DOCUMENT RESUME

ED 133 853 EA 009 107

AUTHOR	Husen, T.
TITLE	International Evaluation of Educational Systems and Its Role of Building Research Competence in Less-Developed Countries. IIEP Seminar Paper: 7.
INSTITUTION	United Nations Educational, Scientific, and Cultural Organization, Paris (France). International Inst. for Educational Planning.
PUB DATE	75
NOTE	26p.; A contribution to the IIEP Seminar on "The Evaluation of the Qualitative Aspects of Education" (September 30-October 4, 1974) ; For related documents, see EA 009 106-111 and EA 009 113-116
AVAILABLE FROM	IIEP Publications, 7-9 rue Eugene-Delacroix, 75016 Paris, France .(3.00 francs)
EDRS PRICE	MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	Academic Achievement; *Achievement Tests; *Developed Nations; *Developing Nations; *Educational Assessment; Elementary Secondary Education; *Evaluation Methods; International Programs; Reading Achievement; *Research Methodology; Sciences; Student Attitudes; Test Construction
IDENTIFIERS	IEA; *International Assn Evaluation Educ Achievement

ABSTRACT

The first three reports in the Six Subject Survey conducted by the International Association for the Evaluation of Educational Achievement reported cognitive and affective outcomes of school education in science, literature, and reading comprehension for students at the 10-, 14-, and 18-year-old levels in over 20 countries, four of which were less developed. The discussion of international evaluation takes place against this background. Discussion is given to the purposes of international surveys of educational systems, misgivings about the appropriateness of employing international evaluation standards, the organization of the international evaluation effort, the mean student performance in science and reading in industrialized and nonindustrialized countries, and the establishment of research competence in education in less-developed countries. (Author/IRT)

IIEP seminar paper: (7)

INTERNATIONAL EVALUATION OF EDUCATIONAL SYSTEMS AND ITS ROLE OF BUILDING RESEARCH COMPETENCE IN LESS-DEVELOPED COUNTRIES

T. Husén

A contribution to the IIEP Seminar on "The evaluation of the qualitative aspects of education"
30 September - 4 October 1974

INTERNATIONAL INSTITUTE
FOR EDUCATIONAL PLANNING
(established by Unesco)
7-9, rue Eugène-Delacroix, 75016 Paris

The opinions expressed in these papers are those of the authors and do not necessarily represent the views of the Institute or of Unesco.

CONTENTS

1. Purposes of international surveys of educational systems

In 1973 the first three reports from the so-called Six Subject Survey conducted by the International Association for the Evaluation of Educational Achievements (IEA) were published (Comber and Keeves, 1973; Purves, 1973; Thorndike, 1973). They reported cognitive and affective outcomes of school education in science, literature, and reading comprehension for students at the 10-, 14-, and 18- year-old levels in some twenty countries. Most of these countries were highly industrialized, but four less developed countries (LDC's) also participated. Three more subject areas will be reported in the near future, namely English and French as foreign languages and civic education.

The evaluation was carried out by giving representative samples of students at the three levels mentioned achievement and attitude tests devised by international committees who had spent three years designing and trying out these instruments. The IEA mathematics project conducted in 1962-65 (Husén, 1967a) was the first attempt on a large scale to obtain objective measurements of student performance for a broad array of countries, all of them, however industrialised.

The IEA survey has been a huge enterprise in terms of time, money and number of individuals involved. The Six Subject Survey comprised some 250,000 students in 9,700 schools. One could, indeed, ask what kind of rationale could be advanced for such a project with all its technical complexities and far reaching administrative complications.

When the IEA research was launched some 15 years ago, the National Centres involved simply wanted to take advantage of international variability with regard to both the outcomes of the educational systems and the factors which accounted for differences in these outcomes. In a way, the world was conceived of as one big educational laboratory where different practices in terms of school organization, curriculum content and methods of instruction were experimented with. But before trying to analyse cross-nationally the 'effects' of various input factors on educational outcomes, it was necessary to devise internationally valid evaluation instruments. Not until the IEA research was launched were such instruments available. Therefore the prime concern during the first years of IEA research was the construction of appropriate measuring techniques that could result in the establishment of adequate international yardsticks. These were, indeed, badly needed, not least for evaluating school reforms in all the countries and particularly technical assistance programmes in education in the LDC's.

Pure 'head-counting', for instance enrolment and graduation statistics (see e.g. Harbison and Myers, 1964), was often used as a criterion of evaluation, lacking qualitative indicators, such as student competence achieved in various subject areas. The efforts at the beginning of the IEA research to devise instruments by means of which international standards could be established unfortunately gave some people the false impression that the main purpose of the exercise was to conduct some kind of international horse race or 'cognitive olympics'. But the development of new evaluative techniques and the setting up of the international co-operative machinery that went with it were preprequisites for establishing international standards in a series of subject areas, such as mathematics and reading. Not until the IEA reading survey, which also comprised three LDC's (Chile, India and Iran), were any comparative assessments of the level of literacy among representative groups of students in such countries available.

