues in the field of education.

The series was designed for rapid consultation “on the run” by busy senior decision-makers in Ministries of Education. These people rarely have time to read lengthy research reports, to attend conferences and seminars, or to become engaged in extended scholarly debates with educational policy research specialists.

The booklets have been (a) focused on policy topics that the Academy considers to be of high priority across many Ministries of Education – in both developed and developing countries, (b) structured for clarity – containing an introductory overview, a research-based discussion of around ten key issues considered to be critical to the topic of the booklet, and references that provide supporting evidence and further reading related to the discussion of issues, (c) restricted in length – requiring around 30-45 minutes of reading time; and (d) sized to fit easily into a jacket pocket – providing opportunities for readily accessible consultation inside or outside the office.

The authors of the series were selected by the International Academy of Education because of their expertise concerning the booklet topics, and also because of their recognised ability to communicate complex research findings in a manner that can be readily understood and used for policy purposes.

The booklets will appear first in English, and shortly afterwards in other languages.

Four booklets will be published each year and made freely available for download from the websites of the International Academy of Education and the International Institute for Educational Planning. A limited printed edition will also be prepared shortly after electronic publication.

## This booklet

One of the key features of the worldwide discussion and debate concerning the need to achieve Education for All (EFA) has been a broader interpretation of this challenge to ensure that increased access to education is delivered in association with improvements in the conditions of schooling and student achievement levels.

This major focus on “quality” has been encouraged by the emerging understanding that education systems can act as pathways to national economic development in an increasingly globalized world.

Many nations have now established national assessment mechanisms with the aim of monitoring and evaluating the quality of their education systems across several time points. Some of these initiatives have been embedded within international assessments – such as TIMSS organized by the International Association for the Evaluation of Educational Achievement (IEA), the PISA Study organized by the Organization for Economic Cooperation and Development (OECD), and the Grade 6 Surveys organized by the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ).

This booklet examines the key issues and decision-points for ministries of education that are (or will soon be) involved in national assessments. It commences with a discussion of what and who should be measured in national assessments, and then proceeds to an examination of initial planning requirements and research implementation considerations related to the collection, analysis, and reporting of national assessment data. Finally the booklet proposes a framework for classifying types of educational policy and action that have been prompted by national assessments, and then concludes by mapping the common and unique aspects of national and international assessments.

## **T. Neville Postlethwaite** --- **(United Kingdom)**

*is Professor (Emeritus) of Comparative Education at the University of Hamburg, and a Fellow of the International Academy of Education. He holds a Doctoral degree in Comparative Education from Stockholm University, and is the author of many research reports and articles in the area of international comparative studies of the quality of education. He was co-editor of the first and second editions of the 12-volume International Encyclopedia of Education. He has been presented with a number of awards by universities and governments to honour his lifelong contributions to the advancement of educational research. He is currently working for UNESCO and the World Bank as an educational consultant in Africa and Asia.*

---

## **Thomas Kellaghan** **(Ireland)**

*is Director of the Educational Research Centre in Dublin. He holds a Doctoral degree in Psychology from Queen's University, Belfast, and is a Fellow of the International Academy of Education. His areas of research include assessment, programme evaluation, and educational disadvantage. He has worked on a wide range of educational research and development projects in Africa, Asia, and Latin America. He was co-editor of the International Handbook of Educational Evaluation.*

---

*This publication has been produced by the International Academy of Education (IAE) and the International Institute for Educational Planning (IIEP).*

*It may be freely reproduced and translated into other languages. Please send a copy of any publication that reproduces this text in whole or part to the IAE and the IIEP. This publication is available on Internet in its printed form, see: <http://www.unesco.org/iiep>*

*The author is responsible for the choice and presentation of the facts contained in this publication and for the opinions expressed therein which are not necessarily those of IIEP (UNESCO) and do not commit the Organization.*

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

*Jointly published by:*

*The International Institute for Educational Planning (IIEP)  
7-9 rue Eugène Delacroix  
75116 Paris  
France*

*and:*

*The International Academy of Education (IAE)  
Palais des Académies  
1, rue Ducale  
1000 Brussels  
Belgium*

*Design and layout by: Sabine Lebeau*

© UNESCO 2008

ISBN: 978-92-803-1321-5

# Table of contents

<b>1.</b>	What is a national assessment of educational achievement?	1
<b>2.</b>	Educational policy questions addressed by a national assessment	3
<b>3.</b>	Reporting student performance for a national assessment	5
<b>4.</b>	The frequency of national assessments	8
<b>5.</b>	Initial planning for a national assessment	11
<b>6.</b>	Collecting, analyzing, and reporting data for a national assessment	17
<b>7.</b>	Choosing an agency to conduct a national assessment	21
<b>8.</b>	Using data from a national assessment for policy purposes	23
<b>9.</b>	International assessments as national assessments	27
	References	31



information, national assessments can be considered as an essential component of the professional administration of any education system.

