DOCUMENT RESUME

ED 112 500 EA 007 535

AUTHOR	Husen, Torsten
TITLE	Multi-National Evaluation of School Systems. IIEP Occasional Papers No. 37.
INSTITUTION	United Nations Educational, Scientific, and Cultural Organization, Paris (France). International Inst. for Educational Planning.
REPORT NO	Occas-Pap-27
PUB DATE	Jan 75
NOTE	24p.
AVAILABLE FROM	IIEP Publications, 7-9 rue Eugene-Delacroix, 75016 Paris, France ($1.00, distribution charges)
EDRS PRICE	MF-$0.76 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	*Academic Achievement; Developing Nations; Educational Assessment; Elementary Secondary Education; Evaluation Criteria; *Evaluation Needs; *International Programs; National Surveys; *School Surveys; Statistical Data; *Tables (Data)
IDENTIFIERS	IEA

ABSTRACT

This paper highlights some of the main conclusions from the Six Subject Survey of education in 19 countries conducted by the International Association for the Evaluation of Educational Achievements. Beside presenting some findings of the survey, the author reviews the aims of conducting such multinational comparisons and discusses the future of evaluating national systems of education, with particular reference to the needs of developing countries. (Author/JG)

IIEP Occasional Papers No. 37

MULTI-NATIONAL EVALUATION OF SCHOOL SYSTEMS

Torsten Husén

Unesco: International Institute for
Educational Planning

IIEP Occasional Papers

The studies in this series include papers contributed by the Institute's staff, visiting fellows, interns and consultants. Some of the studies have originally been prepared as part of the training programme of the Institute; others have previously appeared as working papers for the Institute's seminars and symposia. All of them, in the Institute's views, are of sufficient interest to merit being re-issued and distributed on a wider scale.

By their very nature these papers are less formal and have not been given the full editorial processing customary for IIEP 'official publications'.

The opinions expressed in these papers are those of the authors and do not necessarily represent the views of the Institute. The use, adaptation or reproduction, in whole or in part, of these papers is limited to institutions and persons specifically authorized by IIEP.

$1.00

Printed in France by
International Institute for Educational Planning
7-9, rue Eugène-Delacroix, 75016 Paris
January 1975

CONTENTS

I. AIMS OF CONDUCTING MULTI-NATIONAL COMPARISONS

In May 1973 the first three reports from the 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 evaluations of school education in 19 countries by drawing upon criteria in Science, Literature and Reading Comprehension, respectively. Within a short time, the three remaining subject areas will also be reported, namely English and French as foreign languages and Civics (Lewis and Massad, in press, Carroll, in press; Farnen, Oppenheim and Torney, in press). IEA in the first stage of its research on evaluation focused on mathematics, which was reported some years ago (Husén, 1967).

One could, indeed, ask about the rationale for embarking on a venture which has included 250,000 students in 9,700 schools in 19 countries with all its far-reaching administrative implications and formidable technical complexities. When the IEA research was launched some 15 years ago, those who were involved simply wanted to take advantage of the 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 could be conceived of as one big educational laboratory where different practices were experimented with in terms of school organization, curriculum content and methods of instruction. But before trying to analyze 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 did such instruments become 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 certain technical assistance programs in education in the LDC's. Pure 'head-counting', for instance enrollment and graduation statistics (see, e.g. Harbison and Myers, 1964), was 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 an international cooperative machinery that went with it was a prerequisite for establishing international standards in a series of cognitive domains, 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 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 widened the population of classrooms from one particular school within one particular national system to a representative set of classrooms within several national systems. Thus, 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, generalize, the basis for such an operation can be broadened by including inter-system and inter-country variables which allow cross-national generalizations 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 that so far have been conducted 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 (Marklund, 1962). The fact, however, that class size within these countries covers a rather narrow range makes generalizations about such a relationship 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 behaviors (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.

Anderson (1969) points out that the prime advantage in international cooperation in educational research lies in overcoming undue generalizations or 'under-generalizations' as well as distortive cultural bias.

> "Scientific research in education, as in the behavioral sciences in general, is a search for empirically valid and theoretically interesting generalizations about the behavior of human beings. This search is hampered by many obstacles, not the least of which is the problem of cultural bias and distortion. These problems are illustrated by two types of errors ... one is the error of 'over-generalization'. We assume that what we discover to be true of learning-teaching behaviors of some part of human species is true of the behaviors of all of the species, when in fact it is not."

