These views are those of the author and do not necessarily reflect those of the IBRD.

INTERNATIONAL BANK FOR RECONSTRUCTION AND DEVELOPMENT

Staff Working Paper No. 200

March 1975

HOW EFFECTIVE IS SCHOOLING IN PROMOTING LEARNING?

A REVIEW OF THE RESEARCH

John Simmons

The purpose of this paper is to review the results of educational research in both developed and developing countries on the issue of schooling effectiveness. The policy implications of the findings are discussed.

Population & Human Resources Division
Development Economics Department
Development Policy Staff.

TABLE OF CONTENTS

DIAGRAMS

TABLES

I. SUMMARY

To what extent are schools effective in promoting learning in the developing countries? Ten years ago, this question would not have occurred to most educational observers. They were concerned mainly with increasing the supply of school places or improving the quality of educational inputs. Now this question is the one most frequently asked by an increasing number of educators, parents, and planners.

The major reason that doubt has arisen is the result of recent research on education in both developed and developing countries. On the other hand, the research has shown that schooling is not as effective in promoting learning as measured by achievement tests as educators had assumed. Moreover, the significance of traditional factors, such as class size, teacher training, and unit costs per student, had been overestimated. In many countries, no relationship could be established between these school variables and how much students tended to learn as measured on achievement tests. On the other hand, other factors, including parental behavior, nutrition, students' personality, and chance, were more significant than previously assumed.

Some controversy surrounds the recent research, partly because it contradicts what some observers would like to believe is happening in the schools, and partly because others have not read and understood the studies they are talking about. All observers have their biases, expecially those whose lives have been committed to schooling.

While the author has read the major studies of the literature and has contributed his own (with inconsistent results), the judgement formed from his experience is only one factor in shaping the findings reported here. This study, to mimimize observers' biases, draws heavily on reviews made by other authors whose politics and disciplines vary. To avoid distortion of their position, we cite extracts from their reviews, and studies, especially the recent international efforts. In fact, we were surprised to find such a high degree of agreement among the varied sources.

The paper is organized into five major sections:

(1) An introduction suggesting the relationships between educational research and policy, as well as the scope and limitations of the review;

(2) A review of the methods and findings from five different approaches to the same question of school achievement;

(3) International comparisons of cognitive achievement, based on studies that have just been published;

(4) Two central policy questions for planners concerned with poverty and social inequality, namely:

(a) how to reduce cognitive inequality among individuals, and;

(b) how to reduce the inequality in the years of school that individuals attain;

(5) A final section dealing with the limitations of the available research and how it can be interpreted for policy decisions.

Introduction

In most developing countries, the educational research on the effectiveness of schools in promoting learning is only beginning. Furthermore, quantitative research on other education questions has received even less attention. Of the developed countries, the United States has the best data, but good data are also available for several European countries, especially Great Britain. A recent study that compares European, American, and developing countries is discussed in detail in Section V. The better European studies are consistent with the results obtained in the United States. Because our depth of understanding of the learning process is much clearer with the American data, we concentrate our review on the American data.

Why should we first focus our review on American data, when our concern is developing countries? How legitimate is drawing cross-national conclusions? Our objective in educational research, as it is with research in agriculture and other sectors, is for the developing countries to draw the optimal benefits from research that is already done, thereby enabling them to reduce their future research expenditures. The basic assumption is that there is more variation within a country than across countries in how and why individuals learn. There is significant supporting evidence for this assumption from such research as the comparative study of cognitive achievement mentioned above, and studies of child psychology and adult mental health. If the assumption is true, then much of the educational research in developed countries is relevant for the developing countries. The tendency toward universality of learning phenomena is comparable to the universality of economic phenomena.

Types of Educational Research

(1) The input-output, or production-function, approach is only one of five different types of educational research; nevertheless, it is the one that has received the most publicity in the past several years. This approach measures the effects of different school expenditures on test scores. It

first received major public attention in the United Kingdom in 1965 with the publication of the Plowden Report (Central Advisory Council, 1965), then in 1966 in the United States with the publication of the Coleman Report, and finally in developing countries in 1973 and 1974 with the publication of the twenty-three-country study under the auspices of the International Association for the Evaluation of Educational Achievement (IEA) (e.g., Thomdine, 1973).

