

Conceptualizing a comparative educational research framework - Tjeerd Plomp

Tjeerd Plomp (The Netherlands). *Professor of Education at the University of Twente (The Netherlands) and Chairman of the International Association for the Evaluation of Educational Achievement (IEA). Former Dean and former Director of the Institute for Applied Education Research, Department of Education, University of Twente. Author or co-author of numerous publications on educational technology, evaluation and international comparative research, and the use of computers in education. His most recent books include The Use of Computers in Education Worldwide (1991) (in English, with W. J. Pelgrum) and Design of Education, and Training (1992) (in Dutch).*

Why comparative international research?

Why should countries participate in international survey research? Would it not be more beneficial for a country to test educational achievement and educational processes in the context of its own educational system, instead of spending much time and energy on costly international comparisons? Are fair international comparisons indeed possible? Are not the well-known educational statistics on fiscal and other resources, such as numbers of teachers, student participation, numbers of graduates and the like, sufficient for policy-makers' purposes?

While answers to these questions are debatable, the most important question is indeed the 'why' question. Of course, nobody doubts the usefulness of national or regional comparisons (state by state, province by province), or the relevance of 'traditional' educational statistics. But there are at least two good reasons 'why' countries or educational systems (both terms will be used interchangeably) should participate in international comparative surveys which measure student achievement.

The first and most important reason is to improve understanding of educational systems. Since there are no absolute standards for educational achievement, comparative studies are essential to provide policy-makers and educators with information about the range of educational quality in relation to other national systems. In this way studies contribute to setting realistic standards for educational systems, as well as monitoring school quality.

The second reason is that comparative studies may also be helpful in understanding the causes of observed differences in student performance, by exploring cross-nationally relations between school achievement and such factors as curricula, amount of time spent on school work, teacher training, class enrolment, parental involvement and many other possible explanatory measures. These comparisons can be twofold: straight comparisons of effects of education, using total scores or subscores on international tests, which directly reflect the curricula of participating countries; and comparisons of how well a nation's intended curriculum ('what should be taught in a particular grade') is implemented in the schools and mastered by pupils.

The International Association for the Evaluation of Educational Achievement (IEA) collects the sort of data policy-makers can use as a basis for decision-making to improve education. This paper briefly describes IEA's mission and history, the design and structure of a typical IEA achievement study, and provides some examples of comparative outcomes obtained which identify student achievement patterns.

IEA: mission and history

IEA is an independent international co-operative of research centres from 45 countries, which any country may join. It has taken as its mission the conducting of comparative studies focusing on educational policies and practices in order to enhance learning within and across systems of education. IEA has committed itself to a cycle of studies of learning in the basic school subjects and to additional studies of particular interest to its members. In its studies, it focuses not only on measuring educational achievement, but also looks at the effects of the curriculum and the organization of schools and classrooms on learning; the relationship between achievement and attitudes; the effects of certain subject-matter practices, such as laboratory work in science, and time spent in class, composition teaching in the mother tongue; different educational practices; and the attainment of special groups.

After its foundation in 1959, IEA's first international study of educational achievement was in elementary and secondary mathematics education (13 countries); this was completed in 1966. IEA was the first international organization that used the same objective cognitive tests in more than one country. Other studies were the Six-subject Survey (in science, reading, literature, civic education, and English and French as foreign languages), with a varying number of countries, from 8 (French) to 19 (science), completed in 1973-74; the Second International Mathematics Study (20 countries; data collection in 1981); the Classroom Environment Study (10 countries; 1982); the Written Composition Study (13 countries; 1984, 1985); and the Second International Science Study (24 countries; 1983). Ongoing studies are the Pre-primary Project (15 countries), the Computers in

Education Study (20 countries); the Reading Literacy Study (31 countries), and the Third International Mathematics and Science Study (more than 50 countries; data collection in 1994 and 1998).

The *IEA Guidebook* (Hayes, 1991) describes the activities, institutions, and people involved in IEA. A special issue of *Comparative Education Review* (Postlethwaite, 1987b) focused on the first twenty-five years of IEA research. IEA studies appeal to a wide range of audiences including: (a) policy-makers in education, who wish to investigate linkages between indicators of school performance and, for example, economic growth and labour productivity; (b) educational researchers and funding agencies, who wish to employ improved measures of educational performance and other educational indicators; (c) curriculum specialists and teachers, who wish to improve their understanding of the nature of the relationship between curricula and student outcomes.

Benefits of IEA participation

Countries or educational systems which participate in IEA studies benefit in several important ways:

They obtain valid national results in the context of international comparisons, which form a basis for decision-making.

They assist in understanding observed differences between one country and other educational systems.

They train national research teams in survey research.

They receive an evaluation design for national assessment.

They gain access to an international network of experts in research and education.