Once suitable measuring instruments were available, the next step was to identify the salient factors which accounted for cross-national differences. Since this could be done in a replicative way at the various levels of the single national systems and across these systems, a much more multi-faceted picture of factors accounting for differences in student attainment between school systems could be obtained. The comparative approach implied that we widen the population of classrooms from one particular school within one particular national system to a representative set of classrooms within several national systems. IEA shared the ambitions prevalent in the social sciences in general, that is to say, to arrive at generalizable findings. By repeating surveys and analyses over many countries which differed with regard to important social and economic factors, a more detailed picture of what accounted for differences in 'productivity' between these systems could be arrived at. Since the ultimate aim of research in the social sciences is not only to identify and describe but to explain and predict, that is to say, to generalise, the basis for such an operation can be broadened by including inter-system and inter-country variables which allow cross-national generalisations and also make it possible to study how intra-system and inter-system variables interact.

We can take as an illustration how class size is related to student performance. Practically all the sample surveys so far have been carried out in the United States and some West European countries. These studies

consistently indicate that class size and performance tend to be positively correlated at the level of 0.10-to 0.20. The fact, however, that class size within these countries covers a rather narrow range makes generalizations about such a relationship extremely awkward. In a multi-national study one can take into account variables such as teacher competence, school resources, and socio-economic structure, which vary widely between countries. This provides an opportunity for obtaining not only a more diversified descriptive picture but also for opening up new avenues of analysis.

One overriding purpose of the IEA Six Subject Survey has been to study the relationship between input factors in the social, economic and instructional domains and output as measured by international tests covering both cognitive (student performance) and affective behaviours (student attitudes and motivation). These relationships have been studied in some twenty national systems of education and, as a rule, at three different levels within each system.

After the completion of the IEA mathematics survey, two international meetings resulted in the report, "Toward a cross-national model of educational achievement in a national economy" (Super, 1970). The aim was to develop an input-output model that could serve as a theoretical framework for the next survey, where achievement criteria from six subject areas were going to be developed. Researchers from the various social science disciplines were brought together to review both national and international research already undertaken and to advance new hypotheses which could be tested in further research. They were also asked to suggest the inclusion of independent variables to a social and economic nature that should be included in the proposed survey.

A key problem in conducting cross-national evaluation studies, where comparisons are made between student performance by means of standardized achievement test, has to do with comparability per se (Husén, 1967b). Two major comparability problems are encountered: the drawing of strictly comparable samples of students and the construction of measuring instruments that are 'fair' in terms of their content, matching the students' opportunity to learn the subject-matter tapped by the tests. The technical aspects of these problems have been dealt with in detail in the IEA international reports (see, e.g. Peaker, in press; Comber and Keeves, 1973, p.42 et seq). IEA has

succeeded in establishing a system whereby national random samples, be they age samples or grade samples, can be drawn. Once the target populations have been defined (e.g. 14 year-olds) and the sampling design has been drawn up, the problem of executing the sample is mainly an administrative one. In several countries, both developed and less developed, the conduct of the Six Subject Survey was the first occasion when nationally representative samples of students were drawn. The experiences gained in countries like Iran and India, for instance, can be drawn upon in the future when procedures of evaluating entire national systems by means of random samples are going to be established as routines.

2. Misgivings about appropriateness of employing international evaluation standards

One criticism levelled against the IEA mathematics study by mathematics educators in a special issue of the Journal for Research in Mathematics Education (Findley, 1971) was a lack of comparability due to considerable differences between countries in terms of the amount of exposure to the teaching of the various topics covered by the items in the international mathematics tests. Country means of teachers' ratings of 'opportunity to learn' and student achievement are indeed rather highly correlated over countries (see, e.g. Comber and Keeves, op.cit. p.158 et seq.). But it should be kept in mind that rank order correlations between country aggregates could be quite high, and indeed are. When countries were correlated over item difficulties, it was found that the overlap in achievement structure was remarkable, that is to say, country differences were only to a minor extent accounted for by dramatic differences in particular topics or sub-areas within one subject but rather by systematic differences over the whole range of topics and items. At least in subjects like mathematics and science, where the subject matter by its very nature is rather universal, the differences between national systems seem to affect all topical areas in a systematic way and not just a few.

The machinery that went with the construction of the international achievement tests in a way served as a safeguard against undue cultural bias. An international committee was set up for each subject area. These committees being composed of subject matter specialists, teachers, test developers and curriculum specialists, were responsible for the construction of the test instruments and for the development of questionnaires related to their respective fields (see, e.g. Comber and Keeves, 1973, p.27 et seq.).

Contact with the participating countries was effected through the national research centres and subject committees set up in each country. The analyses of the curricula, the proposing of test exercises and the try-out of the items were carried out in the participating countries. IEA headquarters served only as a coordinating centre and a clearing house.