> "A second error is found in our tendency to 'under-generalize'. In this case we assume that what we discover to be true of the behavior of some given part of mankind is uniquely true of only that part, when in fact what is true of the part is also true of the whole. Thus, the search for reliable knowledge about the process of human education in large measure is a matter of progressively eliminating generalizations which erroneously assume either more or less communality in our species' learning-teaching behaviors than do in fact exist."(Anderson, 1969, p. 144).

The replication aspect of cross-cultural research in education is also emphasized by Gage (1963), who hopes that by advancing theory in education it might be possible to identify 'laws' or principles of teaching that would cut across subject areas, grade levels, and teacher categories. One could in this context refer to the model of teaching advanced by Beeby (1966) which is an attempt to relate the level of development of formal schooling to the overall level of development reached by the nation where the teaching takes place. Another illustration might be Flander's study of teaching behavior and student achievement in Minnesota and New Zealand which was conducted on the hypothesis that such 'laws' or principles could be identified (Flanders, 1970).

After the completion of the IEA Mathematics survey (Husén, 1967a), two international meetings were held under the title "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 more powerful theoretical framework for the next survey, where achievement criteria from six subject areas were going to be developed. Researchers from 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 in this connection to suggest the inclusion of independent variables of 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 tests, has to do with comparability per sé (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, page 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, such that each student has a specified non-zero chance of entering into the sample, 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, for instance, countries like Iran and India could 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.

One main 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 that there were considerable differences between countries in terms of the amount of exposure the students had had to teaching of the various topics covered by the items in the international mathematics tests. Truly enough, country means of teachers' ratings of 'opportunity to learn' and student achievement tend to be rather highly correlated over countries (see, e.g. Comber and Keeves, 1973, page 158 et seq.). But it should be kept in mind that rank order correlations between country aggregates could be quite high, and they are indeed. 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 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 construction of international achievement tests and the machinery that went with it in a way served as a safeguard against undue cultural bias. An international committee was set up for each of the subject areas included in the Six-Subject Survey. 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, page 27 et seq.). Contact with the participating countries was effected through the National Research Centers and subject committees set up in each country. The analyses of the curricula, the proposing of item material and the try-out of the items were carried out in the participating countries. The IEA Headquarters served only as a co-ordinating center and a clearing house.

Furthermore, since evidently the main purpose of achievement tests is to measure differences in achievement, complete equality in terms of exposure to teaching and opportunity to learn would make the administration of such tests pointless. The same applies to intelligence tests, where individual and group differences unavoidably also reflect differences in terms of opportunity. As has been spelled out in another connection (Husén, 1967b), the administration of achievement tests internationally differs only in degree and not in principle from the administration of them nationally. Within a given country there are differences between school districts and regions due both to differences in student background and school resources. Very few 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 objectives 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 socially privileged have higher scores than underprivileged is per sé not to be interpreted as an act of discrimination against those who socially and pedagogically have been subjected to the less favorable conditions. The establishment of the factual differences in terms of these criteria, once the latter have been agreed upon, is in itself of informative value. It can, as in the case of the IEA research, serve as a basis for analysis of

what the factors are that account for differences in performance and ultimately can be used for more adequate educational policy. The data collected can also serve as a basis for evaluating how far the 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-industrialized and more or less agricultural economies, in brief to comparisons between developed and less-developed countries (LDC). 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, sometimes are spectacular.

One might well raise the question of the worthwhileness of an elaborate exercise like the one pursued by IEA to develop international standards of evaluation, considering the tremendous differences between the two categories of countries in terms of culture and tradition. But if the goal in the LDC's is to achieve 'modernization', i.e. among other things, to bring about an infrastructure of knowledge and skills conducive to an economic development which has led to affluence in the industrialized countries, then there is much to be said for attempts to measure, for example, basic reading skills and the knowledge in Science that is basic to the creation of a modern technology.

II. ORGANIZATION OF INTERNATIONAL EVALUATION OF EDUCATIONAL OUTCOMES

To conduct multi-national evaluation surveys is 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 centers have to take decisions about subject areas and problems they want to investigate. A uniform design guiding the construction of instrument, 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 (Foshay, 1962) 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 appropriate to describe briefly its organizational 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 'standardized' 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 (eight West European countries, the United States, Israel, Australia, and Japan). Student achievement in Mathematics was chosen as the 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 one hundred per cent of the relevant age-group was still in full-time schooling; and

(b) pre-university grade students.