(2) Process research concentrates on analysis of classroom behavior by focusing on student-teacher interaction.

(3) Organizational research focuses on schools as educational organizations that have to satisfy multiple goals and demands from various constituencies. The basic question is how to make the schools more innovative, adaptive, and flexible as social demands increase and as the composition of the student body changes. In the organizational studies, the outputs are measured by innovativeness and responsiveness, while the inputs focus on the rules of the school, incentives, procedures, and leverage.

(4) Rather than examine the effect of any particular intervention, as the input-output studies do, evaluation studies ask whether large-scale intervention in the schooling process has had an effect in general. These studies, in contrast to the input-output studies, increase the amount of resources devoted to each child and then measure the effect on student achievement.

(5) The experiential approach describes the way that the experience of schooling affects the student in relation to himself, his peers, authority, and social institutions. The measure of the experience is not educational outcome, as indicated by achievement tests, but rather the effect of the school experience on people's lives, both while they are students and afterwards.

(6) A final dimension of the research on schooling and learning focuses on the non-cognitive outcomes of schooling. This review is shaped by available literature concentrating on schooling in relation to achievement test scores. Because we have defined learning as that learning which is measured by cognitive achievement tests, we cannot dwell on the subject of non-cognitive learning. Yet it is increasingly believed that the non-cognitive outcomes of schooling--the changes in student behavior that are not measured by the achievement tests--are more important in terms of lifetime earnings, productivity, and satisfaction than the cognitive outcomes of schooling. Therefore, we cannot bypass an essential dimension of learning and must look at the literature that relates school inputs to personality development in children of such traits as discipline, persistence, flexibility, and creativity.

The combined data for each of these types of studies is reviewed in the limited space of twenty pages, and represents a summary of more than five hundred pieces of research done over the last twenty years in the United States alone.

International Comparisons

There were four developing countries among twenty-three nations that were tested for school achievement. These included Chile, India, Iran and Israel.

(1) The level of reading comprehension of both ten-year-olds and fourteen-year-olds in developing countries was surprisingly low. Significant numbers of even the fourteen-year-olds had to be considered semiliterate. They did not approach being able to read and understand national newspapers. This is a very disappointing finding since the majority of students who get any education at all in developing countries do not stay in school past the age of twelve or fourteen.

(2) The data provide interesting information on the value of automatic promotion of students from one year to the next. 1/ In the United States, where mass education is most extensive and automatic promotion virtually universal, the top 9 per cent of the students did better than their peers in all other countries. Thus, opening education to the mass of the population did not seem to affect the quality of the best students. Conversely, systems that were highly selective and promoted only a tiny number of upper-income, high-IQ students appear to pay a high price in terms of lost talent and social dislocation. In Germany, for example, only one percent of children from lower-income homes were found in school at the age of 18 whereas there were fourteen times as many in the United States.

(3) The test of science achievement indicates that boys show a greater interest in science than girls and also perform better on the science tests. These differences exist across all countries, and the differences increase as the students grow older. Since there is no difference in IQ between the boys and girls, it is socialization which has oriented girls away from science; thus a significant pool of talent is lost that could be used for increasing the number of scientists. This is particularly iron in the poorer countries that have such an undersupply of scientists, both for teaching and application.

(4) The input-output studies of achievement within countries provide results that are consistent with findings in the United States and England. School variables, like the type of school, the teacher's experience, and school equipment, with several important exceptions, tended not be be significant in predicting achievement test scores. The home background of the student, which included father's and mother's education, father's occupation, number of books in the home, and the size of the family, was significant.

The major distinction between the IEA study and the earlier input-output studies was that the IEA study tested a greater range of cognitive abilities: for example, English as a foreign language and science. The

---

1/ The consequence of automatic promotion is that students who are terminatting the last year of secondary school have lower mean IQs and social-economic status than do students in countries where automatic promotion is not carried out.

previous studies had concentrated on reading comprehension, since reading comprehension scores tend to correlate highly with other scores of cognitive achievement. The exception noted above and the major distinction between the previous studies and the IEA studies were that science, and both French and English as a foreign language, tended to be significantly related to amount of schooling and learning conditions. Why should these differences between learning conditions be more importnat for science, French, and English than for reading comprehension, literature, and civics? Is it that the first three are more school-oriented in that specialized knowledge is being learned where there exists little knowledge in the home and therefore little backing? Or is it that reading comprehension is not systematically taught in schools once decoding and the mechanics of reading have been accomplished? The answers are not clear.