National assessments also provide important information to other stakeholders – teachers, parents, and the general public. Although it has been known for governments to suppress the results of national assessments because of anticipated embarrassment, the long-term advantages of an open information system (including increased public support for education and a stimulus for reform) are likely to outweigh any short-term disadvantages.

A national assessment can have wide-ranging implications for: (a) social and economic policy regarding the overall quality and performance of the education system, including its role in achieving social and economic objectives (for example: equality of opportunity, gender parity, and improving the performance of students from disadvantaged backgrounds); (b) the organization and management of an education system (for example: the provision of public and private education); and (c) learning conditions (for example: instructional time, resources, teacher education, and family support) (Husén, 1987).



- How well are students learning in the education system (with reference to general expectations, the aims of the curriculum, or preparation for life)?
- Is there evidence of particular strengths and weaknesses in students' knowledge and skills?
- Do certain sub-groups of students in the population perform poorly? For example, are there disparities between the achievements of: boys and girls, students in urban and rural locations, students from different language or ethnic groups, and students in different regions of the country?
- What factors are associated with student achievement? That is, to what extent does student achievement vary with the characteristics of the learning environment (for example: school resources, teacher preparation and competence, and type of school) or with students' home and community circumstances?
- Do the achievements of students change over time? And in particular, has student achievement improved, stayed the same, or declined in the time period covered by the introduction of important education reforms? (Kellaghan and Greaney, 2001; Kellaghan and Greaney, 2004).

## Reporting student performance for a national assessment

- 
- 
- ***Modern approaches to reporting student performance in terms of “learning hierarchies” provide more useful information about the precise knowledge and skills that have (or have not) been mastered by students***
- 
- 
- 
- 
- 
- 
- 

Similarities are evident in the subject matter studied in national assessments across the world. Most have been carried out at the primary-school level and nearly all have included an assessment of reading comprehension (in the language in which students are being instructed) and mathematics or numeracy achievement. Some have included science, which has either been based on general knowledge about the world in which students live or focused on topics in biology, chemistry, and physics. A few national assessments have included art, music, or civic education. At the secondary-school level, some interest has been shown in the skills of reading, writing, listening, and speaking a foreign language.

A ministry of education must decide what areas of student achievement to study and at what levels. The popularity of reading comprehension is based on its centrality in the educational process; if children cannot read, they cannot access textbooks in other curriculum areas.

It is important for a ministry of education to say how it wishes the results of a national assessment to be displayed. Some national assessments present results in terms of the percentage of test items (questions) that were correctly answered by students, but this provides no indication of

what students can and cannot do. Alternatively, results are sometimes presented as a distribution of student test scores (often standardized to a mean of 500 and a standard deviation of 100). Again, this approach provides no indication of what students can and cannot do.

A more accessible and meaningful approach for displaying national assessment results is to describe student performance in terms of a hierarchy of skills – so that it is very clear what students can and cannot do. An example of this approach has been presented below for mathematics from the Laos 2006 Grade 5 National Assessment Survey (Sisouk & Postlethwaite, 2007).

It should be noted that the six levels in the table are hierarchical. That is, students at Level 1 would be unlikely to be able to answer correctly items associated with the skills shown at higher levels. Those at Level 2 would be likely to be able to do what was required at Levels 1 and 2, but not at Levels 3 to 6. When presented with such information, ministry personnel and other stakeholders can immediately see what students can and cannot do.

Some countries also test the teachers of the students. This was undertaken in the Vietnam Grade 5 Survey in 2001 (World Bank, 2004) and also the international surveys of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) (Nzomo & Makuwa, 2006). Testing teachers can be used to identify gaps in their knowledge about school subject matter – which, in turn, can help to point to areas that can be addressed in in-service teacher training.

Level	Mathematics skill levels	% of students at level	SE*
1	May recognise and classify basic shapes. Familiarity with numbers described in word and numeric form. Understanding of place value for whole numbers.	16.3	1.0
2	Emerging ability to perform single-step arithmetic operations including addition, subtraction, multiplication, and division. Recognises fractions in both numeric and visual representations. Some concept of symmetry emerging.	49.1	1.1
3	Emergence of arithmetic applied to problem solving. Multiple-step arithmetic operations. Understanding place value for decimals. May read a value from a simple bar graph. Familiarity with inequalities and ability to order decimal numbers by magnitude. Developing understanding of proportional fractions. Basic conversion of linear units such as length, weight, and time.	19.7	0.8
4	Developing the ability to solve word problems requiring a fraction or percentage operation. Developing more sophisticated arithmetic including BODMAS, long division, and multiplication with decimals. Conversion between various units of weight, time, and volume. Deals with elementary spatial problems involving 2-dimensional displacement.	10.3	0.7
5	Beginning to combine and summarise multiple pieces of information from charts. Developing an understanding of spatial concepts such as rotation and reflection. Conversion of units for weight, time, area, and volume.	3.6	0.5
6	Use of rules and symmetry to solve geometric and measurement problems. Strong command of unit conversion for everyday measures. Capacity to solve word problems using a range of appropriate arithmetic operations.	1.0	0.2