Design of a typical IEA study

Over the past three decades, IEA studies have varied in form and content, but overall they have retained curriculum-based, explanatory designs employing large-scale survey methods. Most IEA studies have tested national samples of age or grade in school populations of 9-10, 13-14 and 17-19-year-olds. A typical IEA study starts by developing a conceptual framework that clarifies the issues to be addressed, suggests appropriate methods of investigation - which results in validated measures of educational outcomes and processes - and uses those analytic tools that can best elucidate key factors and issues.

Conceptual framework

As IEA studies are designed to contribute to the explanation of differences in learning outcomes, those variables chosen must be derived from a conceptual model of how educational systems work and reflect the curriculum-based character of learning. A typical example of an IEA framework is the one applied in the Second International Mathematics Study (SIMS) found in Travers and Westbury (1989). Figure 1 illustrates how the curricula can be examined at the system, the school/classroom and the student level.

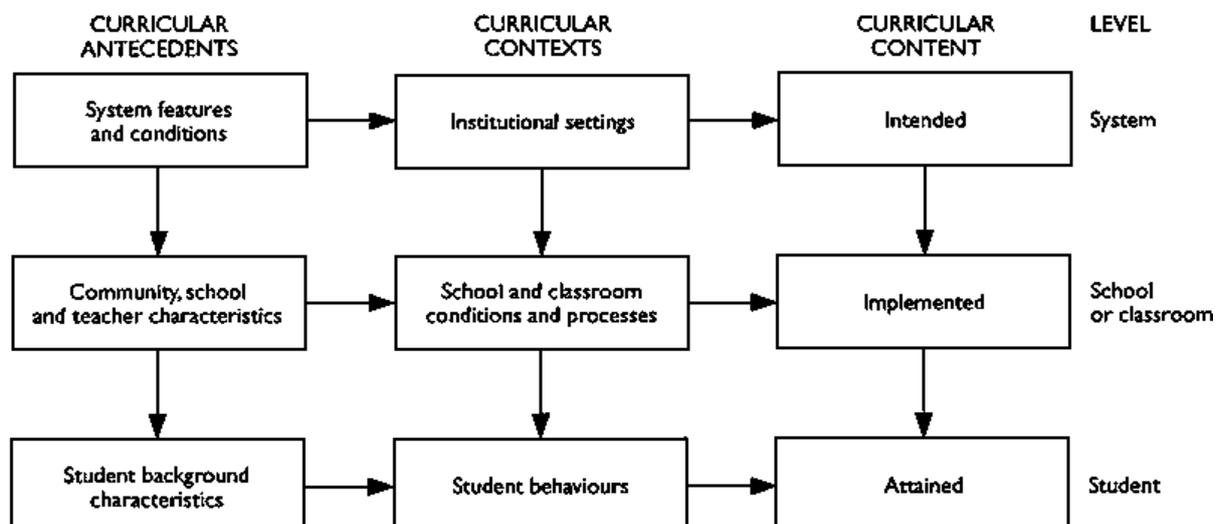


FIG. 1. An IEA research model

At the macro level of the educational system (country, region, school district), there is the set of intentions found within the curriculum: the official goals and the ideas and traditions of subject-matter specialists and educators. This collection of intended outcomes, together with course outlines, official syllabi and textbooks, forms the intended curriculum.

At the level of the school and the classroom, one finds the implemented curriculum being taught by the teachers. Finally, we have the attained curriculum represented by knowledge, skills and attitudes acquired by students and measured by tests.

This conceptual framework allows us to generate important questions when monitoring the quality of education. For example, how closely does the implemented curriculum match the intended curriculum: Do teachers teach the syllabus? Is teacher perception of the intended curriculum in agreement with the intentions at the system level? Are the intended topics covered, or are some left out because of, say, time constraints? Equally important is the question of possible discrepancies between the intended and the attained curriculum, that is, whether the students learn what they are expected to learn according to teacher intentions; and, if this is not the case, whether the explanation can be found in discrepancies between curricula.

Curricular differences refer to different curricular contexts, such as the range and depth of topics, organization and rules governing who can study what topic in what order, etc. Furthermore, background or antecedents of the curriculum influence the curricular contexts, as well as the curricular contents. For example, at the system level, the wealth of a society may affect the retentiveness of the school system; the community and family wealth of students may influence the amount of mathematics taught and learned.

A model or framework such as the one given in Figure 1 provides the starting-point for the further design of the study. The rows indicate the level at which relevant data can be collected, and hence indicate the nature of variation in the data. For example: sources of variation may be found (a) between educational systems (for example Japan and Germany); (b) between schools (comprehensive and vocational schools); (c) between classrooms (remedial versus advanced classes); and finally (d) between students (rich and poor, rural and urban, boys and girls, highly or poorly motivated).

From a general research model it is possible to develop and refine predictive models that incorporate data collected from schools, teachers and students into one large paradigm which includes background influences on schooling, school processes and student outcomes. Figure 2 illustrates a model employed by the IEA Reading Literacy Study, which links the many facets that influence student reading.