Findings

In considering the following conclusions one should have the limitations of the research clearly in mind. These limitations are discussed both in the review of the individual types of studies and in Section VI, "Limitations of Available Research."

(1) The first major implication of the educational research is that research has not identified a variant of the existing system that is consistently related to students' educational outcomes. (A "variant" of the system is used to describe the broad range of alternative education practices, which includes changes in school resources, processes, organizations, and aggregate levels of funding.) It is important to emphasize that this conclusion does not suggest that nothing makes a difference, or that nothing would improve student achievement. Rather, we are saying that the research has found nothing that consistently and unambiguously makes a difference in student cognitive outcomes. The literature contains numerous examples of educational practices that seem to have affected student outcomes significantly. The problem is that there are invariably other studies, similar in approach and method, that find the same educational practice to be ineffective. And we have no clear idea of why a practice seems to be effective in one case and is apparently ineffective in another.

We also should emphasize that we are not saying that school does not affect student outcomes. We have little knowledge of what student outcomes would be were students not to attend school at all. They might be better off over their lifetimes, or they might be worse off, if they were to skip some or all formal schooling. People who have learned outside of school have not been sufficiently studied to yield any generalizations.

It is important to emphasize that the educational practices for which school systems have traditionally been willing to pay a premium do not appear to make a major difference in student outcomes. Teachers' experience and

teachers' advance degrees, the two basic factors that determine salary, are not closely related to student achievement. Reduction in class size seems not to be related to student outcomes.

(2) Increasing expenditure on traditional educational practices is not likely to improve educational outcomes substantially.

(3) There seem to be opportunities for significant reduction or redirection of educational expenditure without deterioration in educational outcomes. As we have suggested above, none of the variants of the existing educational system seem to improve educational outcomes consistently. A corollary of this is that few variants have been shown to lead to significantly worse outcomes; If these variants are not all equally expensive, then choosing the least expensive provides opportunities to redirect and even reduce costs without reducing effectiveness. We should note that while educational research has concentrated on effectiveness studies, there have been very few cost-effectiveness studies.

(4) The smaller the school system, the more innovative, responsive, and adaptive the schools are to educational and social needs. The larger the school system, in terms of the number of schools, teachers, administrators, and so forth, the more reform depends upon exogenous shocks to the school system.

(5) Educational research is seriously deficient in terms of the size, scope, and focus of the research efforts, and in the integration of research results for policy decisions. In comparison with other major economic sectors, the amount of research activity devoted to educational problems is surprisingly small. In the United States alone, the amount of resources allocated to agricultural research and development is more than four times as much, and to health research thirteen times as much, as education receives.

(6) The educational research tentatively suggests that improvement in student outcomes, both cognitive and non-cognitive, may require sweeping changes in the organization, structure, and conduct of educational experiences. This inference follows from the first four conclusions we have cited above, as well as from the testimony of the experiential approach.

Implications for Poverty Planning

(1) To Reduce Cognitive Inequality. To what extent could you equalize the cognitive ability across individuals if various inputs were equalized in the United States? This simulation approach to a basic policy question produced several interesting results: 1/

1/ Cited from Jencks et al, (pp. 109, 110, 1972).

(a) If we could equalize everyone's genes, inequality in test scores would probably fall by 33 to 50 percent (equalizing genes would still have little effect on equalizing lifetime earnings);

(b) If we could equalize everyone's total environment, test score inequality would fall by 25 to 50 percent;

(c) If we merely equalize everyone's economic status, test score inequality would fall by 6 percent or less;

(d) Equalizing the amount of schooling people get might reduce cognitive inequality among adults by 5 to 15 percent;

(e) Equalizing the quality of elementary schools would reduce cognitive inequality by 5 percent or less;

(f) Equalizing the quality of high schools would reduce cognitive inequality by one percent or less;

(g) Eliminating racial and social segregation in schools might reduce the test score gap between rich and poor children by 10 to 20 percent;

(h) Additional school expenditures are unlikely to increase achievement, and redistributing resources will not reduce test score inequality.