Since the main purpose of achievement tests is evidently to measure differences in achievement, complete equality in terms of exposure to teaching and opportunity to learn would make the administration of such tests rather pointless. The same applies to so-called intelligence tests, where individual and group differences unavoidably also reflect differences in terms of opportunity. As has been spelled out in another connexion (Husén, 1967b), the international administration of achievement tests differs only in degree and not in principle from their administration on a national scale. Within a given country there are differences between school districts and regions due both to differences in student background and school resources. Very few are those who would dispute the worthwhileness of administering the same test of achievement to all the children at the same grade level in a given country, once the test measures the main objective it is purported to measure. For instance, the finding within a given country that children in urban areas perform better than children from rural areas or that the socially privileged have higher scores than the underprivileged is per se not to be interpreted as an act of discrimination against those who socially and pedagogically have been subjected to the less favourable conditions. The establishment of factual differences in terms of agreed criteria of performance is in itself of informative value. It can, as in the case of the IEA research, serve as a basis for analysis of what factors account for differences in performance and can ultimately be used for more adequate educational policy. The data collected can also serve as a basis for evaluating how far students have been brought under the prevailing conditions and for analyses of what could be done in order to improve these conditions.

The rationale indicated above also applies to comparisons between highly industrialised and more or less agricultural economies, in brief, to comparisons between developed and LDC's. So far, no representative comparative information with regard to student competence in LDC's has been available. Those who have first-hand experience have intuitively felt that differences between students who grow up in countries where there is a long tradition of literacy, and those whose parents in most cases are illiterate, are sometimes quite significant.

Misgivings have been expressed in some quarters about the worthwhileness of an exercise where national school systems in LDC's have been evaluated according to the same standards as those in the industrialized countries with their tradition of universal formal schooling that now is some hundred years old or more. These misgivings range all the way from objections about 'comparing the incomparable' to pointing out that the LDC's can be expected to suffer from certain handicaps because of the format and methodology employed in conducting the evaluation.

It would in this connexion take us too far to discuss in detail the adequacy - or lack of adequacy -/of the methodology. I shall therefore limit myself to spelling out the rationale for establishing a common standard of achievement in an attempt to evaluate national systems of education in both industrialised and non-industrialized countries, the latter allegedly attempting to develop their economies in the same direction as the former. I shall also point out certain flaws evidenced by the Six Subject Survey, which - it should be kept in mind - was the <u>first systematic attempt</u> to evaluate primary and secondary education in LDC's according to some kind of international norms.

The introduction of universal elementary schooling in, for instance, Western Europe during the 19th Century, when in most countries certain basic schooling by state legislation was made compulsory (frequently with

opposition from the peasants), has to be viewed in its economic context. Most of the countries were in the midst, or at the beginning, of a 'great leap' in industrialization. Apart from the task of taking care of children in urban areas whose parents were working long hours in the factories, the school was supposed to provide the literacy and numeracy required by the labour force in industry. To be sure, most LDC's are not yet at the stage of industrialization reached by the West European countries by, say, 1870. Subsistence from agriculture is still far more widespread, and this of course raises some doubts about the adequacy of institutionalized elementary schooling when the children in rural areas by tradition work with their parents. But if the goal behind the efforts to build up an educational system in the LDC's is to achieve 'more modernization', that is, among other things, to build up an infrastructure of knowledge and skills conducive to an economic development which has radically changed the standard of living in the industrialized countries, then much can be said for attempts to assess the competence achieved in, for instance, reading and science that is basic to modern technology. Such competencies have been defined in the IEA survey as the result of cooperation between the participating research institutions in both developed and developing countries. Instruments for their measurement were constructed and tried out conjointly before being administered to representative samples of students in the respective countries.

The format of the achievement tests employed constituted a serious handicap for students in the LDC's. Psychological studies have shown that children brought up in cultures where sustained efforts in pursuing assigned tasks have not been an everyday part of their training have difficulties in mobilizing the motivation that is required to complete a test examination with increasingly more difficult test exercises. The tests were so-called paper-and-pencil ones, that is, the students had to read the exercises and then respond by blackening the space on an answer sheet that corresponded to the

correct alternative, of which there were five as a rule. The most serious drawback among students at the 10- and 14-year-old levels was their frequent lack of the reading competence necessary to understand the test exercises. A high proportion of those who either gave wrong responses or omitted responses did so because they were unable to understand the questions. Thus, one important lesson learned from the Six Subject Survey is that in evaluating cognitive competence, be it skills in the three R's or basic items of information in the content subjects, such as science or civics, one would have to develop new formats for the examinations which would reduce the handicap inherent in a low level of reading competence. On the other hand, since a certain level of reading comprehension is instrumental in acquiring knowledge in other subject areas, it could be argued that lack of sufficient skill in reading should not be regarded as a serious handicap.

I am fully aware of the objections raised by some of my colleagues in the international assistance agencies that the comparisons have been 'invidious', because they might not have taken fully into account the explanatory factors underlying the very significant differences in achievement between developed and less developed countries. A more 'pluralistic' approach would have seemed to be in order. Apart from the fact that the IEA survey, as was emphasized above, was not intended to be an international olympics, the crucial point is to what extent it is justified to apply <u>one</u> standard of comparison across countries so different in their social and economic structure, not to speak of the tremendous differences in culture and traditions. The point made above for the 'unidimensional' approach is that if one wants to achieve 'modernization', then certain consequences are entailed, such as the establishment of certain competencies conducive to industrialization.