\* SE = Standard Error of Sampling

# 4 The frequency of national assessments

- 
- 
- ***Ministry of education officials need to decide on the frequency with which they wish to conduct national assessments. This decision will depend on several factors.***
- 
- 
- 
- 
- 
- 
- 
- 

In some countries, a national assessment in a particular curriculum area is carried out every year. This is the case in England where the assessment is part of an official annual accountability system. It is also the case in Uganda where assessment instruments are administered every year, but background data about students are collected only every three years.

Decisions concerning how frequently to conduct a national assessment depend upon several factors:

## **a. The intended use of the results**

In some countries, national assessment data are used to determine if achievement has improved, remained constant, or deteriorated at national (and perhaps regional) levels over a period of time. In such a case, data collected every five years from a sample of schools would seem sufficient since education systems change slowly. In other countries, a major political concern is to hold schools accountable and to provide parents with information that they can use to make a choice of schools for their children. In this case, it is quite common to conduct an assessment every school year for all schools, and to ensure that the results are widely publicised.

## **b. How much money is available**

National assessment studies can cost a lot of money and, if a ministry wishes to have regional as well as national estimates of achievement, the costs can escalate. To illustrate: During 2001 in a South East Asian country where separate estimates of the conditions of schooling and the quality of education were required for each of 60 provinces, the data collection cost the equivalent of 500,000 U.S. dollars. However, during 2007 in an African country with only four provinces the data collection was carried out for the equivalent of 30,000 to 50,000 U.S. dollars. When all students in an education system are tested, as in England, the costs can be very high and therefore such “census-based” approaches may not be affordable in poorer countries.

In countries where many schools have been built in relatively remote and inaccessible locations, the cost of data collections can be high. In extreme cases, it can actually cost half of the assessment budget just to conduct a data collection in remote schools. In this case, students in more remote areas may be excluded from the target population before schools are selected for the data collection.

## **c. The linkage of a national assessment to an international assessment**

In the international study organised by the Organisation for Economic Co-operation and Development (OECD) known as PISA (Programme for International Student Assessment), the cycle of data collection is every three years. This cycle length was selected because if the cycle length was longer, the national research teams might be disbanded, which would necessitate a lot of time and work to build and train new teams.

## **d. The demand for assessment information**

There must be a political will to have the kind of information that emanates from assessment studies. A change in government, in which the incoming government does not accord as high a priority to the collection of assessment data

as the outgoing government, may result in a reduction in the frequency of assessments.

**e. The time required to fully train a multi-disciplinary staff to conduct an assessment**

This can affect the timing of the first assessment and the interval before the next survey. In general, it is suggested that a 5-year cycle is sufficient.

## Initial planning for a national assessment

- 
- 
- ***National assessments must begin with detailed planning that will ensure that data collections are: relevant, useful, based on scientific sampling procedures, and generated from valid data collection instruments.***
- 
- 
- 
- 
- 
- 
- 

It has occurred that, when the results of a national assessment have been published, top ministry of education officials have asked questions such as “Do students in smaller classes really achieve less than students in larger classes?” or “What is the optimal size of a primary school for high achievement and good behaviour?” Then it is discovered that no questions were asked about class size or school size and no measures of student behaviour were included in the survey instruments, thus making it impossible to answer the officials’ questions.

The lessons to be learned are that the exact research questions for a study have to be planned from the very beginning, and that care must be taken to ensure that the information required to answer these questions is actually collected in the national assessment. This may sound like common sense, but many national assessments have been undertaken in which ministry of education officials were not involved in stating the research questions, and many others have collected large amounts of data that were not focused on the concerns of officials.

In general, the planning of a national assessment should start at least two years before the main data collection.

There are eight main steps involved in the initial planning for a national assessment. Each of these steps needs to be addressed in a systematic fashion.

### **a. Decide on the purposes of the assessment**

The purposes should be expressed in the research questions to be answered, and from these questions it should be possible to plan the data tabulations in the form of blank tables (sometimes called dummy tables) to be completed when the assessment has been completed.

Senior ministry of education officials and the researchers conducting the assessment should meet to discuss the research questions so that the officials can be certain that the study will provide the information that they require, and so that it is crystal clear to the researchers what they have to do. The research questions will almost certainly require information about the achievements of students, and also information about the factors in the home and in the school that might affect achievement. This exercise takes time, and since top officials are often very busy people, time must be scheduled well in advance.

