Evaluation of Education and Training: What Room for the Comparative Approach ?

George PSACHAROPOULOS
World Bank, Washington

Lauwerys Lecture

Abstract

The evaluation of education and training systems has intensified in response to the economic squeeze of the 1980s and to the ever increasing social demand for education. This paper examines a series of conceptual issues involved in evaluating school systems, such as the nature of the evaluation criteria, who should be the evaluator and what should be evaluated in the first place. It also discusses alternative methodologies, emphasizing the shift away from descriptive and towards statistical techniques. Since any rigorous evaluation must involve a comparison of some kind (against a reference standard or control group), attention is paid to the international comparative approach, using student achievement in particular as a key performance indicator.

International comparisons of student achievement do serve an important role, in the sense that they alert the policymaker to potential problems within an educational or training system. They also help to train local researchers in analytical (rather than descriptive) evaluation techniques. However, in today's world where survey data on individual students are becoming increasingly available within each country, the value of international comparisons as an evaluation tool might be a little *passé*. More emphasis could be put on cross-sectional or over-time comparisons within countries because this would achieve better control and standardization for the *other* factors that enter into the cause-effect relationship in which the evaluator is interested.

Author's address: The World Bank, Washington, DC 20433, USA. I am indebted to Norberto Bottani, Stephen Heyneman, William Loxley, Constantin Soumelis and Jack Schwille for providing me with key material while writing this paper, and to Andreas Kazamias and Zafiris Tzannatos for commenting on a first draft. The views expressed here are those of the author and should not be attributed to the World Bank.

Contents

List of Displays

I. Introduction

Gone are the days when an educational system could proceed unchecked on a momentum of its own. Following the economic crunch of recent years, "evaluation" has become the buzzword in Ministry of Education offices, state budget corridors and factory shop floors alike.

Why this sudden surge of interest in evaluating education and training programs? The explanation can be found in the simple arithmetic between the increasing demand for education and training on the one hand and the limited amount of resources to meet such demand on the other. As students and their families have demanded more and better education, mainly financed by the government or private firms, the way in which educational resources are used has come under more scrutiny.

Evaluation in education is not new.[1] What is new, however, is that financial and economic considerations dominate the evaluation scene today. "Cost-effectiveness" and "value for money" are commonly used terms. The rhetoric of "obtaining education for its own sake", "achieving excellence" or "improving quality" is followed up by questions such as "at what cost?" and "who is going to pay?".

The financial underpinning of the evaluation debate today explains why economists are now so heavily involved in an area that was once the domain of educators, psychologists and sociologists. This paper is no exception to that trend. But that is not to say that other disciplines should be excluded from the process of educational evaluation. As I will go on to argue, an interdisciplinary approach is essential for evaluation to be successful. At the same time, and with all due respect to my non-economist colleagues, I would not be content to leave the conduct of an evaluation entirely in their hands because, as the current financial situation has proven, their evaluation might overlook the economic constraints.[2]

How can comparative education assist today in improving educational systems? What is the role of "evaluation" in this respect, and what should be evaluated in the first place? I assume that the most fundamental interest of this group is: (a) to understand school-related phenomena, mostly in a cross-national context, and (b) to draw conclusions that may lead to educational improvement.

---

[1] For example, see Walberg and Haertel (1990) for an extensive review of the literature.

[2] For example, without the restraint of an economist constantly asking, "at what cost," or "what alternatives are being sacrificed", an educator may arrive at non-replicable, Rolls Royce solutions to education problems.

The *raison d'être* of the field of comparative education is the belief that by studying education internationally one's understanding of one's own educational system is enhanced and thus one will be better equipped to improve it (Sadler, 1900). Hence, a very popular question is: What is the best educational system in the world? If there was agreement on the answer to this, then the simplest solution ought to be for every other country to copy it.[3/]

Someone once said that the ideal world is one in which the cooks are French, the policemen are British, the mechanics are German, the organizers are Swiss and the lovers are Greek.[4/] With the free movement of labor within the European Community, such a paradise may well materialize. But where are the best educators?

A Newsweek article in 1991 attempted an answer to this question: Send your young children to kindergarten in Reggio Emilia, Italy. Then, somehow split them between New Zealand and the Netherlands to enable them to acquire a solid base in reading and mathematics respectively. Once literate, let them learn their science in Japan and their foreign languages back in the Netherlands. If you can split them further, then consider relocating them for high school in Germany or at least expose them to German-trained teachers. The foundations of their arts education should be obtained in the United States, and so should their graduate education. When they become adults, any further education should be obtained in Sweden.

Comparative educators sometimes behave like economists who, after all, might not be alone in practicing a dismal science. Comparative educators may also look "where the light is", rather than where the keys were lost. If shipwrecked, they may also "assume they had a can opener." And, of course, they often do all this "at the government's expense."

There is no question that "evaluation", "standards", "quality" and "achievement" are current buzzwords. I will pass over the intricacies of their definitions and subtleties [5/] so that we can focus on the broader picture.

In the following sections, I discuss a number of conceptual issues such as what exactly should be evaluated and by whom. Section III deals with the methodology of evaluation and in particular the role of international comparisons in this respect. Section IV summarizes where we stand and presents an agenda for research on questions that are still wide open.

[3/] All too often, when visiting a country, I am asked: "Why not copy the educational system of another country where things work better?" ... as if it were that easy!

[4/] Alternative versions claim Italians or Spaniards for the latter role.

[5/] For that see OECD (1989), Psacharopoulos (1991).

I am being deliberately cavalier with the mainstream educational evaluation literature. Or rather, I take this literature as given and try to answer the question: Given the fact that resources are limited, not only for education but also for evaluation research, what types of evaluation should have priority? How can scholars, whether "comparative" or not, contribute to improving a country's educational system?

II. Conceptual Issues

There are several considerations at the outset: (a) what should be the criteria of evaluation, (b) what exactly should be evaluated and (c) who should conduct the evaluation. Another major issue, the evaluation methodology, will be treated in a separate section below.

Evaluation criteria

In an educator's domain, a common criterion of evaluation is the learning outcome of alternative curricula. Although important, to the economist this is an extremely narrow evaluation criterion. Starting from the axiom that education and training aim to improve human wellbeing, [9] an economist would take a wider approach to the evaluation question and try to express social wellbeing as a function of education and training activities:

$$WELLBEING = f(EDUCATION, TRAINING).$$

At this abstract level, wellbeing may be split into efficiency and equity components, for example, how education and training spending translates into more food and other products or services available to the consumer at large (what economists call "economic or per capita income growth"), and how such spending contributes to distributional equity in a given society:

$$WELLBEING = g(EFFICIENCY, EQUITY).$$

The task of the evaluator is then to model the above relationships, to observe and measure proxies for the grand arguments in the social wellbeing function, to assign normative

[9] Given the multi-disciplinarity of my audience I explicitly avoid the more correct "social welfare" economic jargon attached to this criterion in order to avoid confusion with the common usage of the latter term.

weights [7] to the efficiency and equity components and to provide answers to the question of what level and types of education and training [8] contribute most to human wellbeing.

This modeling, for example, could be done by using national income (GNP) as an (admittedly partial) measure of efficiency, a standard income inequality measure (such as the Gini coefficient) and arbitrary weights alpha (α) and beta (β) for the efficiency and equity components respectively:

$$WELLBEING = GNP^{\alpha} (1 - GINI)^{\beta}.$$

To put it into words, a certain allocation of funds to curriculum X rather than to curriculum Y or to higher education rather than to primary education will have a differential impact on economic growth and income distribution in a given society. Such effects can and have been measured. [9]

What to evaluate?