In all 133,000 students from 5,400 schools were tested and completed questionnaires in the 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 organization 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 organizational features (Postlethwaite, 1967), were also conducted.

After the completion of the feasibility study and the first main study (in Mathematics) the participating research centers 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'. According to the statutes its principal aims are:

(a) to undertake educational research on an international scale;

(b) to promote research aimed at examining educational problems common to many countries in order to provide evidence which can help in the improvement of educational systems; and

(c) to provide, within the framework of the Association, means whereby research centers, which are members of the Association, can undertake co-operative projects.

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 Organization (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, when they were moved to Stockholm and are at present accommodated within the Institute for the Study of International Problems in Education. 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 Center. The number of members is presently 23, consisting of ten West European countries (Finland, Sweden, Federal Republic of Germany, Scotland, England, Ireland, Netherlands, Belgium, France, and Italy), three East European countries (Poland, Hungary and Romania), and ten non-European countries (Israel, Iran, India, Thailand, Australia, New Zealand, Japan, Chile, 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 center 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 of research. As mentioned above, one international committee in each subject area in which survey research is undertaken is appointed by the Council. Further, the Council set up a Technical Committee which was responsible for overall decisions taken on technical problems pertaining to sampling, data collection, 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 250,000 students, 50,000 teachers, and 9,700 schools in some 20 countries were involved in testing and completion of questionnaires. The data were made available to the data processing center 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.

III. SOME MAJOR FINDINGS IN THE IEA SIX-SUBJECT SURVEY

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 at the time of testing;

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

Population IV - all students in the terminal year in full-time secondary school programs which were either pre-university programs or programs of the same length (this gave the National Centers some latitude of interpretation, which means that in some countries only those students who are 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 programs 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 occasion when qualitative comparisons between industrialized and LDC's have been made according to agreed-upon international yardsticks.

Table 1, on the following page, 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 industrialized and non-industrialized countries. The latter are consistently far behind the former in average achievement over subject areas and levels of schooling. In Science the LDC's 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.

Table 1

Mean Total Score and Standard Deviation in Science and Reading Comprehension Among 10-Year-Olds, 14-Year-Olds, and Pre-University Students in 19 Countries.

	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
Fed. Rep. Germ.	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	9.1
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	8.5	8.3	7.6	9.0	6.6	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[1]	9.9	6.5	15.6	8.1	12.4	6.1	--	--	--	--	--	--

[1]Thailand did not test national sample but sampled schools in the Bangkok area.

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 industrialized 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 with regard to their school systems all are on the same level of 'efficacy'.

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 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 his or her class 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). But 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 which not the least is shown by the consistently substantial correlations between various family background measures and student achievement in all subject areas. Passow, Noah and Eckstein (in press) have, in their report on the National Case Study Questionnaire, 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 U.S. $1,400 - 4,300 in the industrialized countries, whereas it varies from U.S. $90 - 270 in the LDC's which are in the study. The size of the non-primary sector 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, which will be dealt with below, 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 figure for the 14-year-olds are 42 and 26 per cent respectively. What is then 'family background'? 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, page 195). Thus, regardless of the quality of the formal educational system in the the LDC's, we can, on the basis of the impact of the family background factors, predict a large difference in mean achievement between them and the more industrialized 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 color 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 doubts about whether quite a few of the 10- and 14-year-olds in the LDC's had been able to read the Science items and the questions in the student questionnaires.

IV. THE RELATIVE 'EFFECT' OF HOME AND SCHOOL

The Coleman study on "Equality of Educational Opportunity" (1966) was a massive attempt to disentangle the unique 'effect' of the school as compared to the home on student achievement. Notwithstanding the doubtful quality of the criteria of outcomes of instruction, such as a simple reading test which happened to be taken from a subject that, to a rather limited extent is 'school-based', the study gave rise to an intensive technical debate with criticism of the causal

ordering of variables in the regression analyses. The limitations of using cross-sectional data for 'effect-studies' of this type was also pointed out. The IEA Six-Subject Survey by and large falls victim to the same criticisms but can claim the following virtues. In the first place, the multivariate analysis has been conducted over a series of national educational systems and at different levels of the systems. Furthermore, which seems more important, the study covers a wide array of subject areas, both those which a priori can be regarded as highly school-based, such as foreign languages and Science and those which are less school-based such as Reading.