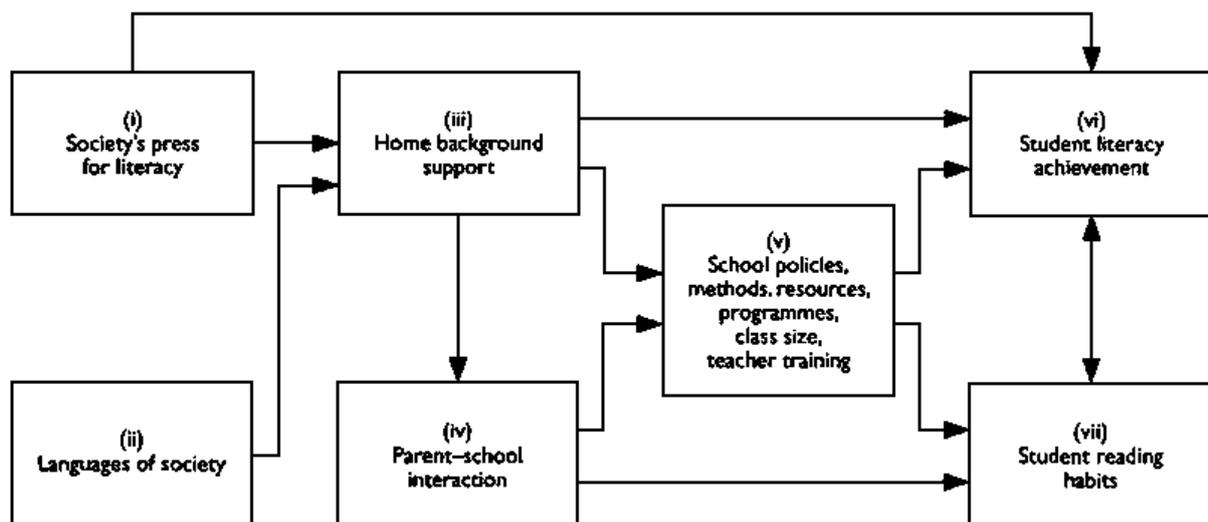


FIG. 2. A reading literacy model.

In this model, background variables are used as control variables to examine the effects of school indicators and school/teacher policy variables on student outcomes. This general model allows a wide range of hypotheses to be tested within and across educational systems. From such a model, it is possible not only to develop a baseline database to estimate the literacy levels of schoolchildren, but also identify which school, teachers and societal factors influence literacy and to what extent this occurs.

Population and samples

Different studies may be directed at different populations. To relate classroom and teacher variables to achievement measures, in a typical IEA study, the population definition is given in terms of age and grade. For example, in the Second International Mathematics Study (SIMS), the population of lower secondary education was defined as all students in the grade in which the modal number of students has attained the age of 13.0-13.11 years by the middle of the school year. Such a population definition allows each participating system to determine a grade level, within which classrooms, and through them teachers and students, can be selected. IEA achievement studies usually have three populations, at the elementary, lower and upper secondary education level respectively.

Careful sampling procedures are applied, striving for the same level of quality control across all participating countries. Each study has a sampling referee - a specialist who determines the actual sampling procedures - who has to approve the sampling plan submitted by each country, and who has to create acceptable solutions for practical problems (see article by Ross hereunder). Such procedures give optimal assurance that the data collected for each nation indeed provide reliable estimates on which to build sound probabilistic samples.

Academic achievement and opportunity to learn (OTL) questionnaires

In every study a variety of tests and questionnaires are developed. For example, in the SIMS several questionnaires were applied: school organization questionnaires; teacher questionnaires on educational job history, attitudes and teaching practices; student questionnaires on background and attitudes; and achievement tests. In the past, many achievement tests consisted predominantly of multiple-choice items. In some cases IEA experimented with other item formats, such as open-ended mathematics items, essays in the mother tongue, performance tests for science and computers in education (word processing), and oral testing in foreign languages. In the Third International Maths and Science Study a variety of item formats will be applied when data are collected in 1994 and 1998.

In all IEA studies, teacher ratings are collected: these ask whether the content needed to respond to each item on achievement tests has been taught to students that year, in prior years, or not at all. This measure is called opportunity to learn (OTL). It provides researchers with a second measure to compare student outcomes, using a score, adjusted for OTL, next to the raw score achievement tests. When reporting national results, whether one uses raw achievement scores or adjusted ones depends on the purpose of the researcher. If policy-makers are intent on comparing their country with other countries, they may be interested in raw test scores.