The achievement test data suggest that most differences in adult test scores are due to factors that schools do not control. It does not follow, however, that schools could not equalize people's test scores if they tried. They probably could. If, for example, it was decided that everyone's reading scores should approximate the current U.S. average, we would provide only one or two years of schooling to very bright youngsters, six years to youngsters who are a bit above average, twelve years to those who are a bit below average, and eighteen or more to the very slow learners. This would probably greatly reduce the inequality of reading scores.

This fact about the effect of differential amounts of schooling has implications for "equality of educational opportunity." Usually, equality of opportunity implies that everyone should get as much schooling as he or she wants. As the above example has suggested, equal opportunity in this sense would guarantee unequal results.

While many observers would reject the idea that schools should try to eliminate all variation in cognitive skill, it does not follow that schools need to accept the present degree of cognitive inequality as inevitable. If people's cognitive skills are far below national norms, they are likely to be at a significant disadvantage, not only economically, but socially and psychologically as well.

(2) To Reduce Inequality in Schooling Attainment. The research suggests that schools have rather modest effects on the degree of cognitive and non-cognitive inequality among adults. Most people find this argument difficult to accept, since highly educated people differ from uneducated people in many important ways. Most people assume that schools cause many of these differences. The research suggests, however, that people who stay in school and attend college would differ from people who drop out even if they all had exactly the same amount of schooling. The literature suggests that schools, including universities, serve primarily as selection and certification agencies whose job is to measure and label people, and only secondarily as socialization agencies whose job is to change people. The implication is that schools serve primarily to legitimate inequality, not to create it.

II. INTRODUCTION

How Effective Is Schooling in Promoting Learning?

A Review of the Research

To what extent are schools effective in promoting learning in developing countries? Ten years ago this question would not have occurred to most educational observers who were concerned with increasing the supply of school places or improving the quality of the educational inputs. Now, it is one of the questions most frequently asked at international conferences and ministerial meetings when future investment in education is discussed.

The results of recent research on education in both developed and developing countries have made this question very important for developing countries. The research has shown that schooling is not as effective in promoting learning as we had assumed. Other factors, including parental behavior, nutrition, students' personality, and chance, were more significant than previously assumed.

The purpose of this paper is to explore the research results that relate schooling to learning--a concept we will narrow to cognitive achievement. Although the recent research provides new insight into the relation between schooling and earnings, labor productivity, job promotion, and worker satisfaction, these relationships are outside the scope of this discussion.

The breadth and depth of the data for such a review vary significantly between developed and developing countries. While the richest data exist for several of the developed countries, some data are available for developing countries; but less than a dozen developing countries have useful data on the question of schooling and cognitive achievement.

The approach of this review is not the classical one of reviewing and categorizing the individual research studies. Instead, we have abstracted portions of the major reviews of the literature that had followed classical methods. In addition, we provide a review of several major studies made since the reviews were published. The major reviews are of the United States research, and the recent studies provide international comparative insight.

Overview of the Research Literature

In most developing countries the research on the relationship of schooling and learning is only beginning. Furthermore, quantitative research on other educational questions has received even less attention. Of the developed countries, the United States has the best data. Good data are also available for several European countries, especially Great Britain, but the European studies have not been examined by previous reviewers, and, with several exceptions, are thus excluded from this paper.

Why should we first focus our review of educational research on the United States, when our concern is the developing countries? The legitimacy of cross-national research is a sufficiently important and difficult subject to warrant separate discussion. Our objective in education, as it is in agriculture and other sectors, is for the developing countries to draw the optimal benefits from research that is already done, thereby enabling them to reduce their future research expenditures and to increase their effectiveness. The basic assumption is that there is more variation within a country than across countries in how and why individuals learn. The assumption is supported by significant evidence: for example, by cross-national surveys of cognitive achievement, and studies of child psychology and adult mental health, including work on low-income groups, as we will discuss in Section V. If this assumption is true, then much of the educational research in developed countries is relevant for the developing countries. The tendency toward universality of learning phenomena is comparable to the universality of economic phenomena.