3. Organisation of international evaluation of educational outcomes

It was pointed out above that to conduct multi-national evaluation surveys is, indeed, a complicated task. A basic prerequisite is the setting up of some kind of machinery that can secure the necessary co-ordination and communication between the participating research institutions. The national research centres have to take decisions about subject areas and problems they want to investigate. A uniform design guiding the construction of instruments, data collection and data processing has to be laid down. A timetable for all these activities has to be agreed upon. Since several languages are involved - in the Six Subject Survey no less than 14 - problems of translation of tests and manuals of instruction have to be properly handled. For instance, to what extent is it possible to avoid cultural biases when tests of reading comprehension are constructed, translated and given in vastly different cultural settings? This problem is a challenging research task in its own. It was dealt with in the feasibility study and was further elucidated in the Six Subject Survey when reading tests were given to students in three developing countries (Thorndike, 1973). However, communication problems are not solved by penetrating language barriers only. Differences in national values and habits can cause difficulties, not least with regard to promptness - or lack of promptness - in responding to letters or sticking to timetables!

Since IEA constitutes the largest network of co-operating research institutes conducting empirical research in education in the world today, it would seem in order to describe briefly its organisational features.

In 1959, a group of researchers from twelve countries, who convened under UNESCO auspices, decided to embark upon a small pilot study to examine to what extent it was feasible and meaningful to undertake multi-national 'standardised' survey research. The pilot study turned out to be rather successful in both respects. It was possible in a series of subject areas to construct achievement tests that could be translated and administered uniformly to students in different countries and to arrive at meaningful interpretations of between-country differences (Foshay, 1962). It was administratively and technically feasible to collect data uniformly and to have them processed in one place. Therefore, it was decided to undertake a more rigorous study using probability samples from twelve countries, of which all were industrialized (Australia, Israel, Japan, the United States and eight West European countries). Student achievement in mathematics

was chosen as the main criterion of output, since this subject by its universal nature seemed to be more readily accessible to international comparisons than other subject areas, possibly with the exception of science.

In the IEA mathematics study, two major levels in the school systems of the twelve countries were sampled (Husén, 1967):

(a) 13-year-olds (both age and grade populations), since this was the last point in all the systems where 100 per cent of the relevant age group was still in full-time schooling; and

(b) pre-university grade students.

In all 133,000 students were tested and completed questionnaires in mathematics study. Furthermore, 13,500 teachers and 5,450 school principals completed questionnaires with information on instruction, curriculum and school resources. The information gathered in this survey was used to test hypotheses concerning: (1) the relationship between different teaching practices in school and outcomes of instruction; (2) the relationship between the organisational features of the systems, such as age of school entry, grouping practices, and student-teacher ratio, to outcomes; and, (3) the relationship between home background and outcomes. Several special studies, for instance one on the relationship between the 'yield' and certain organisational features (Postlethwaite, 1967), were also conducted.

After the completion of the feasibility study and the first main study (in mathematics) the participating research centres in 1967 formed a corporate body. The main reason for this was to establish IEA as a legal entity eligible for research grants. Thus, IEA is now an international non-profit-making, non-governmental association constituted under the name of the "International Association for the Evaluation of Educational Achievement".

The Association is constituted in accordance with the Belgian law of 1919 regarding international non-profit-making, scientific societies, and which was modified by a law of 1954. IEA has from its inception had close relationships with the United Nations Educational, Scientific and Cultural Organisation (UNESCO). The feasibility study and the mathematics survey were conducted under the auspices of the UNESCO Institute for Education in Hamburg, where the IEA working headquarters were located until

1969. At that date they were moved to Stockholm and are at present accommodated within the Institute for the Study of International Problems in Education in the University of Stockholm. IEA has a consultative relationship with UNESCO.

Membership in IEA is restricted to institutions carrying out research in education. In order to be eligible for membership an institute should have a good reputation, qualified staff, ready access to schools in the national school system and the necessary financial resources to carry out the research work to which the institute has committed itself. Membership is upon application decided upon by the IEA Council, which is made up of one representative from each national centre. The number of members is at present 23, consisting of ten West European countries (Belgium (with the Flemish-speaking and French-speaking parts being treated as two separate entities), Federal Republic of Germany, Finland, France, Ireland, Italy, Netherlands, Scotland, Sweden and the United Kingdom), three East European countries (Hungary, Poland and Romania), and nine non-European countries (Australia, Chile, India, Iran, Israel, Japan, New Zealand, Thailand and the United States).

The Council meets, in principle, once a year and determines the general policy of the Association. It elects a Chairman and a Standing Committee consisting of six of its members. The Standing Committee elects two of its members to serve with the Chairman on the Bureau, which meets several times a year and is responsible for the execution of decisions taken by the Council. The centre staff employed by IEA consists of an Executive Director, research officers, technical assistants and secretaries. During the Six Subject Survey two data processing units were established, one in New York for the first stages of processing and one in Stockholm for further processing and the statistical analyses. A data bank has been established at the University of Stockholm.