The range of educational outcomes subject to evaluation is enormous. [10] Given the recent economic rationale for educational evaluation, this range has been increased to include post-school performance, such as occupational attainment and earnings, taking into account the cost (to the individual and society) of achieving these objectives. Since one cannot evaluate everything, where do we draw the line?

In order to answer this question, let us split the temporal sequence of events over someone's lifetime into three major stages, before school, in school and after school, bearing in mind the efficiency and equity components. This gives us six major areas of evaluation, as shown in Display 1.

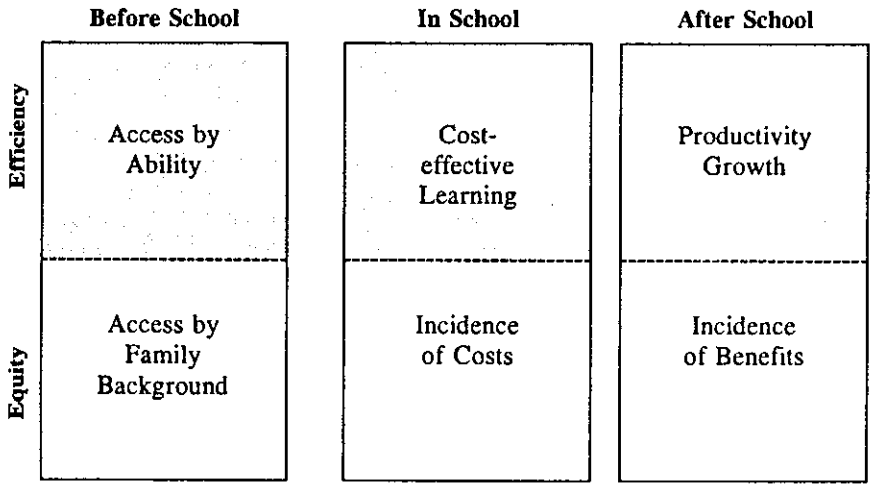
---

[7] Such weights can only be determined by politicians. The role of the evaluator in this respect should be limited to recognizing the tacit existence of value judgements and the estimation of alternative wellbeing outcomes given different hypothetical values of normative weights.

[8] In what follows, I will refer to "education" in a shorthand fashion to include "training". After all, the difference between education and training is one of modality in the transmission of knowledge. When the modality is important to the context, I will discuss training in particular.

[9] For example, see Psacharopoulos (1977).

[10] For example, see Bloom (1956) for an early and, by today's standards, a very partial list.



Display 1. Taxonomy of Key Evaluation Areas

A. *Before school*

- *Access by ability.* Assuming that the more able[11] will benefit most from more schooling (especially higher education)[12], one question related to the efficiency of an educational system is whether, for entry into a particular level of schooling, it really selects those students who are best capable of learning from that level of schooling or type of training rather than throwing it open to all.

- *Access by family background.* From the equity viewpoint, the evaluation question is: to what extent are those coming from a poor socioeconomic background excluded from certain levels or types of education?

---

[11] "Able" here does not have any genetic connotation. It means likelihood to succeed in further schooling, for example, as evidenced by a high Scholastic Aptitude Test (SAT) score before entering university.

[12] In this case, "before school" means access to any level of education and not just primary education.

B. *In school*

- *Cost-effective learning*. At each particular schooling level, the key efficiency evaluation questions are: Does learning in fact take place [13] and is it achieved in a cost-effective way?

- *Incidence of costs*. From the equity viewpoint, the evaluation question is: What is the relationship between the distribution of financial resources for education and the economic situation of the student and his family? In other words, are poor (and able) students excluded from higher education because their families cannot afford the fees?

C. *After school*

- *Productivity growth*. After leaving school, the cardinal evaluation question from the efficiency viewpoint is whether the funds devoted to education have an economic payoff in the form of the increased productivity of the graduate in the labor market.[14]

- *Incidence of benefits*. The equity evaluation question at this stage is: How does the provision of education (in quantity and quality) relate to the distribution of rewards among the population.

Who should evaluate?

In order to answer this question, let us identify three key actors in the educational process that is the subject of the evaluation: (a) the producer of educational services, (b) the financier of such services and (c) the consumer(the student and his/her family.) Also, we must sharply distinguish between private and public school systems. In some instances, the producer of the services may be the same as the financier (as, in a public school system). In other cases, the producer is a distinct entity (as in a private school system). (See Display 2).

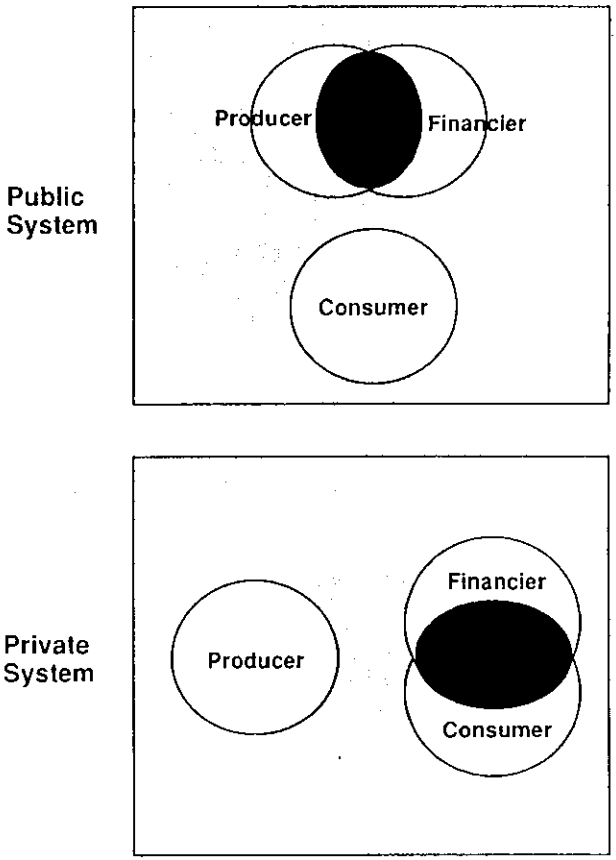
---

[13] "Learning" here is used in a shorthand manner. In reality, the evaluator should assess *incremental* learning, by trying to control for any stock of learning the student had prior to entering the course under evaluation. (On the importance of this value added concept of learning see Hanushek, 1979).

[14] Of course "productivity" does not refer only to monetary rewards. It could also mean increased "home production" in the form of a more educated housewife (outside the formal labor market) contributing to better sanitation conditions in the household. (See Schultz, 1974).

Display 2

Actors/Evaluators in Public and Private School Systems



I argue that the need for evaluation is much greater in a public rather than private school system. In a private school, the large overlap between financier and consumer (lower panel of Display 2) ensures that the consumer will implicitly evaluate the provider of educational services before committing his/her financial resources. This evaluation might not take a statistical form in the sense of comparing achievement increments per dollar spent in a particular private school, but schools soon acquire a certain reputation for being good or bad. The good schools draw resources (in the form of student fees) and prosper, whereas bad private schools close down because of lack of financing.

Actually, there is nothing wrong with a bad private school surviving, perhaps charging lower tuition than a top private school, as long as (a) there are students willing to pay the fees charged by that school, in the full knowledge that they are receiving education of inferior quality or value, [15/] and (b) there exists an independent professional body that licenses people entering critical occupations such as physicians or airline pilots so that unqualified or underqualified people can be excluded.