### **b. Construct the test(s)**

Test construction involves developing test blueprint(s), writing test items, trying them out, conducting item analyses, and assembling the final test(s). A test blueprint describes content domains and skills levels of the actual construct (for example, reading) or knowledge of subject-matter (for example, the science curriculum for sixth grade) that is to be assessed. It is important to ensure that a test does what it is intended to, that is, that it measures accurately the students' knowledge and skills in the areas identified in the test blueprint. This means that it should contain a representative set of tasks covering the knowledge, skills, and strategies needed for the activity or domain being assessed (Bloom, Hastings and Madaus, 1971). The judgment of curriculum specialists is important for this purpose. Tasks should be attuned to the level of competence of the students being assessed. Often, the tasks in national assessments

in developing countries are based on an idealized view of achievement and fail to take adequate account of students' current level of achievement or of the conditions in which learning takes place.

### **c. Decide on educational indicators**

Decide on educational indicators for students, teachers, and schools, the variables they involve, and the questions that will measure each variable, and then produce the required questionnaires. The preparation of questionnaires begins with the listing of all indicators in the dummy tables. Next to each indicator, a list of the variables required to form the indicator can be presented. Next to each variable, the questions needed to obtain the information required to form the variable can be listed. There is no easy way to construct questionnaires. Several trials, or pilots need to be built into the overall timetable in order to produce a satisfactory questionnaire.

### **d. Define the target population**

A decision about the target population should involve ministry of education officials in co-operation with researchers. At the outset, there is a need to decide whether to cover all students in the "desired target population" or to permit some students to be excluded in order to focus on a more restricted "defined target population". Exclusions might cover students in special schools for children with handicapping conditions, or students in certain provinces that are difficult or dangerous to access. In general, it is usually not acceptable to exclude more than 5 percent of "the desired target population".

### **e. Ensure that a good sampling frame exists**

A sampling frame is a list of all schools in the country that contain students in the defined target population. The list should include the names of the schools, their identification numbers, and the numbers of students in the defined target population. Supplementary information (such as urban/rural, public/private, and regions, etc.) can also be included

for stratification purposes. Such lists are normally produced via the Annual School Census conducted by the Education Management Information System (EMIS) Division in a ministry of education. It is, however, surprising how many ministries do not have up-to-date, clear, and accurate lists. It is for senior ministry officials to ensure that such lists are accurate and available. The research questions should be decided before the sampling frame is prepared – just in case extra variables need to be included in the sampling frame. For example, if comparisons are needed between schools serving wealthy and poor communities, then the sampling frame should include a variable that provides a measure of the socioeconomic background of the students.

#### **f. Decide on where “over-sampling” is required and then draw the sample**

Normally, ministry of education officials want estimates of levels and variations in student achievement in each province, state, or region in the country, since this information can help identify parts of the country where greater effort is needed to improve achievement levels. If accurate estimates are required for separate sectors of the education system, then the number of schools in the sample may need to be quite large because small sectors may need to be “over-sampled”. For example, if only 2 percent of students are in private schools, a comparison of student achievement in private and public schools may require the sample size for private school students to be a much higher percentage of the total sample of students. Or if the number of students who attend schools that are attached to secondary schools is small, it may be necessary to select a disproportionately larger sample of students in such schools in order to answer the question: Do students perform better in primary schools that are attached to secondary schools? When students from a sub-group are over-sampled, weighting procedures have to be used when calculating the overall results for the country so as to adjust for differences in the probabilities of selection for different sample members.

Sampling is not a simple matter. A sampling statistician is required. Since it is unusual for a ministry of education to

have a full-time sampling statistician, it has to ensure that one is found who understands the purposes of the study and who will be available at the required time. After the main data have been collected, this person will also be responsible for the calculation of sampling weights and the standard errors of sampling.

### **g. Trial the test items, questionnaires, and procedures**

It is essential that all test and questionnaire items are tried out on a judgment sample of students which spans the variation in schools in the target population. Researchers should be skilled in classical and Rasch item analysis. A Rasch analysis is needed: to examine the technical performance of items used in the trial, to equate scores over time, and to examine the differential functioning of items (DIF) for sub-groups of students (for example, boy/girl, urban/rural, language group membership) so as to ensure that items are fair to all subgroups in the study.<sup>1</sup>

Ministries must ensure that the researchers are competent in the relevant measurement techniques well before beginning the assessment. The procedures to be used in the main testing can also be field-tested at this stage.