In conducting the Six Subject Survey, the Council had to establish various bodies for conducting and reporting the research. As mentioned above, one international committee was appointed by the Council in each subject area in which survey research was undertaken. Further, the Council set up a technical committee which was responsible for overall decisions taken on technical problems pertaining to sampling, data collections and data processing. The international committees interact with national committees set up in the various subject areas. For example, during the

IEA Six Subject Survey some 300 persons spread across 19 countries with 14 different languages were engaged in the construction of instruments. During the mathematics study English and French were used as linguae operandi at international meetings and in correspondence, but in the Six Subject Survey it was decided to use only English.

In the Six Subject Survey the data for 250,000 students were made available to the data processing centre on either cards (in most cases) which could be optically scanned (MRC cards), tapes or punched cards. The MRC card-reading took place in Iowa City. The editing, sorting, filing, item analysis and run-off of univariates was done in New York at Columbia University, and the bivariate and multivariate analyses were conducted at the University of Stockholm. Data on some 2,000 variables were collected, most of these being input variables. The variables in any one subject area at any one level of the school system amounted to between 200 and 500. To be sure, there were too many to be manageable in multivariate analyses and they had to be considerably whittled down on the basis of analyses of the intercorrelation matrices.

4. Mean performance in science and reading in industrialised and non-industrialised countries

The following three target populations were sampled in the Six Subject Survey:

Population I : All students in full-time schooling aged 10:00-10:11;

Population II: All students in full-time schooling aged 14:00-14:11; and

Population IV: All students in the terminal year in full-time secondary school programmes which were either pre-university programmes or programmes of the same length (this gave the national centres some latitude of interpretation, which means that in some countries only those students who were about to complete courses which in a narrow sense qualify for university entrance were included, whereas in other countries those who are about to complete qualified vocational programmes were also included.)

It would indeed be preposterous to try to condense the findings from the comprehensive Six Subject Survey into a few pages: the report series will upon completion consist of nine volumes ! We shall therefore confine

ourselves here to a presentation of some findings which seem to have a particular bearing on the evaluation of education in LDC's, particularly since this is the first time that qualitative comparisons between industrialised and LDC's have been made according to agreed-upon international yardsticks.

Table 1 shows the means and standard deviations in total science score and total reading comprehension score in the 19 participating countries, of which four are mainly less developed. We have limited ourselves to these two cognitive criteria, since data on them are available for four and three LDC's respectively. The only LDC which participated in literature was Chile, which also participated in English and French. Iran was the only LDC participating in civics.

The most dramatic difference is the one between the industrialised and non-industrialised countries. The latter are consistently far behind the former in average achievement over subject areas and levels of schooling. In science the LDCs' score was roughly one standard deviation or more below the more developed. This means, then, that in science the average student in a LDC scores between the 10th and 12th percentile in a developed country. The difference is even more pronounced in reading comprehension, where only some 5 to 10 per cent of the students in the LDC's score at the level of the average student in a more developed country. Chile participated, as mentioned above, in the survey of French and English as foreign languages and Iran in civics. The mean cognitive scores in both cases turned out to be on the same relative level as in science and reading.

What explanations can be advanced for such big differences ? In the first place, we must emphatically caution against any premature conclusions about the 'productivity' or 'efficacy' of the school systems in the two types of countries on the basis of the mean scores presented in Table 1. The differences that we find between the industrialised countries are negligible in comparison with the gap between the two categories of countries. There is, however, no reason to believe that the rich countries all are on the same level of 'efficacy' as regards their school systems.

A first-hand explanation that would seem plausible is that the tests are not doing justice to the children in the LDC's. The tests might draw upon knowledge and learning experiences that are more predominant in the rich countries. Furthermore, the test situation as such and the format of assessing the outcomes of learning might imply a certain cultural

Table 1. Mean total score and standard deviation in science and reading comprehension among 10-year-olds, 14-year-olds, and pre-university students