In the first place, the total variance accounted for was consistently larger than in previous studies to which reference has been made in the debate on the relative effect of home and school (Coleman, 1966; Jencks, 1972). Secondly, the school factors or 'learning conditions' at school tended to be of increasing importance as one moved from lower to higher levels in the system. Finally, Reading tended to differ considerably from Science and foreign languages in terms of the role played by the home. As pointed out above, this would seem to be the main explanation why the gap between the LDC's and the industrialized countries is larger in Reading than in other subject areas. Thorndike (1973, p. 177) sums it up like this:

> "A dominant determiner of the outcome from a school in terms of reading performance is the input in terms of students that go to school. When the population of a school comes from homes in which the parents are themselves well educated, economically advantaged, and able to provide an environment in which reading materials and communications media are available, the school shows a generally superior level of reading achievement."

Rank order correlations between means should, as indicated above, be looked upon with suspicion and interpreted with great caution, since they tend to boost heavily relationships that are much weaker at the level of the individual. But the following series of rank order correlations between mean achievement in Reading and various home background factors in the 15 countries which participated in the Reading survey casts some light on the statement quoted above and which was based on a broader spectrum of evidence:

Father's education	0.60
Mother'education	0.73
Expected (own) education	0.67
Parents' help with homework	0.50
Parents' encouragement to read	0.56
Number of books at home	0.85
Number of magazines at home	0.71
Hours listening and watching radio/TV	0.92

For the IEA-Harvard Graduate School of Education meeting on the implications of the IEA findings Coleman (1973) collated the outcomes of the between-student analyses reported in Comber and Keeves (1973), Purves (1973) and Thorndike (1973) for the six countries which tested both 10- and 14-year-olds in all the three subjects which were covered in Stage II (Reading, Science and Literature). It should be mentioned that the Literature score refers to the ability to comprehend literary prose. A comparative study of the outcomes of the between-student multivariate analyses is, as was pointed out earlier, of greatest interest because of its replicative nature. Parallel analyses have been conducted in a variety of countries which provide a broader perspective and facilitate meaningful interpretations.

Table 2

Relative Contribution of Home and School Variables in Accounting for Between-Student Differences at the 10-Year and 14-Year Old Level

	Chile		England		Finland		Italy		Sweden		U.S.		Average	
	10	14	10	14	10	14	10	14	10	14	10	14	10	14
Total Home Background Effects														
Science	0.20	0.36	0.46	0.48	0.37	0.47	0.20	0.32	0.40	0.42	0.42	0.47	0.34	0.42
Reading	0.12	0.45	0.47	0.52	0.42	0.45	0.31	0.32	0.34	0.40	0.45	0.47	0.35	0.44
Literature	--	0.38	--	0.50	--	0.43	--	0.33	--	0.39	--	0.43	--	0.42
Total Direct School Effects														
Science	0.30	0.26	0.18	0.30	0.21	0.34	0.20	0.26	0.23	0.28	0.32	0.28	0.24	0.29
Reading	0.29	0.28	0.13	0.19	0.18	0.23	0.22	0.19	0.18	0.18	0.21	0.28	0.20	0.22
Literature	--	0.32	--	0.22	--	0.26	--	0.18	--	0.26	--	0.30	--	0.26

Source: Coleman (1973)

It has been indicated above that there is some consistency among the five more developed countries that home effects account for more than school effects. As far as the 10-year-olds are concerned that does not apply to Chile. As can be studied in more detail in Professor Thorndike's report (1973, page 88 et seq.), the R-values and the per cent of added variance for Block I in the regression analysis (home background, age and sex) are much lower, particularly in India and Iran, than in the other countries. This indicates a relatively greater importance of school factors in these countries as compared with the richer ones.

V. EDUCATIONAL OPPORTUNITIES AND RESOURCE ALLOCATION

Since information was available on parental occupation and parental education, a comparative study could be made on the degree of equity that went into a national system, or, conversely, how priorities were reflected in the social selection that took place when the students moved up to the pre-university level.

One overriding educational policy problem in all LDC's has been to what extent and how fast formal school education should be made universal and how much education, i.e., how many years of schooling that could be provided to how many students. This problem can be resolved in a more egalitarian or a more elitist direction. It has, among educational planners in LDC's, often been advocated that in the long run the educational system would provide a better yield if the scarce resources were not spread thin and (at least in theory) made available to all children at primary school age. One should give first priority to educating an elite which would then build up the infrastructure needed for universalizing primary education in a remote future.