Generally, IEA held back reporting only raw scores, because such scores easily make a study into a kind of 'Olympics' or 'horse race'. On the other hand, scores adjusted for OTL may provide more insight into the quality of the educational processes. OTL offers, in principle, an important tool for carrying out nuanced comparisons which tell why children do poorly on tests or test items. This can be illustrated by drawing on an example from the French test in a Foreign Language Study (from Postlethwaite, 1987a). In any foreign-language testing it is usual to test the four skills of reading comprehension, listening comprehension, writing and speaking. It is also possible that these four skills are emphasized differently in classrooms by teachers depending on either the intended national curriculum, or on how well teachers themselves have mastered each skill. Speaking a foreign language is a skill that carries great weight in some countries but is of little importance in others. Therefore, would it be fair to compare spoken proficiency in Country A, with high opportunities for speaking, with scores in Country B, which places low emphasis on speaking outcomes? Perhaps scores should be adjusted automatically for opportunity to learn, and then profiles across all major skill subtests reported so strengths and weaknesses can be viewed side by side.

Development of achievement tests

An important aspect of making valid international comparisons of achievement concerns accuracy in judging the fairness of the achievement tests for all participating students. Test development at least needs to reflect the intended curricula in participating countries. The procedure for the development of achievement tests, which IEA follows to obtain optimal fairness, is based on a careful curriculum analysis. In the Third International Mathematics and Science Study, this curriculum analysis is a study in itself (see the article by Schmidt on curriculum mapping). We can illustrate the steps in the development of achievement tests with the SIMS.

The starting-point is to see that tests reflect as well as possible the intended curricula of all participating educational systems. On the other hand, tests need to display across countries sufficient commonalities in mathematical content, so that meaningful comparisons between the participating systems are possible. In earlier studies, the organizing framework for this was a content-by-behaviour grid. For example, the content dimension for the population of 13-year-olds consisted in SIMS of 133 entries under five broad categories: arithmetic, algebra, geometry, statistics and measurement. The behaviour dimension was divided into four levels: computation, comprehension, application and analysis.

Each participating system was asked to describe its curricula using the format of this grid. For each cell in the grid, national centres were asked to report on: (a) coverage (i.e. whether the mathematical skills and knowledge defined by the cell were part of the curriculum for 'all', 'some' or 'none' of the students in the system); (b) emphasis given to each topic of the curriculum; and (c) the importance of each topic found in the grid of the curriculum within each country. National centres were also requested to supply items which they considered appropriate for ten preselected cells in each grid. On the basis of the completed returns received from each educational system, the International Mathematics Committee of the study discussed the differences found and reached a consensus. An example of such a rating is the grid for algebra found in Figure 3.