	SCIENCE						READING COMPREHENSION					
	10-year-olds		14-year-olds		Pre-university students		10-year-olds		14-year-olds		Pre-university students	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Australia	--	--	24.6	13.4	24.7	10.7	--	--	--	--	--	--
Belgium (Flemish)	17.9	7.3	21.2	9.2	17.4	8.1	17.5	10.2	24.6	9.7	25.0	9.3
Belgium (French)	13.9	7.1	15.4	8.8	15.3	7.9	17.9	9.3	27.2	8.7	27.6	9.2
England	15.7	8.5	21.3	14.1	23.1	11.5	18.5	11.6	25.3	11.9	33.6	9.0
F.R.G.	14.9	7.4	23.7	11.5	26.9	8.9	--	--	--	--	--	--
Finland	17.5	8.2	20.5	10.6	19.8	9.8	19.4	10.8	27.1	10.9	30.0	7.5
France	--	--	--	--	18.3	8.7	--	--	--	--	--	--
Hungary	16.7	8.0	29.1	12.7	23.0	9.0	14.0	9.8	25.5	9.9	23.8	8.9
Israel	--	--	--	--	--	--	13.8	11.0	22.6	12.8	25.2	10.8
Italy	16.5	8.6	18.5	10.2	15.9	8.8	19.9	8.8	27.9	9.3	23.9	10.2
Japan	21.7	7.7	31.2	14.8	--	--	--	--	--	--	--	--
Netherlands	15.3	7.6	17.8	10.0	23.3	11.1	17.7	9.5	25.2	10.2	31.2	7.0
New Zealand	--	--	24.2	12.9	29.0	11.6	--	--	29.3	11.0	35.4	8.1
Scotland	14.0	8.4	21.4	14.2	23.1	12.1	18.4	11.1	27.0	11.5	34.4	8.2
Sweden	18.3	7.3	21.7	11.7	19.2	10.2	21.5	10.5	25.6	10.8	26.8	9.3
United States	17.7	9.3	21.6	11.6	13.7	9.5	16.8	11.6	27.3	11.6	21.8	12.0
Industrialized Countries	16.7	7.9	22.3	11.8	20.9	9.9						
Chile	9.1	8.6	9.2	8.9	8.8	6.0	9.1	9.3	14.1	11.1	16.0	8.8
India [1]	8.5	8.3	7.6	9.0	6.0	6.0	8.5	9.4	5.2	7.2	3.5	5.8
Iran	4.1	5.4	7.8	6.1	10.2	5.6	3.7	6.9	7.8	6.7	4.4	6.0
Thailand [2]	9.9	6.5	15.6	8.1	12.4	6.1	--	--	--	--	--	--

[1] India samples the Hindi-speaking states or regions only.

[2] Thailand did not test a national sample, but samples schools in the Bangkok area only.

bias against students in LDC's. We certainly cannot entirely refute such hypotheses, but they do not get much support from the empirical evidence we have. In the first place, the content of the tests, i.e. the individual test items, went through a long procedure of scrutiny and try-out before they were 'passed' by all the national subject area committees and included in the international tests. Secondly, the rank order of difficulties of items tended to be highly correlated over countries, which indicates that differences in total scores between countries are not so much accounted for by differences in particular sub-areas or topics of a particular subject as by systematic differences in level of competence. The teachers were asked to rate, on a four-point scale, each item in the tests with regard to what opportunity the students in their classes had had to learn the subject matter that was assessed by the item. As far as science is concerned, the average opportunity tended to be somewhat lower for Populations II and IV in the LDC's (see Comber and Keeves, 1973). However, these differences in opportunity can by no means explain more than a small portion of the difference in mean performance.

The main factor is no doubt the socio-economic gap between the two categories of countries. Education does not operate in a socio-economic vacuum, a fact which is shown not the least by the consistently substantial correlations between various family background measures and student achievement in all subject areas. Passow, Noah and Eckstein have, in their report on the 'National Case Study Questionnaire' (in press), drawn up 'national profiles' for the 19 countries which participated in the first stage of the Six Subject Survey. The size of the per capita GNP varies from about US$ 1,400 to 4,300 in the industrialised countries, whereas it varies from $90 to 270 in the LDC's which took part in the study. The size of the non-primary sector of the economy in per cent of the GNP is in most cases 90 to 95 per cent in the rich countries as compared to 50 to 75 per cent in the LDC's. The difference is even more marked if we measure the size in terms of number of people employed in the primary and non-primary sectors respectively.

Thus, the difference between developed and less developed countries could be expected, considering the overall socio-economic setting for the school systems in the two categories of countries. The outcomes of the multivariate analyses tell us that the total effect of home background variables in both science and reading is greater than the total effect of

all the school variables. Among the 10-year-olds, 35 per cent of the variation between students can be attributed to family background and 22 per cent to school factors, including, of course, all the instructional factors. The corresponding figures for the 14-year-olds are 42 and 26 per cent respectively. What is 'family background' then ? After a careful study of some 20 variables that could be considered as candidates for an overall measure of social background, the following were selected to form a composite 'School Handicap Score' (SHS): (1) Father's Occupation, (2) Father's education, (3) Mother's education, (4) Use of dictionary at home, (5) Number of books at home, and (6) Family size. It is pointed out, in the international report in science, that the "effectiveness of the education provided by the school must be assessed by what is achieved, after allowance has been made for the nature of the community in which the school is operating" (Comber and Keeves, 1973, p. 195). Thus, regardless of the quality of the formal educational system, we can, on the basis of the impact of the family background factors, predict a large difference in mean achievement between the less and the more industrialised countries. Parents in the former type of countries are in most cases illiterate and no reading material is available at home. On the whole, the verbal environment in which the children grow up is almost entirely oral and there are rather few occasions in which reading skills picked up at school can be reinforced by experiences at home.

A simple reading speed test was developed in order to measure to what extent the mechanics of reading skills had been acquired. The items consisted of short paragraphs of two or three simple sentences, and the students by checking the right answer of a choice of three had to indicate that he had understood what he had read. The items were like this:

> "Peter has a little dog. The dog is black with a white spot on his back and one white leg. The colour of Peter's dog is mostly:
>
> black brown grey."