In attempting to evaluate a national school system one or more of the following criteria could be employed. One could try to assess to what extent the system is taking care of the most able, the average and the less able students. One could look at the attrition rate in terms of grade-repeating and drop-out, which usually is very high in most LDC's. One could assess student attitudes toward further learning and try to find out how motivated they are. One could follow the students up through the system and assess how open or closed the system is in terms of options between types of programs and tracks. One could try to measure the amount of social bias that goes into the processes of attrition and selection.

The IEA data lend themselves to elucidate one major aspect of the problem of universalization vs. elitism or selectivity, namely the amount of social bias that goes into the selection procedure and the standard of the elite in a selective as compared to a more comprehensive or universal system.

By comparing the distribution of socio-economic status, as indexed by father's occupation, for the 10-year-olds with the one for the 14-year-olds and the pre-university students respectively, we can make an estimation as to what extent the selection that operates from one level to another is correlated with social background. As far as the industrialized countries are concerned, the overall outcome of the analyses is this (Husén, 1973). In national systems with high retention at the secondary level selectivity on social basis is less predominant than in systems with low retention rate and more strict selectivity. On the basis of the proportion of upper and lower stratum representation at the 14-year-olds level (when in the industrialized countries practically all children are still in full-time schooling) and at the pre-university level respectively, an index of social bias can be calculated (Husén, 1973). This index is unity when upper and lower strata have equal representation. It turns out to be 1.3 for the United States and 2.4 for Sweden, two countries with relatively high retentivity (75 and 45 per cent of the relevant age-group still in school). Social bias in the enrollment at the senior secondary school in England is 7.9 and in the Federal Republic of Germany as high as 37.7, two countries where the retention at that level is relatively low (20 and 9 per cent respectively).

The students were, according to parental occupation, classified in nine categories. The classification scheme, which has originally been developed by the International Labour Organization in Geneva, could, however, not be employed uniformly over all the countries (Comber and Keeves, 1973). Therefore, to the extent that the categorization has been consistent within the countries, comparisons can be made between various levels of the system in terms of the social structure of enrollment. The high proportion of fathers with professional and clerical occupations in the LDC's

Table 3

Family Background (in per cent) in Terms of Father's Occupation of Students at Various Levels of the Educational System in Chile, India, Iran, and Thailand

Occupational Category	Chile Age			India Age			Iran Age			Thailand Age		
	10	14	pre-univ.	10	14	pre-univ.	10	14	pre-univ.	10	14	pre-univ.
Professionals and Managerial	4	8	19	8	10	16	20	20	24	3	3	9
Clerical	19	21	34	24	22	27	19	20	19	30	33	34
Skilled manual	30	27	17	52	52	43	46	44	42	45	46	53
Semi-skilled and Unskilled-manual	33	28	8	6	3	5	7	6	8	18	8	3
Unclassified	14	16	22	10	13	9	8	10	7	4	8	2
Total	100			100			100			100		

indicate that those children who on the whole enter school are a socially select group. This seems to be the case, for instance, in Iran. This also explains why in Iran the social composition of the pre-university students does not differ very much from the one at the primary level. The most marked social selection takes place in Chile, which differs from the other three LDC's in terms of the size of the non-primary sector of the economy. The percentage of the upper stratus increases from 4 to 19 when one moves from the 10-year-old to the pre-university level, at the same time the number of semi- or unskilled workers decreases from 33 to 8 per cent.

Evidently, when an evaluation is made of the standard of the elite students one has to take into consideration what proportion of the relevant age-group we are looking at. It is pointless to limit the comparison to the mean performance at that level, simply because we are dealing with a highly variable portion of the age-group. Among the IEA countries it varied in 1970 all the way from 75 per cent in the United States to less than 10 per cent in Iran. Therefore, it would be not only more 'fair', but also more informative to compare equal proportions of the age-groups. This has been done in Table 4, where we present the means for the entire samples at the pre-university level and the means for the top 9.5 and 1 per cent of the entire age-group. The comparison between countries in terms of total Science test score is based on the assumption that those who at this age level are not in school would not have scored in any of the three top categories had they been accessible for testing. There are indications that in the industrialized countries the means arrived at in the top groups would not have been significantly affected. This is even more valid for the LDC's.