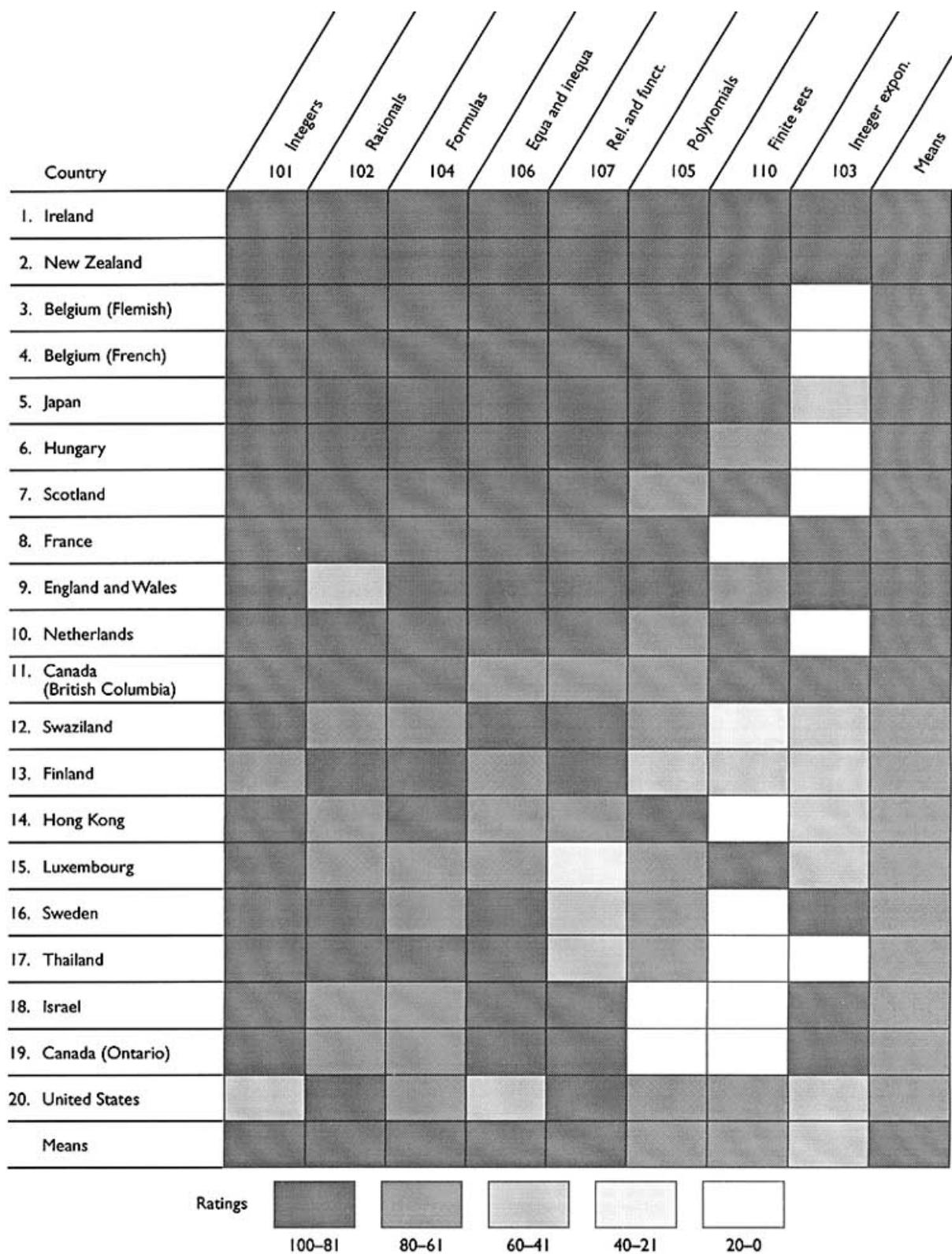


FIG. 3. The match between national and international algebra grids (after Travers and Westbury, 1989).

Given the differences in ratings between national curricula, it is clear that the curriculum of a particular educational system will be reflected in the international grid more or less equally, depending on the extent to which the system uniquely covers all topics of mathematics taught to specific age/grade groups. As the grid serves as the basis for test construction, it is important to get some information on the 'goodness of fit' between the grid and each system's curriculum. This information is obtained by comparing the international grid with each system curriculum as described above, which can then be used to interpret national test scores in the context of an internationally scaled test covering internationally agreed-upon test items.

Exemplary outcomes of IEA studies

Results of many IEA studies are reported in numerous scholarly volumes. In many other publications are reports about national and international results obtained from these and previous studies (Degenhart, 1990). Several contributions in this and the following issue of *Prospects* provide examples of how national achievement data can be interpreted within an international context. I will restrict myself here to some illustrative examples of how IEA data can be analysed and used.

TOTAL TEST SCORE AND OPPORTUNITY TO LEARN

As stated earlier, until recently IEA has been hesitant to publish raw national test scores. The publication of such comparisons does not do justice to the complexity of educational systems and extenuating circumstances facing each individual. Furthermore, such 'Olympics' or 'horse races' clearly do not serve research purposes. It is for this reason that, for example, Robitaille and Garden (1989) did not report the outcomes of school mathematics using raw scores. Clearly, IEA is not interested in newspaper headlines such as 'Sweden Excels Norway', or 'USA Ranks Poorly' (see Postlethwaite, 1987a).

On the other hand, policy-makers are interested in some total score international yardstick. This is clearly illustrated by President Bush of the United States who, in his 1990 State of the Union Message, announced as one of the six national goals for education that 'by the year 2000, US students will be first in the world in science and mathematics achievement'. The speech went on to explain this goal as follows: '... while no international comparisons of student achievement to date are considered adequate, available measures suggest that US 13-year-olds perform near the bottom in science and mathematics compared to their peers in other industrialized countries' (White House, 1990). The document further stated the need for a permanent international framework for coordinating international assessments that compare the performance of United States students in mathematics and science to that of their counterparts in other industrialized countries. In this context, reference was made to the IEA Third International Mathematics Study, which was at that time in preparation.

If IEA wants to ensure relevancy to policy-makers, then it must publish test results. But by using the opportunity to learn measure to adjust differences to total scores, it is possible to present a score that reflects student exposure to subject-matter.

TABLE 1. Total raw achievement means and adjusted means for opportunity to learn

Country	Raw mean	Adjusted mean	Adjustment
Australia	60.25	60.47	0.22
Canada (English)	60.50	60.38	-0.12
Canada (French)	57.25	59.65	2.40
China	63.25	59.07	-4.18
England	55.50	56.02	0.52
Hungary	70.50	64.24	-6.26
Italy (Grade 8)	52.50	53.60	1.10
Italy (Grade 9)	62.00	64.50	2.50
Japan	66.50	64.26	-2.24
Republic of Korea	61.50	63.46	1.96
Netherlands	60.25	62.04	1.79
Nigeria	40.25	43.02	2.77
Philippines	40.25	40.23	-0.02
Singapore	57.75	59.06	1.31
Thailand	56.50	54.74	-1.76
MEAN	57.65	57.6	0.00

Source: Postlethwaite and Wiley (1992, p. 105).

COGNITIVE YIELD

In many IEA studies, the average level of student achievement across nations, weighted by the proportion of the student age cohort attending school, is calculated. When both level and proportion are considered together, it is possible to determine an average yield. This yield is the distribution of correct items made by the proportion of the age-group in school. For example, on a test of thirty items, if an entire age cohort of children were enrolled in school and all answered every item correctly, there would be a rectangular distribution (e.g. 100 per cent area under the curve in Fig. 4).