On the average, 10-year-olds in Europe had an error rate of about 10 per cent on items such as the one cited. At the 14-year-old level the rate had gone down to about 4 per cent. For the three LDC's the rates were:

	10-year-olds	14-year-olds
Chile	26%	16%
India	36%	33%
Iran	52%	20%

Therefore, there is some justification for what was said earlier that quite a few of the 10- and 14-year-olds in the LDC's have not been able to read the science items and the questions in the student questionnaires.

5. The establishment of research competencies in education in LDC's

The IEA survey research, conducted over more than ten years, is indeed a highly sophisticated one. Therefore, doubts have been raised, not least in international agencies involved in technical assistance in one way or another in LDC's, as to whether the techniques developed by IEA might not be too sophisticated to become part of routine evaluation procedures in these countries.

Since four LDC's participated in the Six Subject Survey along with 15 more or less industrialised countries, it would seem in order at this juncture to take stock of the experience which has been gained.

In the first place the participating institutions have accumulated a vast experience in terms of research strategies and techniques related to the evaluation of national systems of education. The IEA international headquarters as well as the national centres have over the years cooperatively built up a considerable amount of collective competence with regard to the conceptualisation of evaluation research, the appropriate techniques for dealing with different kinds of problems and the modes of feedback to policy makers in the countries concerned. The completed studies have had an impact on purely pedagogical matters, such as curriculum development and the provision of instructional facilities, but also on considerations related to the structure of the school systems.

In spite of the obvious limitations and drawbacks that the application of the IEA methodology had in some LDC's and which have been dealt with above, I think that for two major reasons the experiences gained (which we, of course, have to take stock of) make a case for further developmental work that would in the long run make these techniques a routine procedure in evaluating the systems of education in LDC's.

In the first place, the major advantage that I see as the most encouraging experience from the IEA Six Subject Survey is its contribution to the build-up of research competence in the participating national centres. Those of us who were responsible for the technical and administrative co-ordination of the project have found how, in spite of scepticism and bad odds, the research competencies in the LDC's especially were tremendously developed.

In some countries this was the first time a sample survey in education had been conducted and by bringing together the technical officers to international seminars and briefings or by dispatching experts from the IEA headquarters, the technique of drawing nationally representative samples had to be learned by actual practice. Examinations are in some countries conducted nation-wide with instruments which cannot be quickly and objectively scored, since they are essay examinations. The development and tryout of the IEA international achievement tests in these countries was another lesson learned by actually ca.rying out the procedure. Finally, the techniques that can be used in data processing in education and making such data available to statistical analyses conducive to finding out what factors account for between-student and between-school differences in achievement nad to be learned in the same way.

It was, however, a matter not only of trying to build competence in conducting evaluation surveys but also of making those who were involved aware of certain features of their own national system of education by broadening their perspective to encompass a series of other systems. The National Case Study Questionnaire, which had to be completed by each of the national centres, aimed at collecting information not only about overall features of the respective national systems of education as such, but also about the social and economic settings in which the systems were operating.

Those who were responsible for conducting the survey, not least those in the LDC's, learned a lot about their own systems which they did not have a concrete idea about before. A national survey of the educational system in a country, with all its limitations and technical snags, provides findings which can be brought to bear on educational policy and planning. So far we have known very little about what factors account for differences between schools and students in achievement. We have, for instance, not been aware of the fact that the various factors in the home background do not play the same rôle in many LDC's as they do in the highly industrialised

countries. Western 'standard' background variables, such as father's occupational status and parental education seem to account much less for differences between students in achievement in the LDC's than in the highly industrialised countries.

A detailed analysis of student performance in a particular subject area can provide valuable feedback to curriculum developers. This is particularly useful for curriculum development in LDC's, since there has been a strong tendency to adopt subject matters as defined by textbooks in the industrialised countries without closer consideration of the particular needs and circumstances in the borrowing country.

Finally, it should be pointed out that the evaluation techniques employed by IEA are in principle applicable to both the formal and informal educational system. Individuals have to be sampled in a representative way. Yardsticks of performance as well as of attitudes have to be developed. Questionnaires administered to students, teachers and administrators have to be devised in order to collect relevant background information. Such information, by the way, is not always available, simply because it might never have been the object of any kind of surveyor census.

6. Concluding remarks

It is by no means a coincidence that international co-operative survey research in education started with evaluation problems. Before one can begin to investigate to what extent various factors account for differences between classrooms, school and entire national systems of formal education, it is necessary to develop international criteria of evaluation. The construction of international instruments that can be used in evaluating both the cognitive and non-cognitive outcomes of instruction is in itself an important research accomplishment. But it is only the first step on the way to the ultimate goal which is to identify the salient factors which account for differences between systems and to explain why they differ. By means of such research it will be possible to establish international indicators of the qualitative outcomes of school education. One would thereby also be able to inform planners and policy-makers about what indicators are worthwhile to manipulate in terms of policy action.