Educational research has not been well funded in either developed or developing countries. Lack of research is a major reason why there are so few nationwide productivity gains in reading comprehension or science achievement. While the importance of agricultural research is well recognized for its contribution to productivity increments and is systematically funded in most developing countries, educational research is not. 1/

Education Research and Educational Policy 2/

While a considerable literature on educational effectiveness had existed, and new results were constantly being published, it was not until 1972 that many policy-makers in the United States and elsewhere became aware of it with the publishing of the research review by Christopher Jencks and his Harvard colleagues entitled Inequality: A Reassessment of the Effect of Family and Schooling in America. A second review with a narrower focus, by Harvey Averch, John Pincus, and their colleagues at the Rand Corporation was completed earlier

---

1/ Even in developed countries, educational research has been under-funded compared to research in other sectors. In the United States between the period 1953 to 1973, government expenditures on research among sectors was: $1 billion for education, $7 billion for agriculture, and $14 billion for health. For the United States, a direct relationship exists between research expenditure by crop and the increase in average national yields over time.

2/ The following sections of the paper draw heavily on the work of Averch, et at 1972, pp. 1-15, including passages that are cited in full with only minor editing.

in the same year and reached similar conclusions. Comparable published reviews of the literature do not exist for Europe or the developing countries. 1/ While the Report of the Edgar Faure Commission, sponsored by UNESCO, discusses education in both the rich and poor countries, a review of the quantitative research is virtually absent. 2/ Research on the developing countries has not yet enjoyed a synthesis similar to the recent ones in the United States, because the quality of the research is often poor and the studies few.

Four reasons help explain why educational research only recently began to have an impact on educational policy. First, there are many diverse streams of educational research. In terms of traditional disciplines, research on educational effectiveness appears in economics, political science, psychology, and sociology, as well as in education. Researchers have tended to vfollow relatively narrow, intradisciplinary paths. There have been few attempts to connect these paths; nor is there a good map for any given path. Policy-maker and researcher alike, therefore, have found it difficult to draw policy implications from these various disciplines.

Secondly, the sheer magnitude of the literature on educational effectiveness makes it difficult to keep up to date on the research being conducted in any one field, let alone to keep up with what is being produced across the entire range of educational research. More than 2,000 studies are published per annum in the United States.

Thirdly, educational research has seldom been explicitly policy-oriented. A considerable volume of research has been aimed at increasing understanding of how and why learning takes place. But the research design has rarely acknowledged the needs of decision-making.

Fourthly, the synthesis of research findings contains contradictory findings. The policy-maker thus finds himself constantly basing his decisions on controversial and disputed research results.

The Averch group determined for each research effort whether the study was internally valid (did the researcher pursue proper methods for the questions he addressed?). If it was valid, the group asked whether the results were credible in the light of accumulated knowledge (were the findings consistent with those of other studies in the area?).

The credibility, or external validity, of the research findings is as important as the proper methodology, or internal validity, of the research project. There is always some chance that a particular variable or set of variables that appears to have a significant effect upon achievement is in

---

1/ For an unpublished review of the achievement literature of developing countries see Alexander and Simmons (1974).

2/ For a review of the Faure Report, see Simmons (1973).

fact unrelated to educational outcome. For example, shorter students are higher achievers. For this reason, educational policy decisions cannot rely on only one study. Whether studies say anything about actual educational outcomes depends, then, on results that appear consistently throughout a number of studies. If an educational resource or procedure shows up as significant in a number of well-designed studies, then we should be able to state with considerable confidence that this resource or procedure should be seriously considered by policy-makers.

An examination of credibility serves three distinct purposes. First, it provides a way of summarizing numerous disparate studies. Secondly, it addresses the question of what should be believed in the face of inconsistent or conflicting results. Essentially, we resolve such conflicts by "adding up" the evidence on each side of a dispute. Thirdly, consideration of external validity enables us to deal with the avalanche of research results. No review, Averch and Jencks included, could possibly consider every study since some are unpublished and inaccessible. But if a large number of internally valid studies yield consistent results, then one can be fairly sure that the omission of one study would not have changed one's conclusions substantively. This does not mean that a new theory and evidence could not significantly change the conclusions: the history of scientific investigation suggests this is a possibility. But any drastic rewriting of conclusions would imply a significant change in either how children learn or how learning is measured.