### **h. Plan the data collection and train data collectors**

The data collection may involve a lot of schools (for example, there were 3,660 schools in the 2001 Vietnam Grade 5 Survey; World Bank, 2004), or relatively few (about 100) if schools within a country do not differ much from each other. In any national assessment exercise it is important that student non-response to test and questionnaire items of schools and students be minimized and the amount of missing data virtually eliminated. This requires very careful

---

1. Differential Item Functioning (DIF) analysis is undertaken to determine if there are differences between population subgroups of *equivalent overall ability* in their performance on particular items (for example, boys/girls, urban/rural students). Where there is significant difference, the item is said to exhibit DIF. Efforts to explain the origin of DIF as either *impact* or *bias* should then be made. Impact describes DIF arising from advantageous or impeding subgroup characteristics directly associated with the measurement variable. *Bias* describes DIF arising from factors unrelated to the measurement variable.

data collection. It can also require several data collectors in every school. For example, in the SACMEQ III Project data collection during 2007 (see [www.sacmeq.org](http://www.sacmeq.org)) there were three data collectors per school, and data collection in each school lasted two days. A lot of training of data collectors, which can last up to four days, may be required. The manuals for data collection must be written well in advance and checked and rechecked. The writing of manuals is complex, and researchers with experience are needed for the task. All of this must be paid for and, if the Ministry wants a study with good data, the training and per diems for data collectors to stay at or near the selected schools must be provided.

# 6 Collecting, analyzing, and reporting data for a national assessment

*The research steps involved in the main data collection for a national assessment usually proceed according to the following sequence: field work, data preparation, data analysis, and reporting results.*

The main research component of a national assessment should proceed as a sequence of four connected steps: “field work” – where data are collected in a manner that will maximize validity and response rates; “data preparation” – which includes computer-based cleaning, management, and merging of data files so that they are ready for analysis; “data analysis” – which concentrates on the construction of new variables and the generation of tabulations; and “reporting results” – where the aim is to produce different types of report for different target audiences.

## **a. Field work**

If a sample of students needs to be selected within sampled schools then the data collectors need to be able to apply strict random sampling procedures according to a pre-prepared set of mechanical rules. On no account must teachers or school principals be allowed to select the students – because they always select the best students! All questionnaires must be checked for missing data and consistency of responses to ensure that all required data are provided. Often, a small group of “evaluators” visits a selection of schools to ensure that the data collection has been satisfactory. Again, this has to be organised and paid for. It must be emphasised that

the data collectors must visit the sampled schools. Some ministries of education have been known to mail tests and questionnaires to schools, but this has proved to be a very bad idea. The wrong students are often tested, there is a lot of missing data, and the response rate is low.

To reduce the amount of missing data, it is often the case that when the data collectors take the completed tests and questionnaires from a school to the ministry of education district offices, the instruments are checked once again to see if the data are complete, and if they are not, the data collectors are sent back to the relevant schools to collect the missing data.

### **b. Data preparation**

The data have first to be entered into a computer-stored database. There are programs that are available for this. One program that is widely used and can be adapted more or less for any data set is WinDEM (Windows Data Entry Manager). It is usually a good plan to conduct double data entry (in which the data for each data collection instrument are entered twice into computers and then compared electronically) because this results in greater accuracy. After data entry, the data files must be cleaned and weighted, and school files, teacher files, and student files have to be merged to produce a large master file. All of these steps require considerable technical expertise and time. If the data entry has not been well done, the cleaning can take a long time, as much as one year. For the weighting of data it is usually wise to have the assistance of a sampling statistician (the person who drew the sample). Ministry of education officials should ensure that the researchers have been trained in WinDEM, in data cleaning, and in file merging before the study begins.

### **c. Data analysis**

Often the data for some variables have to be recoded, and single variables have to be combined to form new derived variables. For example, an Index of Socioeconomic Background will need to be constructed for each student by

combining data about the characteristics of the students' parents (such as education levels) and the characteristics of the students' homes (such as the number of books in the home). All of this takes time. When complete, the data analyses can begin. Normally, the first step is to run those analyses that are required to complete the blank or dummy tables mentioned above. As the data analyses proceed, clues to the formulation of new research questions or hypotheses will usually arise, in which case new tables will need to be drawn up, and appropriate analyses carried out.

To get through all these steps, the research team carrying out the national assessment needs to have appropriate computers and software. It is also essential that they have the required statistical knowledge and experience with statistical software. All of this sounds self-evident, but administrators are often unaware that up-to-date hardware and software are needed to manage and analyse large complex datasets. Their ignorance has often resulted in projects not being able to run certain analyses and in delays to projects. Such ignorance can also exist in some research centres where the director is more an administrator than a researcher.

#### **d. Reporting results**

A ministry of education should decide on the kinds of report that will be required. One detailed report could be designed for reference in future assessments and to meet the needs of other educational researchers. Another could be shorter and less technical for a wider audience. A less technical report with practical implications for teachers is also often prepared. One could also prepare media releases or make television or radio programmes showing major results.