Closely related to this is the problem of how the 'productivity' of a national system of school education should be assessed. Too long have we tended to evaluate the outcomes in terms of the number of individuals who are enrolled at a particular stage in the system or in terms of

how many years they have completed rather than by the competence they have achieved. A certain amount of schooling in terms of number of years or a particular certificate can by no means be regarded as comparable quantities from one system to another. Furthermore, it is not satisfactory, when evaluating its quality, to limit oneself to the end products of a system. One has also to consider its power to take care of and impart competence in all students who enter the system. Since attrition, particularly in terms of drop-outs, is in many systems very high, one basic question that needs to be answered in evaluating a system is: How many students are brought how far ?

As far as the evaluation of national systems of education in the LDC's is concerned, the IEA research has brought about the accumulation of strategies and techniques which can begin to be utilised routinely. Methods of analysing national curricula in terms of the goals which are to be achieved have been developed. Similarly, techniques have been devised by means of which instruments can be constructed to measure these goals. Procedures for drawing probability samples from target populations under consideration have been developed. Routines for data collection in the schools have been tried out in a wide variety of contexts. Finally, experience has been gained in data processing of particular relevance to nation-wide evaluation surveys.

The IEA international headquarters, as well as the national centres, have over the last ten years built up a considerable amount of collective competence with regard to the conceptualisation of research problems connected with evaluation, the techniques employed and the different modes of feedback to policy-makers in the countries concerned. The co-operative machinery that has been built up could be utilised to provide training programmes for students from regions of the world where particular strengths and competencies in evaluation are still developing. From the IEA international network one could set up task forces to work with centres in LDC's. Such forces could co-operate with local researchers on designing evaluation surveys.

BIBLIOGRAPHY

Beeby, C.E., The Quality of Education in Developing Countries, Cambridge, Mas., Harvard University Press, 1966.

Carroll, John B.,(in press) French as a Foreign Language in Seven Countries, Stockholm and New York, Almqvist & Wiksell and Wiley-Halsted Press.

Coleman, James S., "Effects of School on Learning: The IEA Findings." Paper presented at the Conference on Educational Achievement, Harvard University, November 1973.

Comber, L.C. and Keeves, J.P., Science Education in Nineteen Countries, Stockholm and New York, Almqvist & Wiksell and Wiley-Halsted Press, 1973.

Farnen, Russell F., Oppenheim, A.N. and Torney, J.V.,(in press) Civic Education in Ten Countries, Stockholm and New York: Almqvist & Wiksell and Wiley-Halsted Press.

Findley, Warren G.,"Critiques of the International Study of Achievements in Mathematics: The Problem of Perspective." Journal for Research in Mathematics Education, 2:2, 1971, p. 104-106.

Foshay, Arthur W. (Ed.), Educational Achievement of Thirteen-Year-Olds in Twelve Countries, Hamburg, UNESCO Institute for Education, 1962.

Harbison, Frederick H. and Myers, Charles A., Education, Manpower and Economic Growth, New York, McGraw-Hill, 1964.

Husén, Torsten (Ed.), International Study of Achievement in Mathematics: A Comparison Between Twelve Countries, New York and Stockholm, John Wiley & Sons and Almqvist & Wiksell, 1967.

Husén, Torsten, "Establishing International Criteria of Comparison in Comparative Education" in: General Education in a Changing World, Amsterdam, Nijhoff, 1967b.

Husén, Torsten, "Implications of the IEA Findings for the Philosophy of Comprehensive Education". Paper presented at the Conference on Educational Achievement, Harvard University, November 1973.

Lewis, E. Glyn and Massad, Caroline E., (in press) English as a Foreign Language in Ten Countries, Stockholm and New York, Almqvist & Wiksell and Wiley-Halsted Press.

Passow, A. Harry, Noah Harold J. and Eckstein, Max, (in press) The National Case Study: An Empirical Comparative Study of Twenty-One Educational Systems, Stockholm and New York, Almqvist & Wiksell and Wiley-Halsted Press.

Peaker, Gilbert F., (in press) An Empirical Study of Education in Twenty-One Countries: A Technical Report, Stockholm and New York, Almqvist & Wiksell and Wiley-Halsted Press.

Postlethwaite, T. Neville (Ed.), Comparative Education Review, 18:1. Special Issue Reporting National Analyses of IEA Data. 1974.

Postlethwaite, T. Neville, School Organization and Student Achievement: A Study Based on Achievement in Mathematics in Twelve Countries, Stockholm and New York, Almqvist & Wiksell and John Wiley, 1967..

Purves, Alan C., Literature Education in Ten Countries, Stockholm and New York, Almqvist & Wiksell and Wiley-Halsted Press, 1973.

Super, Donald E. (Ed.), Toward a Cross-National Model of Educational Achievement in a National Economy. A report of the Lake Mohawk Conference. Teachers College, Columbia University, 1970.

Thorndike, R.L., Reading Comprehension Education in Fifteen Countries, Stockholm and New York, Almqvist & Wiksell and Wiley-Halsted Press, 1973.