To put it blandly, in a private school system, there is no need for evaluation as a way of ensuring standards. The consumer largely assumes the evaluator's role by means of controlling the resources that flow into schools of different quality. Private school headmasters are accountable to the student and their families, the ultimate evaluators of the system.

By contrast, in a public school system that combines the financing and production of services within the same entity (the public sector), the lines of accountability are less clear. True, taxpayers lobby for and against the resources allocated to a given school district or a particular school, but public schools do not depend as directly on consumer satisfaction for their financing and existence as do private schools. Accountability and incentives for better performance are very diluted in a public school system. Hence the greater need for evaluation in a public school system. But in this case, who should be the evaluator?

Traditionally, it has been the producer itself that evaluates a public school system. As I have already mentioned, the consumer is much more remote from the production line, and the financier is the same entity as the producer. What tends to happen as a result is that a unit located in the Ministry of Education conducts traditional analyses of how students learn in school, the effectiveness of different curricula and, more rarely, the cost-effectiveness of alternative educational inputs.

However, there exist more effective ways to evaluate public schools. One possibility would be to separate the financing from the provision of education. This could be done by

---

[15/] Of course, such a principle will not apply in thinly-populated rural areas where only one school is available.

assigning the evaluation of schools to, say, the Ministry of Finance (or to an independent body reporting to the Ministry of Finance) which would reward the best schools by granting them additional resources. [16] Another possibility would be to tie the promotion and pay of particular teachers to student performance -- a policy that would be difficult, but not impossible, to implement.

Often the evaluation of schools (public or private) by the public sector is based on the argument that consumers are ignorant about what constitutes good and bad schools. But in today's world, especially in Europe, I strongly challenge the notion that parents cannot distinguish between a good school and a bad school.

III. Evaluation Methodology

Having settled what to evaluate and by whom, the question is "how"? Particularly relevant for this audience, how does the "international comparative" element come in?

Educational evaluation is a very tricky subject as it involves many disciplines. It can be descriptive/anecdotal in nature, or it can be more analytical/statistical. (Display 3).

The first type of "evaluation" has been the rule of thumb for many years and is still dominant today. [17] Newsweek's report on the ten best school systems in the world is an example of this type of evaluation. (And, in fairness to the magazine, this fascinating report was not meant to be a scientific evaluation.) The reason I do not believe this constitutes a genuine evaluation is that someone else, given enough local information, could produce evidence on other schools in other countries that are equally good, or even better, than the ones identified by the Newsweek reporters. [18]

---

[16] This policy is a two-edged sword, in the sense that it might have adverse equity effects. There are instances where low quality schools should receive more resources, especially if there are no competing schools in the area.

[17] The quotation marks denote that, in my opinion, type A evaluation leaves much to be desired.

[18] I bypass the collection of "isms" associated with this type of evaluation as the topic has been covered extensively in the August, 1990 issue of the *Comparative Education Review* (see Psacharopoulos et al., 1990).

Display 3. Types of Evaluation Methodology

Type A	Type B
• Descriptive/anecdotal	• Analytical/statistical
• Qualitative	• Quantitative
• Using words/narration	• Using numbers/data
• Opinion surveys	• Fact surveys
• Vague or no hypothesis formulation	• Specific hypothesis testing
• Concerned with "isms"	• Concerned with substance

The second type of evaluation is more demanding. It requires first that an analytical framework be established and that hypotheses be formulated regarding causal links, It next requires that these hypotheses be tested using actual data rather than opinion survey.[19/] The reason why evaluation A is more dominant today is that even though it is not persuasive it is easier to do. It is searching "where the light is", rather than where the keys were lost.

From now on, I will only refer to evaluation type B. A key element in this evaluation is the establishment of a *control group*, i.e., an entity or base measurement against which to judge the effect of an alterable policy variable. This is a notion that unfortunately is missing from the vocabulary of those engaged in descriptive evaluations. For example, suppose children in school district X may outperform children in district Y by a wide margin on the same standardized test. To what extent is the superior performance of children in district X due to the knowledge children already had when they enrolled in school (for example, acquired in the home from educated parents who acted as informal teachers) as opposed to the education they received in school?

A related notion is that of *controlling for other factors* that may produce similar effects. For example, to what extent is the earnings advantage of college graduates due to their education rather than to their inherent ability which allows them to progress to further education? The

[19/] The two types of evaluation are, of course, not mutually exclusive, in other words, quantitative analysis does not preclude qualitative considerations.

type B evaluator would want to control for ability in assessing the external effects of higher education. The descriptive evaluator might not be able to differentiate these two factors.

International comparisons

But how does the comparative element come in? To start with, all evaluation involves a comparison of some kind, whether the evaluator states this explicitly or not. The real obsession with international comparisons in education started with the launch of Sputnik and led to the demand for evaluation.

About ten years ago, a United States National Commission on Excellence in Education produced a frequently cited report, *A Nation at Risk*, describing what was wrong with the United States education system (US Department of Education, 1983). As evidence for the existence of "an education risk", the report listed three kinds of indicators that are very different in nature (pp. 8-9):

(a) *Cross-country comparisons*: The first indicator of risk was that "International comparisons of student achievement...reveal that on 19 academic tests American students....in comparison with other industrialized nations were last seven times." The norm in this case was supplied by other countries. Borrowing a standard from another country is comparative education's bread and butter.

(b) *Absolute standards*: The second and third risk indicators were that 23 million American adults were functionally illiterate, with the incidence of this illiteracy running as high as 40 percent among minority youth. The criterion in this case, literacy, was a widely accepted standard, regardless of place or time.

(c) *Over-time, within-country comparisons*: According to several other indicators in the report, achievement in science and mathematics had been declining over time in the United States. The standard in this case was provided by the score in the same test in the past, in the same country.

More recently, IEA and IAEP studies on achievement in different countries have fueled the so-called International Olympics in Education.[20] Results from studies like these receive a lot of attention in the daily press and typically show that the United States is not doing very well, especially relative to Asian countries.

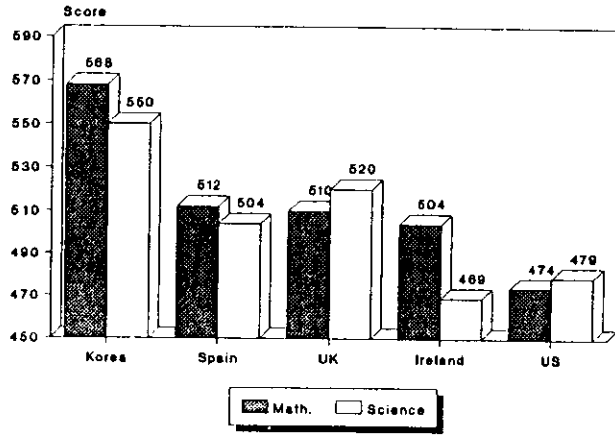
---

[20] See the Appendix for an inventory of who is doing what in international achievement testing and a glossary of the organizations' acronyms. Also see Torney-Purta (1990) for a bird's eye view of their activities. For a good description of the IEA studies see US Department of Education (1992).

For example, according to the second IEA International Mathematics Study, United States twelfth graders scored nearly at the bottom in algebra and geometry relative to their counterparts in 14 other parts of the world (IEA 1989, pp. 22, 24). In the more recent IEA Science II study, ninth graders in the United States scored in the core test at about the same level as their counterparts in Papua New Guinea, 54.8 and 54.5 respectively (Postlethwaite and Wiley, 1992, p. 60). According to the most recent IAEP science study among 13 year olds only 67 percent of the United States children answered correctly versus 78 percent of Korean children (IAEP, 1992a, centerfold). In a similar mathematics study, only 55 percent of the United States children gave correct answers, whereas 73 percent of Taiwanese children did so (IAEP, 1992b, centerfold). On another front, a report from the United States Congress on worker training concluded that: "When measured by international standards, most American workers are *not* well trained" (United States Congress 1990, p.3, emphasis in original).