The reports should highlight strengths and weaknesses in achievement in the nation as a whole, as well as in regions or provinces. It is of the utmost importance that an attempt be made to identify the reasons for achievement differences or for wide variations in achievement levels, though care should be taken in formulating causal explanations when data have been obtained in a cross-sectional survey. Analyses should,

however, give rise to suggestions about what can be done at national, provincial, and school levels to improve the situation. The ability to do this will depend on the quality of background information that has been collected in the study.

Suggestions for improving an education system are usually made in a summary chapter in the national report, with estimates of the cost and time associated with implementing each suggestion. Researchers alone can rarely make these estimates; usually a joint effort of researchers and appropriate ministry personnel is required. Time must be built into the project to allow this to happen.



selected agency has the required technical competence. The question arises: Does the ministry have the knowledge among its staff that allows it to judge whether the technical expertise exists in a particular agency? If not, how will it judge? Much depends on where the technical expertise in the country is to be found.

In several countries (for example, Bhutan, Ethiopia, the Gambia, Kenya, Lesotho, Uganda, and Zambia), the administration of national assessments has been assigned to the body with responsibility for public (external) examinations. This body may be part of a ministry of education or it may have a separate legal status, though subject to the general policy of the ministry. Examination bodies have been asked to carry out national assessments mainly because of their experience in test development and their capacity to manage the logistics of a large-scale testing programme. In fact, the traditions of test development for national examinations differ from the requirements from national assessments. In addition, examination bodies often lack expertise in a number of the areas that are essential in a national assessment, in particular sampling and statistical analysis. To address this issue, the services of consultants (for sampling) and of a university statistics department (for analysis) may be obtained.

In other countries dedicated research organizations or units affiliated to universities have been employed to administer national assessments (for example, in Chile, Ireland, and Sri Lanka). Such bodies have been chosen because of the expertise that they possess (test development, sampling, statistical analysis), or are in a position to acquire. They are also likely to have the computer facilities and space that a large-scale survey requires.

# 8 Using data from a national assessment for policy purposes

- 
- 
- ***Reviews of the situations where national assessments have had an impact on educational policy and practice suggest that these situations may be classified according to nine categories of action.***
- 
- 
- 
- 
- 
- 
- 
- 
- 

In each of the SACMEQ national assessment reports, the final chapter is devoted to suggestions for improving the education system (see, for example, Nkamba & Kanyika, 1998). The aim here was to suggest actions to be taken on the basis of research results and then to classify these according to which person or group was responsible for taking action and their timeframes and estimated execution costs. However, there is evidence from many countries that the information derived from national assessments are often not being fully exploited and therefore have had limited impact on policy formation related to improving the quality of student learning (see, for example, Ferrer, 2006). This is due, at least in part, to the fact that while the need for indicators of student achievement in assessing the quality of education may be recognised, little consideration has been given to how exactly the results of assessments can be used to inform policy or to the strategies that would be required to effect an improvement in student learning.

The situations where national assessments have resulted in an impact on policy and practice do not appear to be associated with similar contexts (see Arregui & McLauchlan, 2005). It has been observed that nine categories of action have been prompted by national assessment results.

### **a. General analyses of the education system**

In these situations, the dissemination of national assessment results has sometimes prompted a general debate concerning the condition of education. The response in many developing countries (for example, Ethiopia, the Gambia, and Ghana) has often been one of shock. The shock is likely to be reinforced if the country has also taken part in an international study (for example, in Ghana which participated in TIMSS) that has provided unwelcome comparisons with other countries.

### **b. Inputs to a general review of policy**

National assessment results have sometimes provided inputs to a national sector study or a national commission (for example, in Mauritius, Sri Lanka, and Zambia). The inputs may have been at a national level or at a regional level.

### **c. Preparation of a reform programme**

A step beyond providing data for a review is to design a reform programme on the basis of national assessment results. In one country (England), a numeracy task force was set up and this, in turn, led to the development of a national numeracy strategy. In another country (Guinea), what were considered to be poor national assessment results prompted a national reading initiative.

### **d. Dissemination and discussion of results with stakeholders**

Practically all countries make some effort to disseminate the results of a national assessment. Activities include sending reports to schools, discussions with teacher unions, and school inspectors, national and regional seminars, and establishing teacher networks to review programmes or to develop strategies to improve students' achievements. Dissemination may involve the use of analytical documents, videos, and workshop material. Procedures to address identified problems may be devised and implemented at a local level. It is critical that parents are informed on a regular

basis of the results of national assessments, as it is often through informed parent pressure that reform begins.