Table 4

Means and Standard Deviations for Science Test Score For Total Pre-University Sample and Equivalent Proportions of the Relevant Age Group

	Per Cent of Age Group in School	Full Sample M	Full Sample SD	Top 9 per cent M	Top 5 per cent M	Top 1 per cent M
Grand Total Score for Industrialized Countries	30	22.0	10.6	32.3	37.1	45.9
Chile	16	9.3	6.3	13.6	16.8	23.5
India	14	6.3	6.1	9.5	12.8	20.8
Iran	9	10.8	5.9	10.8	14.8	21.9
Thailand	10	12.5	6.1	13.6	17.4	23.2

We notice in Table 4 that the average score for the total sample in all the industrialized countries is 22.0, which is more than one standard deviation above the average for the four LDC's. If we then look at the top 9 per cent and 5 per cent, we find that the difference becomes even more marked. The top 1 per cent of the students in the pre-university year in the LDC's score at the level of the average student in the industrialized countries. As far as Science is concerned the selection that has taken place in the LDC's from the lower to the higher level of the system does not seem to have considerably increased the 'productivity' at the upper level of the system.

VI. 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 types of 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 tests 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 and not 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, in many systems is 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 utilized routinely. Methods of analyzing 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, experiences have been gained in data processing that lend themselves to nation-wide evaluation surveys.

The IEA international headquarters, as well as the National Centers, have over the last ten years built up a considerable amount of collective competence with regard to the conceptualization 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 utilized to provide training programs for students from regions of the world where particular strengths and competencies in evaluation are still developing. From the IEA international network it would be relatively simple to set up task forces to work with centers in LDC's. Such forces could co-operate with local researchers on designing evaluation surveys.

REFERENCES

Anderson, L., "The importance of international co-operation in educational research", Educational leadership, 27:2, 1969.

Beeby, C. E., The quality of education in developing countries, Cambridge, Mass.: 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. et al, Equality of educational opportunity, Washington, D.C.: U.S. Department of Health, Education and Welfare, Office of Education.

———— Effects of school on learning: the IEA findings, Paper presented at Conference on Educational Achievement, Harvard University, November 1973.

Comber, L.C. and Keeves, J.P., Science education in nineteen countries, New York: 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.

Flanders, Ned A., Analyzing teaching, Reading, Mass.: Addison-Wesley, Series of Education.

Foshay, Arthur W. (ed.), Educational achievement of 13-year-olds in twelve countries, Hamburg: Unesco Institute for Education.

Gage, Nathaniel L. (ed.), Handbook of research on teaching, a project of the American Educational Research Association, Chicago: Rand McNally, 1963.

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 and Almqvist & Wiksell, 1967.

———— "Establishing international criteria of comparison in comparative education" in General education in a changing world, Amsterdam: Nijhoff, 1967b.

———— Svensk skola i internationell belysning: naturorienterand ämnen (The Swedish school system in international perspective: science), Stockholm: Almqvist & Wiksell, 1973a.

———— Implications of the IEA findings for the philosophy of comprehensive education, paper presented at the Conference on Educational Achievement, Harvard University, November 1973 (Harvard Graduate School of Education), 1973b.

Jencks, Christopher et al, Inequality: a reassessment of the effect of family and schooling in America, New York: Basic Books, 1972.

Lewis, E. Glyn and Massad, Caroline (in press), English as a foreign language in ten countries, Stockholm and New York: Almqvist & Wiksell and Wiley-Halsted Press.

Marklund, Sixten, Skolklassens storlek och struktur (Size and structure of the school class), Stockholm: Almqvist & Wiksell, 1962.

Passow, A. Harry, Noah, Harold and Eckstein, Max (in press), National case studies: 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.

———— 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, New York: 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, New York: Wiley-Halsted Press, 1973.

OCCASIONAL PAPER No. 37 highlights some of the main conclusions from the Six-Subject Survey conducted by the International Association for the Evaluation of Educational Achievements (IEA). It sets out the aims of conducting such multi-national comparisons and discusses the future of evaluating national systems of education, with particular reference to developing countries.

TORSTEN HUSEN (Sweden) is Professor of International Education and Director, Institute for the Study of International Problems in Education, University of Stockholm. The author of some 50 books and a considerable number of research papers, Professor Husén is also Chairman of the Governing Board of the I. I. E. P. and Chairman of the International Curriculum Organization.