But how valid are international comparisons for judging a country's educational system? Let us take as an example the IAEP science and mathematics achievement study conducted in 1988 among 13 year olds in Korea, Spain, the United States, Ireland and Canada (see Lapointe et al., 1989). Display 4, based on this study, shows a typical achievement bar graph that makes headlines in the United States.

Display 4. Proficiency in Science and Mathematics among 13 Year Olds, 1988



Source: Lapointe et al. (1989), Figure 1.1 and 4.1

As shown in the display, Korean students are clearly doing the best in both subjects and U.S. students are doing the worst. It is evidence such as this that has prompted a lot of discussion in the United States about improving the school system (see *A Nation at Risk*, op. cit.) But is it really the school system that should be blamed for the lower performance of American children relative to Korean youngsters? Could it be that, because of cultural factors, Korean parents pressure their children more (a) to work on school-related matters after school hours and (b) to watch less television? Or could it be that the school retention rate for the same age group differs from one country to another? [21]

Or consider the *prima facie* embarrassing performance of United States' high school students relative to their counterparts elsewhere in IEA's Science II study. The top panel of Display 5 shows the scores of "Population 2" in the study. Before sounding an alarm regarding the deleterious state of United States' high schools, the reader should consider three important facts, stated clearly in the study [22] but very unlikely to appear under the newspaper headlines reporting the respective bar graph: (a) In the United States attendance of the reference age group is 99 percent, whereas in Papua New Guinea only 11 percent of the relevant age group is in high school; (b) "Population 2" students in the United States are ninth graders, whereas in Papua New Guinea they are tenth graders; and (c) the mean age of the subject population in the United States is 15 as opposed to 17 in Papua New Guinea. Could such sample selectivity/non-comparability be responsible for the observed gross differences in the achievement bar graph? It is natural to expect that a more selective/older/higher grade sample will result in higher achievement, other things being equal.

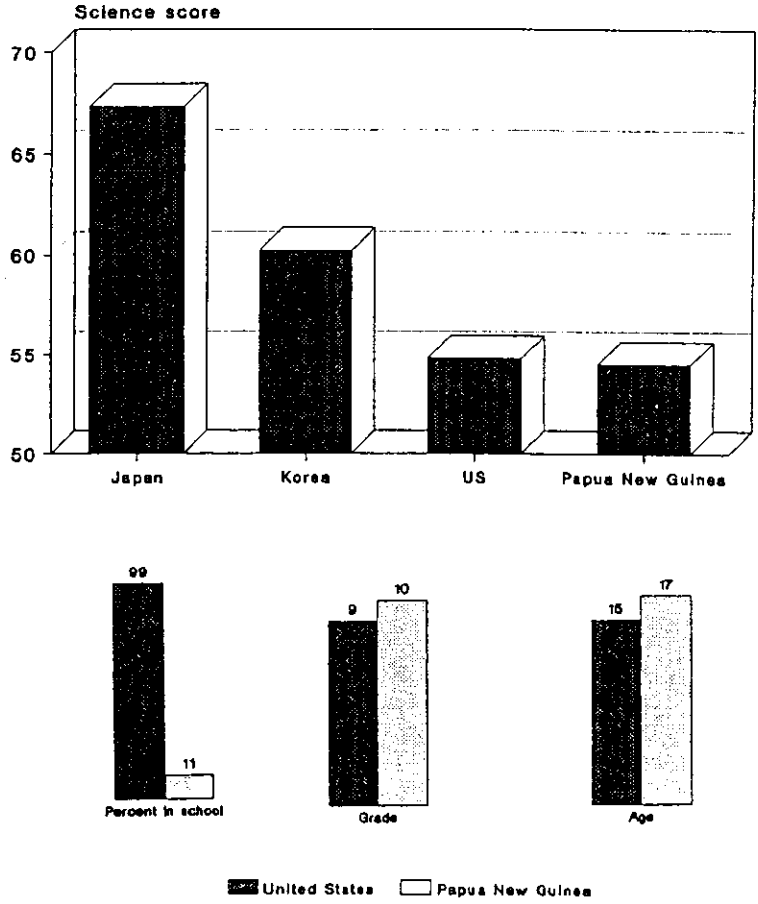
It was reports such as this that moved public opinion in the United States to see the necessity for school reform (see also: US Department of Education, 1992, Figure E.7). But the reports more often than not forget that different school systems have differential retention rates. For example, if the United States covers 90 percent of the age group in the last year of secondary education (as shown in US Department of Education, 1992, Figure C.4), maybe it should not be surprising that having large numbers of students at each age level means that their overall achievement scores will be lower than those of other countries where a small percentage (presumably the brightest) of the population is educated. Also, school systems in different countries differ in what they teach each age group (for example, calculus may be taught later in one country than in another) just as different cultures affect student's attitude towards study (Japanese students are likely to spend more hours on homework than their American counterparts).

---

[21] For the controversy surrounding the validity of international achievement comparisons see Husén (1983), Rotberg (1990) and Bradburn et al. (1991).

[22] See Postlethwaite and Wiley (1992), Table 1.1.

Display 5. Mean Achievement in Core Science Test and Sample Differences, Selected Countries



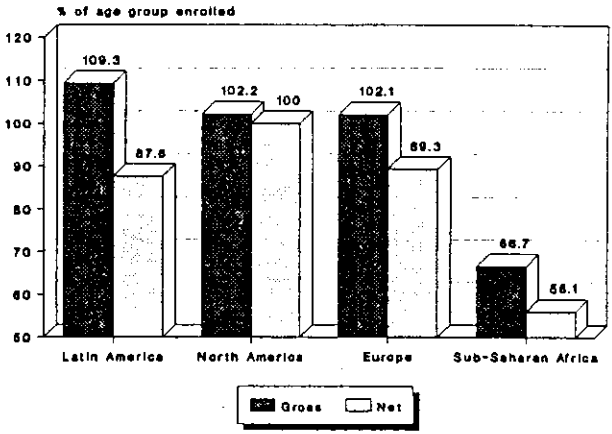
Source: Based on "Population 2" of IEA Second Science Study, Postlethwaite and Wiley (1992), pp. 5 and 60.

More pedestrian indicators

As international education indicators go, comparisons of achievement are among the most sophisticated and scarce. Due to the availability of data, it is more common to compare educational systems in terms of enrollments or educational finance than achievement. Let us take a brief trip into this territory and gauge the limits of this approach.

A good example of how international comparisons can be deceiving is to compare primary education enrollment ratios in selected regions of the world (see Display 6). One might conclude from the display that the primary education systems in sub-Saharan Africa are less developed relative to the rest of the world in the sense that only 67 percent of the respective age groups are covered. But by using this indicator, one would also be forced to conclude that the education systems in Latin America are more developed than those of the United States, Canada and Europe. Such a deduction is surely counter-intuitive.