#### **e. Reform of the intended curriculum**

National assessment results have led to a number of initiatives to reform intended curricula. Usually these have operated at the national level, and have included defining standards or benchmarks – both for student achievement and the provision of resources (for example, Kenya, Vietnam), and have occasionally precipitated a shift in emphasis in curriculum content (for example, Ireland). Another initiative involved an increase in emphasis on mathematics education in teacher education programmes (Norway).

#### **f. Reform of the implemented curriculum**

Here the focus is on the classroom in terms of how teachers are implementing the curriculum. Policy decisions have been taken to provide additional resources for teacher development (pre-service and in-service) with the ultimate aim of affecting how teachers deliver the curriculum (for example, Norway, Paraguay, Peru, Uruguay).

#### **g. Improvement and reallocation of resources**

This involves increasing resources in poorly performing schools (by providing, for example, additional finances, textbooks, classroom libraries, pedagogical materials, or teacher development), and has been a feature of many national assessments (for example, Chile, Uruguay). The reallocation of resources is most feasible following a census-based assessment (that is, when students in all or most schools are assessed) in which it is possible to identify individual schools that are performing poorly. However, it also occurs when resources are provided to a particular category of schools (for example, schools serving isolated or socioeconomically disadvantaged populations) which have been identified as performing poorly in a sample-based national assessment.

## **h. Monitoring the effects of an intervention**

When interventions coincide with a national assessment, the results have been used to determine if student achievement has been affected (for example, following the introduction of a school voucher system in Colombia; and following the expansion of educational provision for Universal Primary Education programmes in African countries).

## **i. Helping donors identify aspects of the education system in need of support**

This may involve the use of national assessment data for the identification of a general curriculum weakness, or to target students in particular circumstances who are in need of special attention.

In summary, it may be said that the use of national assessment results has been varied, ranging from use in policy reviews and the establishment of benchmarks to very specific actions such as the provision of resources to schools and in-service programmes for teachers. What we cannot say, however, is what the long-term impact of the initiatives has been on the quality of student learning.

## International assessments as national assessments

*International assessments of student achievement share many procedural features with national assessments in terms of field work, data preparation, data analysis, and reporting results. However, international assessments pose special challenges in terms of the construction of valid cross-national data collection instruments.*

International assessments of student achievement share many procedural features with national assessments. They are designed to describe and compare the level of student educational achievement for a group of education systems (Beaton, Postlethwaite, Ross, Spearritt, & Wolf, 1999; Husén & Postlethwaite, 1996; Kellaghan & Greaney, 2001). In international assessments a great deal of care is required in the design and development of data collection instruments so as to ensure that differences in language and culture do not prevent researchers from making valid cross-national comparisons.

Participation in an international assessment has a number of practical advantages. First, these initiatives are usually established and managed by teams of experienced research specialists who can assist individual countries that may not have access to trained personnel. Second, the data processing in international studies is usually located in a single institution that has all of the required human and

material resources to manage (on behalf of the participating countries) all the aspects of the data cleaning, data merging, and statistical analysis. Third, staffing requirements and costs may be lower than in a national assessment since costs for many aspects of the design and development of data collection instruments, field manuals, and related research tools are shared among countries.

An important aspect of an international assessment is that it can provide data for individual countries to carry out their own within-country analyses in what becomes, in effect, a national assessment. Indeed, many countries participate in international studies to have good national studies. The national dimension is reinforced in studies such as PISA and SACMEQ in the expectation that each participating country will produce a national report.

The value of an international study is optimised for national-level analyses if variables of national interest are added, a practice usually allowed in international studies. For example, in PISA 2006, 35 of 52 participating countries included “national variables” in the student questionnaire, and 18 in the school questionnaire. In approximately a third of cases, however, these “national variables” were simple adaptations to core items (such as parent education) rather than entirely new variables (J. Cosgrove, personal communication, 22 May 2007). PISA also allows for extension of samples. A grade-based national option is now used by roughly half of the OECD countries, while some countries with federal systems conduct PISA on a regional basis (A. Schleicher, personal communication, 30 April 2007).

A number of limitations of international assessments have been identified. Perhaps the most obvious is that a test designed to serve as a common denominator of curricula in many countries is not going to provide as valid a measure of curriculum mastery for every individual country as tests which take account only of the curriculum of individual countries, as is the case in a national assessment. It may be noted that the task of devising a common instrument is more difficult in some areas (for example, science and social

studies) than in others (for example, reading). Modern developments in educational measurement have helped to address these issues. It is now possible to use “rotated test forms” whereby students in the sampled schools receive “different but overlapped” test booklets containing a mixture of unique and common test items. This allows researchers to collect data on a much larger number of test items so as to improve curriculum coverage, and at the same time score students as if they had all completed the same test.