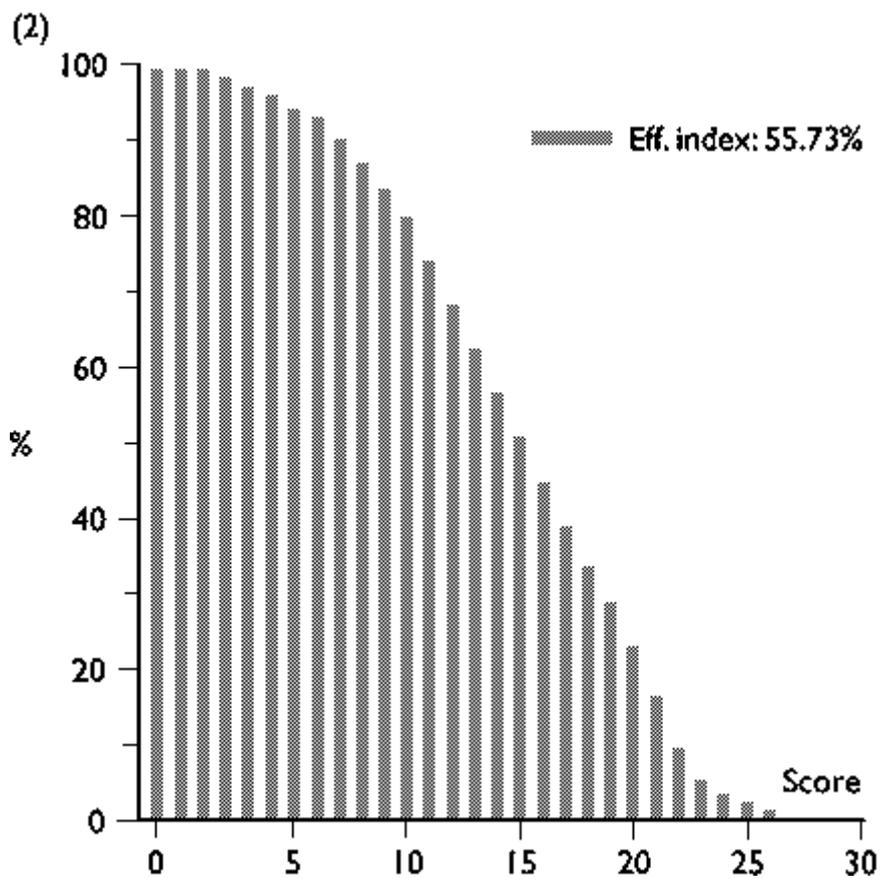
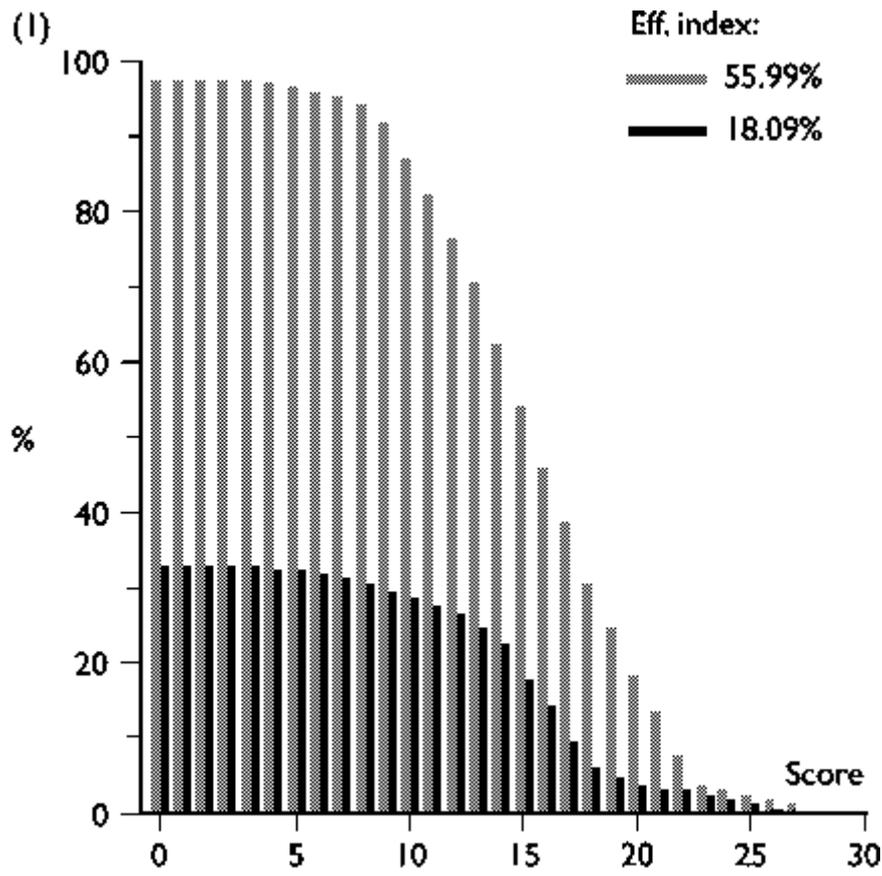


FIG. 4. Cognitive yield in Thailand (1) and the United States (2).