Display 6. Enrollment Ratios in Primary Education by Region, 1990



Source: UNESCO (1991), Table 2.4
Note: Percent of the population aged 6-12 who are in school

The key to this puzzle lies in the use of *gross* enrollment ratios, which neglect the age dimension of those attending school, rather than *net* enrollment ratios.[23] Unfortunately, often only gross enrollment statistics are reported because they are easier to collect this information. Latin American countries top the world in terms of grade repetition (UNESCO, 1990) and consequently, show higher gross enrollment ratios than countries in other regions. Net enrollment in schools (i.e., students of school attendance age attending school as a fraction of the *same* age group in the population) continues to be a very scarce statistic. Hence, the search for "where the light is".

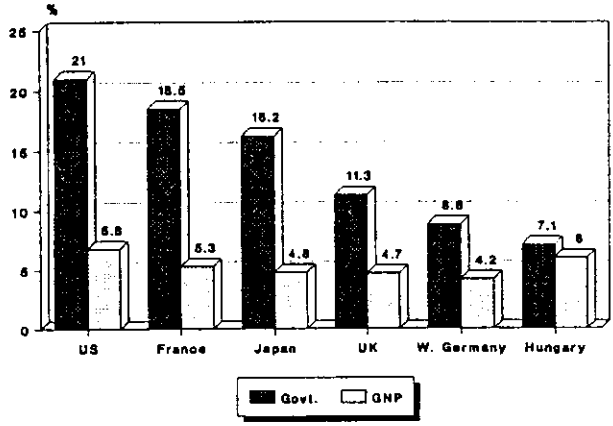
Or consider the amount of resources each country spends on education (Display 7). Here the United States is the leader among the countries displayed in terms of *public* resources devoted to education, both as a percentage of total government expenditure as well as a proportion of the country's national income. On the basis of this indicator, a casual evaluator would rate the United States above Japan and certainly above Germany. But then, how does this information square with the now universally accepted view (which is also supported by Display 5) that Japan's education system is better than that of the United States in terms of achievement performance? Are Japan and Germany more efficient than the United States in translating public expenditure on education into student learning? Or could it be that the most readily available information on *public* expenditure on education masks *private* out-of-pocket expenditures by households, such as spending on cram school/coaching by Japanese parents?

Within-country comparisons

Now let's contrast this "methodology" of using international comparisons to evaluate an education system (which might conclude that the United States is really doing poorly relative to Papua New Guinea) to other types of analysis which focus solely on factors within a given country. In such studies years, schools, states or students provide the source of "comparative" variation. The basis of analysis may be a combination of the following: (a) cross-sectional data (the information referring to one point in time, say 1992, and using states, districts, schools or individual students as the units of observation) or (b) time-series data (comparing the above generated indicators over time within the given state, district, school or grade).

[23] This is also the reason why gross enrollment ratios can exceed 100 percent, although net enrollment ratios cannot.

Display 7. Public Expenditure on Education as a Percentage of National Income and Total Government Expenditure



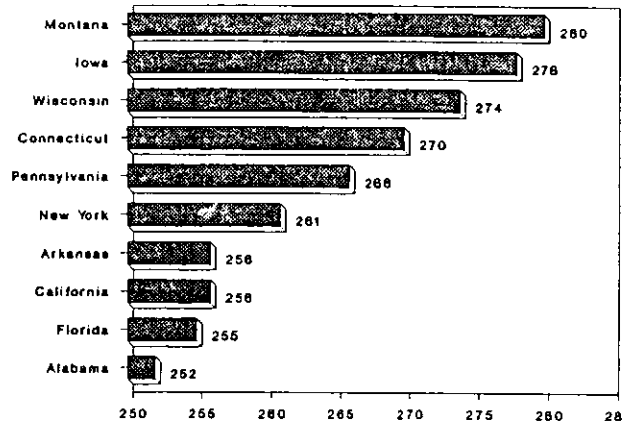
Source: UNESCO, Statistical Yearbook 1989 and 1991.

Cross-sectional comparisons

Display 8 shows a cross-sectional comparison for the United States, where the state of Alabama is clearly at the bottom of the achievement league. Once again, this type of information necessarily begs the question of whether it is the school system of Alabama that produces such appalling results, or it is the state's poor economy relative to the country's Northeast and Midwest? But at least this method ensures automatic control for the myriad of "cross-national" factors in which countries differ. [24]

[24] For an analysis of cross-state data on achievement, see United States Department of Education (1990a).

Display 8. Mathematics Proficiency at Grade 8, Selected States, U.S., 1990

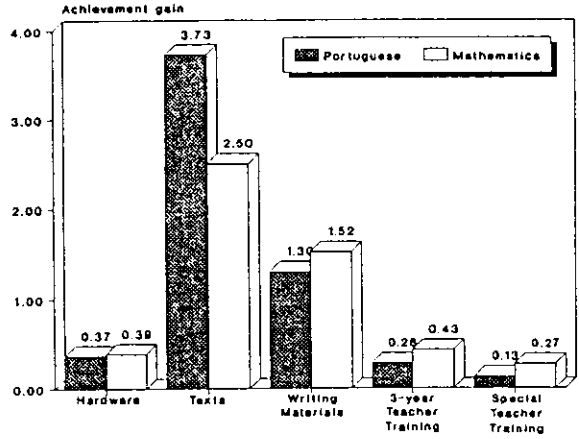


Source: US Department of Education (1990), Table 3.

Using the student as the unit of observation constitutes, in my opinion, a better focus for evaluating what works and what does not work in education. To recall what was mentioned earlier in this paper, the end result of evaluating a school system should not be to produce a hierarchical table of how a country, state, district or student is doing in terms of, for example, achievement *en absoluto*, but to identify the factors responsible for such things as achievement *differentials* among the different units of observation. As such, we consider the example portrayed in Display 9 coming from Brazil's Northeast -- a region with atrocious educational conditions. Harbison and Hanushek (1992) not only fitted educational production functions to determine what inputs affect educational achievement, in a value-added sense, but also analyzed the costs of the alternative interventions and related them to the student achievement gained from each input provided.

This evaluation revealed things that would not be obvious using any other method of analysis. For example, hardware inputs, such as classroom construction and chairs, are not as cost-effective in raising student achievement in Portuguese and mathematics among second graders as are "software" inputs, such as the usage of textbooks and the availability of writing materials.

Display 9. Achievement Gains of Educational Inputs Per US$, N.E. Brazil

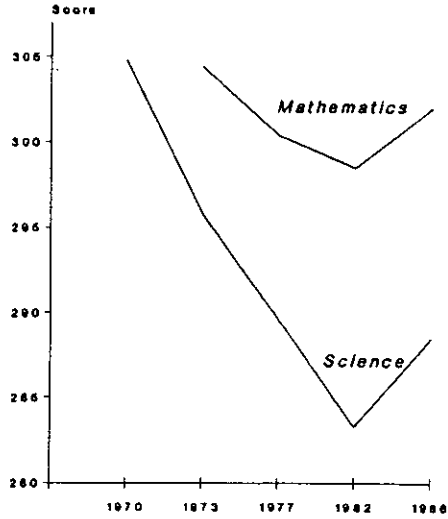


Source: Harbison and Hanushek (1992), p. 138

Time-series comparisons

A time-series comparison can yield additional information, as it at least achieves standardization for the many non-school factors in which countries differ (for cross-national comparisons). Furthermore, this methodology can control for overall parental wealth (for cross-state comparisons). But fine tuning for controlling individual student/family wealth would require longitudinal panel data -- a rare breed indeed. An excellent example of this type of comparison is presented in Display 10. The negative trends exposed by this comparison set off the alarm regarding the state of education in the United States. Regardless of how much better Japan and Korea did in international comparisons, the fact remained that, *within* the United States, achievement dropped sharply over a period of more than one decade.