The problem of curriculum coverage is not so important with studies such as OECD’s PISA Project because PISA assesses “literacy” (in reading, mathematics, and science) of 15-year-old students not in relation to the curricula to which students have been exposed, but rather in relation to expert opinion concerning the knowledge and skills that students will need for full participation in society. This approach gives rise to two problems. First, it is very difficult to determine what knowledge and skills students are going to need, given the rapidity and unpredictability of change in all areas of life in the 21st century. Second, since countries vary in their social and economic systems, it seems unlikely that the required knowledge and skills will be the same in all countries.

A particular problem arises when developing countries participate in international studies which have been designed for industrialised countries. Experience indicates that the tests are too difficult and fail to discriminate adequately between the achievements of students in the less developed countries (see Naumann, 2005).

If a country wishes to obtain achievement data that will permit comparisons with other countries, a regional assessment would seem more appropriate than a world-wide study designed by and for industrialised countries. A number of studies exist in which countries in a region that share many socioeconomic and cultural features collaborate: SACMEQ in Southern and Eastern Africa (see Nzomo & Makuwa, 2006); PASEC in West Africa (see Bernard & Michaelowa, 2006); and LLECE in Latin America (see Casassus et al., 1998).

At the same time, since each country is, or wants to be, part of the globalised economic system, and is in competition with other nations, it may want to know where its human capital stands relative to a wider range of countries. To address this issue, some countries have added items used in an international assessment to their national assessment, allowing them to link their test results to those of other nations.

Before deciding to participate in an international study, policy makers, in addition to determining if the assessment instruments will provide a fair coverage of their school curricula, should satisfy themselves that the purpose of the study is clear, that it will address major policy concerns, that technical aspects of the study are satisfactory, and that adequate contextual information will be provided (Kellaghan, 1996; Postlethwaite, 2006).

## References

- Arregui, P., & McLauchlan, C. (2005). *Utilization of large-scale assessment results in Latin America*. Report prepared for the Partnership for Educational Revitalization in the Americas (PREAL) and the World Bank Institute.
- Beaton, A.E., Postlethwaite, T.N., Ross, K.N., Spearritt, D., & Wolf, R.M. (1999). *The benefits and limitations of international educational achievement studies*. Paris: International Institute for Educational Planning/ International Academy of Education.
- Bernard, J.M., & Michaelowa, K. (2006). How can countries use cross-national research results to address 'the big policy issues'? (Case studies from Francophone Africa). In K.N. Ross and I. Jürgens Genevois (Eds.), *Cross-national studies of the quality of education. Planning their design and managing their impact* (pp. 229-240). Paris: International Institute for Educational Planning.
- Bloom, B.S., Hastings, J.T., & Madaus, G.F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw Hill.
- Casassus, J., Froemel, J.E., Palafox, J.C., & Cusato, S. (1998). *First International Comparative Study of Language, Mathematics, and Associated Factors in Third and Fourth Grades*. Santiago, Chile: UNESCO.
- Ferrer, G. (2006). *Educational assessment systems in Latin America: Current practice and future challenges*. Washington DC: Partnership for Educational Revitalization in the Americas (PREAL).
- Husén, T. (1987). Policy impact of IEA research. *Comparative Education Review*, 31, 29-46.
- Husén, T., & Postlethwaite, T.N. (1996). A brief history of the International Association for the Evaluation of Educational Achievement. *Assessment in Education*, 3, 129-141.

- Kellaghan, T. (1996). IEA studies and educational policy. *Assessment in Education*, 3, 143-160.
- Kellaghan, T., & Greaney, V. (2001). *Using assessment to improve the quality of education*. Paris: International Institute for Educational Planning.
- Kellaghan, T., & Greaney, V. (2004). *Assessing student learning in Africa*. Washington DC: World Bank.
- Naumann, J. (2005). TIMSS, PISA, PIRLS and low achievement in world society. *Prospects*, 35, 228-248.
- Nkamba, M., & Kanyika, J. (1998). *The quality of education: Some policy suggestions based on a survey of schools. Zambia*. Lusaka: Ministry of Education; Paris: International Institute for Educational Planning.
- Nzomo, J., & Makuwa, D. (2006). How can countries move from cross-national research results to dissemination, and then to policy reform? In K.N. Ross and I. Jürgens Genevois (Eds.), *Cross-national studies of the quality of education. Planning their design and managing their impact* (pp. 213-228). Paris: International Institute for Educational Planning.
- Postlethwaite, N. (2006). What is a 'good' cross-national study? In K.N. Ross and I. Jürgens Genevois (Eds.), *Cross-national studies of the quality of education. Planning their design and managing their impact* (pp. 105-120). Paris: International Institute for Educational Planning.
- Sisouk, V., & Postlethwaite, T.N. (2007). *Laos grade 5 national assessment*. Vientiane, Laos: Ministry of Education.
- World Bank. (2004). *Vietnam: Reading and mathematics study* (3 vols). Washington DC: Author.