When the percentage of a test item answered correctly falls, so does the area under the curve (i.e. the cognitive yield). Figure 4 illustrates the distribution of science achievement scores for Thailand and the United States (IEA, 1988). The percentage of correct responses for both countries on the thirty-item test was the same, giving the appearance that achievement on the test in Thailand and the United States was the same. However, one out of three 13-year-olds in Thailand attended school in the early 1980s at the time of this study, while almost all 13-year-olds in the United States attend school. When the cognitive yield takes into account the percentage of an age cohort in school, the area under the curve in Thailand's case diminishes greatly. Based on measures of cognitive yield, educators in Thailand should improve the quantity of schools, and educators in the United States should improve the quality of education in their schools.

TRENDS IN ACHIEVEMENT OVER TIME

By planning its studies according to a cycle, IEA wants to chart and monitor for its members the distribution and change in achievement worldwide over time. An example of how this can be accomplished was calculated by Keeves (1992). Drawing on data from the First and Second International Science studies, this Australian researcher developed a science achievement scale that enables researchers to compare student performance from different countries at different age and grade levels, and different occasions on one international scale (see Fig. 5).

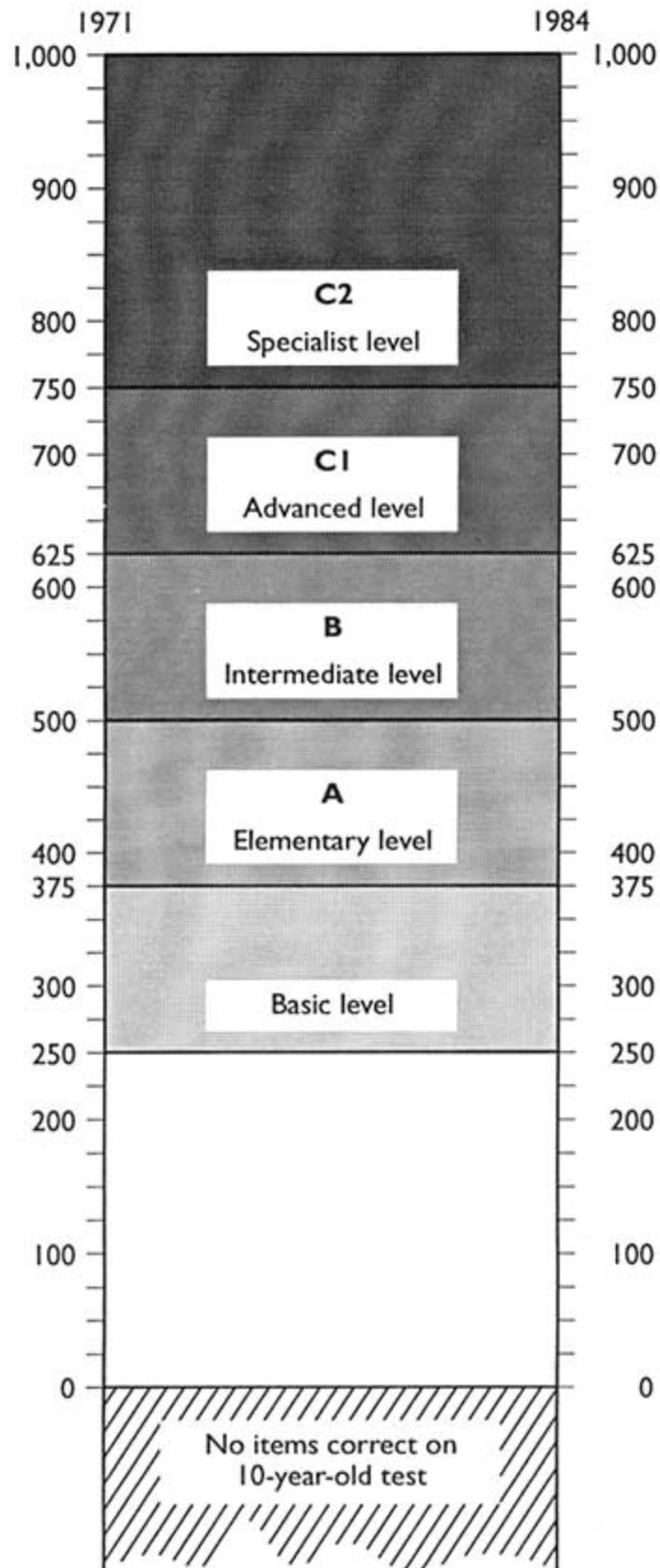


FIG. 5. An international science scale.