Display 10. Proficiency in Science and Mathematics among 17 Year Olds in the United States

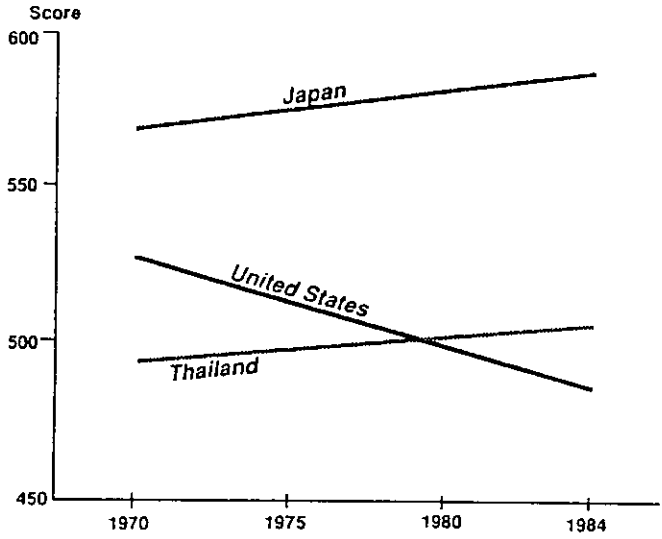


Source: U.S. Department of Education (1990b), Figure 2.1

A good example of within-country, time-series analysis is that contained in Bishop (1989). Using trends in test scores (which declined between 1967 and 1980), he attributed part of the productivity slowdown after 1973 in the United States to the preceding fall in test scores. If the academic achievement of high school graduates had been maintained at the rate of gain prevailing between 1948 and 1953, workers would now be about three percent more productive.[25]

[25] For a presentation and analysis of achievement trends in the United States, see Mullis, Owen and Phillips (1990).

Display 11. Science Score among 14 Year Olds in Selected Countries, circa 1970 and 1984



Source: Keeves (1992)

One can of course combine cross-country and over-time achievement data to arrive at information as was done by Keeves (1992, p. 15) and is illustrated in Display 11. The problem is not so much with the differential level of achievement between countries in any given year, as the drop of achievement in the United States between 1970 and 1984.

A similar example pointing to the superiority, in my opinion, of within-country, time series compared to traditional international comparisons is given in Display 12. Educational expenditure per pupil is a common "evaluation" indicator. For instance, looking across countries in 1986, we see that Japan spends a little less per student than the United States. We also see that the Netherlands spends significantly more per student than do either Japan or the United States. Are we to conclude from this information that the United States has better schools than Japan? Such a conclusion does not necessarily follow, and it certainly does not conform with achievement information so widely publicized in the press which suggests that Japan is doing better that the United States.

Display 12. Total Educational Expenditure per Pupil as a Percentage of GDP per Capita

Year	FRG	Greece	Ireland	Italy	Japan	Neths.	UK	US
1970	20.5	16.0	24.3		23.9	31.0	31.6	21.8
1971	21.8	14.3	24.4	23.6	24.5	31.1	31.5	22.0
1972	22.1	14.8	24.3	23.6	23.6	30.4	32.3	20.9
1973	21.3	14.4	24.6	23.2	23.0	29.3	30.2	20.4
1974	21.6	16.9	23.7	21.8	25.2	28.3	32.7	20.7
1975	22.6	16.6	24.3	21.6	24.6	29.1	31.8	21.8
1976	20.9	16.4	24.8	21.9	24.5	28.2	30.0	21.4
1977	20.3	18.2	24.3	21.3	24.5	27.8	28.3	20.6
1978	20.1	18.7	24.7	19.7	25.5	28.4	27.4	20.2
1979	20.1	17.5	25.1	20.7	25.1	28.5	27.4	20.3
1980	20.4	16.1	24.2	20.2	26.0	29.2	28.3	20.8
1981	20.7	16.8	24.8	21.1	24.1	30.4	28.2	20.5
1982	20.8	18.3	22.9	21.6	23.2	30.9	27.3	21.4
1983	20.4	18.0	23.4	22.2	22.9	30.6	27.3	21.3
1984	19.7	18.6	22.3	23.9	22.1	29.1	26.6	20.7
1985	20.0	19.5	22.1	23.9	21.4	29.5	21.9	21.2
1986	20.1	18.6	23.1	24.7	21.6	31.1	22.9	22.2
1987	20.2	18.6	22.8		21.1	32.8	22.6	22.3

Source: OECD (1992), Table 2.11

But looking vertically, in other words, over time within a given country, we gain different insights. We see, for example, that the United Kingdom has managed to decrease significantly the amount of public resources spent per student during the last decade. Is this due to the fact that during the period under consideration privatization in Britain reached its zenith (including education)? Has education quality suffered by such reductions? Or have the private resources flowing into the system (for example, in the form of increased foreign student fees in higher education) helped improve the quality of instruction? Although such questions require more detailed analysis, beyond what Display 12 can provide, the vertical reading of the table is, in my opinion, more informative than the horizontal "traditional comparative education" reading.

IV. Where Do We Stand?

All countries in the world face, and will continue to face, a broad range of educational problems that "evaluation" can help to solve. One might think that developing countries face different educational problems than those faced by advanced OECD countries. But upon closer examination, it becomes evident that the problems are similar in nature, regardless of location.[26/] Just as there are headaches and pneumonia in Abidjan and Paris, so there are low performers and school dropouts in both places-- hence a need to know which interventions will best cure these ills. Is it better to prescribe an aspirin or penicillin? Just as there is unsatisfied demand for access to higher education in Tanzania, so there is in Greece. Would the creation of "centers of excellence" (whatever this means) help solve these problems? Or perhaps some selective cost recovery, coupled with student loans, would be a move in the right direction?

Certainly, the debate may have different names in different parts of the world. In Europe, it might be hidden under the rubric of "harmonization of educational systems", "student mobility", "cross-country recognition of qualifications" or "linking universities with industry". In the United States, it might be called "declining standards". In poorer countries, the problem might manifest itself as an "educational financing crisis" or "non-relevance of the curriculum." However, the basic issues remain the same, and proper evaluation *can* help to find solutions.

From my perspective, it is encouraging to see that what I have described above as type A evaluation is in decline, while type B evaluation is in ascendance. The efforts by the IEA, OECD and BICSE are all moves in the right direction. All of these efforts involve measurement, control groups and standardization for other factors affecting performance. But within every effort there is still room for improvement.

It often takes five years from the inception of an IEA study for it to be carried out and the results to be published. Educational practitioners may not be willing to wait for international studies to suggest solutions before they take action in their respective countries. In addition, in spite of its importance, the IEA never included in its portfolio an external evaluation of the school system (for example, by tracing graduates out of school and into the labor market). To the extent that the transition from school to work continues to be one of the foremost educational problems, such study would be extremely welcome.

It is very encouraging to see that the OECD-INES project is extending the range of indicators to those relating to the labor market. Information collected under this project will include the educational attainment of the population and the labor force, occupational distribution by educational attainment, the transition characteristics and labor force status of school leavers, after school training of the young and adults and, most importantly, relative earnings from work

---

[26/] An important qualification here is that the institutional capacity to conduct educational analysis and implement potential reforms varies from country to country.

by educational attainment (see OECD-INES, 1991). Let us hope that the empirical results from this project will soon become available to the educational research community and to the practitioners in particular.

International statistics in education started with enrollment data, then began to include public financing data (see *Unesco Statistical Yearbook*, various years). Today, a proper evaluation of a school system needs more information than that. Here is my short list of additional indicators that I hope one day will be readily available for all countries in the world:

- enrollments in private schools
- private contributions to the financing of education
- cost per student by level and curriculum type
- availability of learning materials
- incidence of who pays and who benefits from public educational expenditure
- within country achievement scores, by grade and subject (even without attempting comparability between countries but building time-trends of achievement within countries)
- tracer information on how the output of the school system fits into the world of work (earnings of graduates by educational level, incidence of unemployment, length of time of recent graduates to land a job).

Collecting such data, of course, goes beyond the immediate capacity of a typical Ministry of Education. The collection process would involve within-country surveys of schools and graduates, and that would require resources and commitment. Special units would need to be set up, preferably *not* within the Ministry of Education so as to achieve independence from the state bureaucracy. Funding continuity would be required, as would a critical mass of analysts. These latter would consist of multi-disciplinary teams which would include economists and statisticians in addition to educators. A feedback mechanism would need to be devised to ensure that the data analyzed made its way into the policy formulation process. This is the only way to guarantee rigorous evaluation of an education system.

I welcome international comparisons. In spite of the contrast between Papua New Guinea and the United States given earlier, students in the United States score consistently lower than those in other countries with similar net enrollments, such as Korea and Japan. When you see smoke, somewhere there must be a fire. I see the value of international comparisons mainly as a gross diagnostic tool for pointing out a problem. But in order to design policies to correct the problem, we must take a closer *within-country* look.

Another good feature of international comparisons is that they have an important training function. They not only keep local education researchers informed of alternative techniques used elsewhere, but help keep them on their toes regarding sample selectivity, questionnaire design and the analysis of cause-effect relationships. The little "Green Booklet" of BICSE (1990) provides a wealth of important advice on how to use and improve upon such comparisons, as do reports emanating from the OECD's INES project (see Binkley, Guthrie and Wyatt, 1991).

How can the comparative educator help today in evaluating educational systems? To play with semantics again, this really means the *international* comparative educator, as every rigorous analysis of an educational problem requires comparison of some kind. As an example, let us take an issue that has surfaced very prominently in recent years, namely the "financial crisis in education". There is widespread agreement today that the private production of goods and services is the most efficient system. Many countries that have historically relied on the public sector to plan or supply what people want have recently made 180 degrees turns towards private systems. Does education belong in the same category? The arguments are split.

There are still strong merit good, externality, information, and market failure arguments for the state to be involved in education. But once these are catalogued, there remain significant parts of the educational system which can be carried out better by private means. For example, we seem to agree that although the state should finance some levels and types of education (for example, by giving a voucher or scholarship to the poor), the state does not also have to be the producer of the educational services itself: the students who receive state financial support could use this support to study in a private school.[27] Comparative educators can help in this respect to document instances of how differential modes and mixes of public and private provision contribute to the criteria of efficiency and equity listed earlier.

There has been a large amount of research on all educational problems touched upon in the preceding pages, although some have received more attention than others. For example, sociologists have extensively documented access to education by family background, and economists have done much work on the relationship between education and productivity and economic growth.[28] But there remains a host of issues on which more findings, especially those which are country-specific, would be of great use to educational practitioners. Here is a list of those most pertinent to contemporary Europe: (a) within-country evaluation of achievement with wide dissemination of the results so that exchange students would know to go to country/school X, rather than country/school Y; (b) evaluation of employer-training programs,

---

[27] See Williams (1991) for a review of the inevitable introduction of market elements in higher education in OECD countries.

[28] For a review, see Psacharopoulos (1991).

so that other firms know the best practices elsewhere; (c) evaluation of professional training programs so that employers know which school produces employable graduates (tracer studies/external evaluation); and (d) analyses of who pays for and who benefits from education and training programs so that the state knows which students require financial assistance and which can afford to pay for themselves.

Of course, these evaluations should be rigorous, statistical and analytical, and not merely descriptive. The field could use more quantitative evaluators conversant in analysis of variance, sample selectivity and t-ratios.

A long way from Sadler

Comparative education has changed a lot since Sadler's times. The questions then might have been at what age one should teach Greek and Latin, or how English schools could learn from the teaching of nature in Philadelphia schools? Today's questions are:

- What are the welfare effects of different educational policies?
- What are the effects of school fees and privatization on equity?
- How can incentives be provided to teachers by means of a non-civil service pay scale?
- What is the optimal mix between an academic and a practical curriculum?
- How can resources be raised for education?
- What are the determinants of educational outputs?
- What is the cost-effectiveness of alternative educational inputs?
- How do graduates enter the labor market?

At the same time, over-enthusiastic, die-hard, international comparativists should not forget Sadler's wise words nearly one century ago:

> *"The practical value of studying...the working of foreign systems of education is that it will result in our being better fitted to study and to understand our own.... In studying foreign systems of education we should not forget that the things outside the schools matter even more than the things inside the schools....We cannot wander at pleasure among the educational systems of the world, like a child strolling through a garden, and pick off a flower from one bush and some leaves from another, and then expect that if we stick what we have gathered into the soil at home, we shall have a living plant."* (Sadler, 1900, p. 310)

REFERENCES

BICSE, Board of International Comparative Studies in Education, *A Framework and Principles for International Comparative Studies in Education*, Washington, DC: National Academy Press for National Research Council, 1990.

Binkley, M.R., Guthrie, J.W. and Wyatt, T.J. *OECD International Indicators Project: Network A: Student Achievement Outcomes*, OECD, 1991.

Bishop J. H., "Is the Test Score Decline Responsible for the Productivity Growth Decline?", *American Economic Review*, 79 (1) March 1989: 178-197.

Bloom, B.S., (ed.), *Taxonomy of Educational Objectives: The Classification of Educational Goals*, New York: McKay, 1956.

Bradburn, N., Haertel, E., Schwille, J. and Torney-Purta, J., "A Rejoinder to 'I Never Promised you First Place'", *Phi Delta Kappan*, June 1991: 774-777.

Hanushek, E., "Conceptual and Empirical Issues in the Estimation of Educational Production Functions", *Journal of Human Resources*, 14 (3) Summer 1979: 351-388.

Harbison, R., and Hanushek, E., *Educational Performance of the Poor: Lessons from Rural Northeast Brazil*, Oxford University Press, 1992.

Heyneman, S. and Loxley, W., "Effect of Primary School Quality on Academic Achievement," *American Journal of Sociology* 89(6):1162-1194.

Husén, T., "Are Standards in U.S. Schools Really Lagging Behind those in Other Countries?", *Phi Delta Kappan*, March 1983: 455-461.

IAEP (International Assessment of Educational Progress), *Learning Science*, Princeton, Educational Testing Service, 1992a.

IAEP (International Assessment of Educational Progress), *Learning Mathematics*. Princeton, Educational Testing Service, 1992b).

IEA (International Association for the Evaluation of Education Achievement), *The Underachieving Curriculum: Assessing U.S. Mathematics from an International Perspective*, Champaign, Illinois: Stipes Publishing Co., 1989.

IEA (International Association for the Evaluation of Education Achievement), R. Elaine Degenhart (ed.), *Thirty Years of International Research: An Annotated Bibliography of IEA Publications (1960-1990)*, The Hague, The Netherlands, IEA Headquarters SVO, 1990.

IEA (International Association for the Evaluation of Education Achievement), *IEA Guidebook*, The Hague, The Netherlands, IEA Headquarters SVO, 1991.

Keeves, J.P., *Learning Science in a Changing World*, The Hague: IEA, 1992.