THE RELATION BETWEEN ACHIEVEMENT AND SCHOOLING

Secondary analyses of IEA data can provide many interesting results, of which some examples are mentioned.

Kifer (1989) concludes from student achievement analyses that what is taught is related to mathematics achievement differences, but how students are taught is not as important.

Carroll (1975) shows from the IEA study of French as a second language that the longer students study the subject, the more proficient they become. Amount of time can be reduced if the student has a teacher fluent in French, is taught mainly in French, and the student has serious aspirations to learn French.

Where Carroll's findings have implications for the allocation of time, teacher competence, and teaching methodology, the Kifer finding implies the relative importance of exposure to content as opposed to teaching and curriculum 'sequencing'. This suggests that exposure to the material is the necessary condition for learning; while sequencing, pacing and teaching variables follow in importance.

Turning to gender, the results from the SIMS (13-year-olds) tend to confirm earlier research showing boys to perform better on mathematics than girls (Robitaille and Garden, 1989). A more interesting outcome shows systems in which girls outperform boys on all mathematics subtests (Belgium-Flemish, Belgium-French, Finland, Sweden and Thailand). The interesting lesson to learn from such a finding is that a pattern considered as 'given' may not be so invariant after all. A closer analysis of countries where gender differences are not significant may suggest ways to change mathematics practice in countries where gender differences do exist.

References and bibliography

ANDERSON, L.; DORIS, W.; RYAN, W.; SHAPIRO, B. 1989. *The IEA Classroom Environment Study*. Oxford, Pergamon Press.

BURSTEIN, L. (ed.). 1992. *The IEA Study of Mathematics III: Student Growth and Classroom Processes*. Oxford, Pergamon Press.

CARROLL, J. B. 1975. *The Teaching of French as a Foreign Language in Eight Countries*. Stockholm/New York, Almqvist & Wiksell/John Wiley.

DEGENHART, R. E. 1990. *IEA bibliography*. The Hague, IEA Secretariat.

ELLEY, W. 1992. *How in The World Do Children Read?* The Hague, IEA Secretariat.

GARDEN, R. A. 1987. The Second IEA Mathematics Study. *Comparative Education Review*, Vol. 31, No. 1, pp. 47-68.

GORDON, T.; PURVES, A.; DEGENHART, R. 1988. *The IEA Study of Written Composition I*. Oxford, Pergamon Press.

HAYES, W. A. (ed). 1991. *IEA Guidebook 1991: Activities, Institutions and People*. The Hague, IEA Secretariat.

IEA. 1988. *Preliminary Science Results in Seventeen Countries*. Oxford, Pergamon Press.

KEEVES, J. P. 1992. *Learning Science in a Changing World: Cross-national Studies of Science Achievement (1970-1984)*. The Hague, IEA Secretariat.

KIFER, E. 1989. What IEA Studies Say about Curriculum and School Organization. In: A. C. Purves (ed.), *International Comparisons and Educational Reform*, pp. 51-72. Association for Supervision and Curriculum Development.

LUNDBERG, I. 1992. *The Teaching of Reading in 30 Countries*. The Hague, IEA Secretariat.

PELGRUM, W.; PLOMP, T. 1991. *The Use of Computers in Education World Wide*. Oxford, Pergamon Press.

POSTLETHWAITE, T. N. 1987a. Comparative Educational Achievement Research: Can it be Improved? *Comparative Education Review*, Vol. 31, No. 1, pp. 150-8.

POSTLETHWAITE, T. N. 1987b. *Comparative Education Review*, Special Issue, Vol. 31, No. 1.

POSTLETHWAITE, T. N.; ROSS, K. 1992. *Effective Schools in Reading: Implications for Educational Planners*. The Hague, IEA Secretariat.

POSTLETHWAITE, T. N.; WILEY, D. E. 1992. *The IEA Study of Science II: Science Achievement in Twenty-three Countries*. Oxford, Pergamon Press.

PURVES, A. 1987. IEA: an Agenda for the Future. *International Review of Education*, Vol. 33, pp. 103-7.

ROBITAILLE, D. F.; GARDEN, R. A. (eds.). 1989. *The IEA Study of Mathematics II: Contexts and Outcomes of School Mathematics*. Oxford, Pergamon Press.

TRAVERS, K. J.; WESTBURY, I. (eds.). 1989. *The IEA Study of Mathematics I: Analysis of Mathematics Curricula*. Oxford, Pergamon Press.

TORNEY, J. V.; OPPENHEIM, A. N.; FARNEN, R. F. 1976. *Civic Education in Ten Countries*. Stockholm, Almqvist & Wiksell.

WHITE HOUSE, THE. 1990. *National Educational Goals*. Washington, D.C., The White House, Office of the Press Secretary.