Lapointe, Archie E., Mead, Nancy K. & Askew, Janice M., *Learning Mathematics*, ETS, Princeton, N.J., 1992.

Lapointe, A.E., Mead, N.A., and Phillips, G.W., *A World of Differences: An International Assessment of Mathematics and Science*, Princeton: Educational Testing Service, 1989.

Mullis, I.V.S., Owen, E.H. and Phillips, G.W., *Accelerating Academic Achievement: A Summary of Findings from 20 Years of NAEP*, U.S. Department of Education, 1990.

Newsweek, "The Best Schools in the World", *Newsweek*, December 2, 1991.

OECD, *Schools and Quality: An International Report*, Paris: OECD, 1989.

OECD, *Educational Expenditure, Costs and Financing: An Analysis of Trends: 1970-1988*. Paris: OECD, 1992 (forthcoming).

OECD-INES (1991), "Phase 2, Network B Report", September, 1991.

Postlethwaite, T.N. and Wiley, D.E., *The IEA Study of Science II: Science Achievement in Twenty-three Counties*, Pergamon Press, 1992.

Psacharopoulos, G., Adams, D., Benson, J.K., King, E. and Paulston, R.C., "Colloquy on Comparative Theory", *Comparative Education Review*, 34(3), August 1990: 169-404.

Psacharopoulos, G., "Measuring the Welfare Effects of Educational Policies," in *Public Economics and Human Resources*, Proceedings of the 31st Congress of the International Institute of Public Finance, eds. A.J. Culyer and V. Halberstadt. Cujas, 1977: 75-94.

Psacharopoulos, G., "Comparative Education: From Theory to Practice: or, Are you A:\neo.* or B:*.ist?", *Comparative Education Review*, Vol. 34, (3), August 1990: 369-380.

Psacharopoulos, G., *The Economic Impact of Education: Lessons for Policy Makers*. San Francisco: International Center for Economic Growth, 1991.

Psacharopoulos, G. and Arriagada, A. M., "The Educational Composition of the Labor Force: An International Update", PHREE Background Paper No. 92/49. The World Bank, January 1992.

Rotberg, I., "I Never Promised You First Place", *Phi Delta Kappan*, December 1990: 296-303.

Sadler, M., "How Far Can We Learn Anything of Practical Value from the Study of Foreign Systems of Education?", Notes of an address given at the Guildford Educational Conference on Saturday, October 20, 1900, as reported in George Bereday, "Sir Michael Sadler's 'Study of Foreign Systems of Education'," *Comparative Education Review*, February 1964.

Schultz, T.W., *The Economics of the Family*, National Bureau of Economic Research, 1974.

Stevenson, Harold W. & Stigler, James W., *The Learning Gap*, New York, Summit Books, 1992.

Torney-Purta, J., "International Comparative Research in Education: Its Role in Educational Improvement in the U.S.", *Educational Researcher*, Vol. 19, N. 7, October 1990: 32-35.

UNESCO, *The State of Education in Latin America and the Caribbean, 1980-1987*. Santiago, OREALC, 1990.

UNESCO, *World Education Report 1991*. Paris: Unesco, 1991.

United States Congress, Office of Technology Assessment, *Worker Training: Competing in the New International Economy*, OTA-ITE-457. Washington, DC: US Government Printing Office, September 1990.

United States Department of Education, *The State of Mathematics Achievement*, National Center for Educational Statistics, 1990a.

United States Department of Education, *Accelerating Academic Achievement*, National Center for Educational Statistics, 1990b.

United States Department of Education, *International Mathematics and Science Assessments: What Have we Learned?*, National Center for Education Statistics, 1992.

United States, National Commission on Excellence in Education, *A Nation at Risk: The Imperative for Educational Reform*, US Government Printing Office, 1983.

United States Department of Education, *The State of Mathematics Achievement*, National Center for Education Statistics, ETS, Princeton, NJ, 1991.

Walberg, H.J. and Haertel, G.D., *The International Encyclopedia of Educational Evaluation*, Pergamon Press, 1990.

Williams, G., "Markets and Higher Education", *Higher Education Management*, 3 (3), November 1991: 214-225.

APPENDIX: International Achievement Testing Efforts

Name: IEA, International Association for the Evaluation of Academic Achievement.

Remarks: A consortium of research institutions in over 40 countries specializing in cross-national surveys of educational achievement and related factors. Extensive country coverage since 1959 in mathematics, science, languages.

Reference: *IEA Guidebook*, The Hague, The Netherlands, IEA Headquarters SVO, 1991.

R. Elaine Degenhart (ed.), *Thirty Years of International Research: An Annotated Bibliography of IEA Publications (1960-1990)*, The Hague, The Netherlands, IEA Headquarters SVO, 1990.

Name: NAEP, National Assessment of Educational Progress.

Remarks: A nationwide test in the United States funded by the Department of Education and administered by the National Center of Education Statistics (NCES) under the direction of NAGB (National Assessment Governing Board). It has measured student achievement in basic subjects since 1969.

Reference: *The State of Mathematics Achievement*, ETS, Princeton, NJ, 1991.

Name: IAEP, International Assessment for Educational Progress.

Remarks: Unit of NAEP situated at Educational Testing Service (ETS), Princeton. Its purpose is to link US national assessment to other countries. Started in the mid-1980s.

Reference: Archie, E. Lapointe, *A World of Differences*, Educational Testing Service, Princeton, New Jersey, 1988.

Archie, E. Lapointe, Nancy K. Mead, & Janice M. Askew, *Learning Mathematics*, ETS, Princeton, N.J., 1992.

Name: UM, University of Michigan

Remarks: Case studies in Minneapolis, US, Taipei, Taiwan and Sendai, Japan examines primary school processes that affect achievement performance.

Reference: Harold W. Stevenson & James W. Stigler, *The Learning Gap*, New York, Summit Books, 1992.

Name: BICSE, National Research Council, Board of International Comparative Studies in Education.

Remarks: Its purpose is to advise United States federal agencies and to consult with others on needs standards, priorities and the value of international educational research. Organized by the National Research Council, which is the operating arm of the National Academy of Sciences and the National Academy of Engineering.

Reference: BICSE, *A Framework and Principles for International Comparative Studies in Education*, Washington, D.C., 1990.

Name: OECD/CERI, Organization for Economic Cooperation and Development/Centre for Educational Research and Innovation.

Remarks: Since 1987, this Centre has encouraged the production of national educational indicators from various sources for its 24 member countries including achievement indices. The Centre does not generate its own achievement data.

Reference: M.R. Binkley, J.W. Guthrie and T.J. Wyatt, *OECD International Indicators Project: Network A: Student Achievement Outcomes*, OECD, 1991.

Name: EEC, European Economic Community.

Remarks: The Task Force on Education and Human Resources in Brussels is organizing limited achievement data collection in the area of the proficiency in a second/foreign language of 15 year olds in the 12 Community member states.

Reference: Not available.

Name: UNESCO, United Nations Educational, Scientific and Cultural Organization.

Remarks: Compiles aggregate statistics on education and is considering the use of achievement data available for limited countries.

Reference: Not available.

Name: ECIEL, Programa de Estudos Conjuntos de Integración Económica de América Latina.

Remarks: Organized in Rio de Janeiro during the early 1960s, this program carried out a research survey in six South American countries using IEA reading and science tests and their own questionnaire instruments. Now closed.

Reference: S. Heyneman and W. Loxley, "Effect of Primary School Quality on Academic Achievement," *American Journal of Sociology* 89(6):1162-1194.