None of the many ways of synthesizing evidence from dissimilar studies 1/ is entirely satisfactory. In the Averch review of the literature, each study can be considered as a witness presenting testimony. The test of internal validity can be compared to cross-examination. Inter-study consistency is determined subjectively, judgment being based upon the accumulated evidence presented by many diverse research reports. The Averch group counted the number of studies in which an independent variable was examined and calculated the probability of significance, but they did not simply weight each study equally. Rather, they were more persuaded by studies that used larger samples or more replications, better designs, a greater number of controls for intervening variables, or more accurate measures of the variables in question.

The Scope and Limitation of the Analysis

An education system has many functions and outputs. Some outputs relate directly to the student, others hardly involve him at all. For example, the school system must interact with the community and must provide a number of outcomes relevant to the community. In doing so, the school may sometimes act in ways that seem to operate against outcomes desired by the student. For example, the community may need mechanics but the student's abilities might

---

1/ See Light and Smith (1971) for a discussion of this problem.

not be mechanical and he or she might prefer office work. The school also has a political role and must provide outcomes that allow it to compete within a political system for power, money, and position. Whatever importance one assigns to political and social functions, it seems that they are not the school's primary objective, which is to educate students. The review focuses on research into the determinants of student learning.

What exactly does student learning mean? The easiest and perhaps the first definition that comes to mind is to interpret learning as the acquisiton of knowledge and cognitive skills. In practice, this has mainly been narrowed by using standardized tests for measuring retention of specific subject matter; higher cognitive processes (abstract reasoning, problem solving, and creative thinking, among others) are seldom measured (Klein, 1971). Teacher grades and essay examinations are sometimes used as measures of broad cognitive abilities, but these measures are extremely unreliable. Along with the general failure to measure cognitive achievement adequately, there is an almost total failure to evaluate and identify "non-cognitive achievement." 1/ Thus, of the many and diverse kinds of behavior that students learn, like discipline, creativity, and cooperation, most are not measured by standardized and popular tests.

By and large, researchers have not employed broad measures of student learning, nor have they resolved the important problem of the priorities among educational outcomes. One does find, however, that many of these same researchers who have not been able to resolve this problem analytically, nevertheless frequently discuss the importance of priorities and individual differences in priorities. It is becoming increasingly clear that different educational objectives and values exist as well as individual differences in types and levels of ability. We must, therefore, realize that research based on limited measures, and accounting for relatively few objectives, cannot lead to conclusive generalizations about educational outcome.

Although implicit criteria are inevitable whenever effectiveness is discussed, the Averch review avoided any explicit discussion of the aims of education, for two reasons. First, a study of the aims of education was not part of their charter. Secondly, the Averch group is reticent to address the aims of education because the researcher is no more competent to solve these issues than is any other citizen. The question is one of values. Nonetheless, certain issues are necessarily raised:

(1) To what extent should education be an agent of social reform as compared with a force for social stability?

1/ This expression is employed here because it is gaining use in the educational literature. Actually, however, it would be more accurate to talk about "non-cognitive growth," rather than "achievement," but debate over terms seems relatively unproductive so long as it is generally understood what the term "non-cognitive achievement" means. In particular, in "non-cognitive achievement," we include the concepts traditionally described by the term "affective domain." For a review of the literature that does identify "non-cognitive achievement," see Gintis (1971).

(2) To what extent should education be oriented toward vocations, to personal development, to the pursuit of knowledge, to screening people by ability categories?

These issues are inseparable in the consideration of any educational policy.

III. TYPES OF EDUCATIONAL RESEARCH

Methods and Findings

The Input-Output Approach

Method. Input-output studies are distinguished by a view of the educational process that holds that a student's educational outcome (defined as cognitive achievement) is determined by: (1) the amount of resources that his school makes available to him; (2) the personal, family and community characteristics that influence his learning -- typically grouped under the term "background factors"; and (3) the influences of his peers. This approach views the school in which the student is enrolled as affecting his outcome only to the extent that the school serves as the channel through which resources flow to him. It neglects, in particular, the structure and organization of the school and classroom.

The educational "production function" is a formal representation of the relationship between school resources and background factors on one hand, and student outcomes on the other. It is commonly expressed in the form of an equation:

$$0 = g(r1,\ldots, rn, f1,\ldots, fm, p1,\ldots,pk)$$

in which there are assumed to be n relevant school resources, m relevant background factors, k relevant peer-group influences, 1/ and in which

0 = a student's output (e.g., his score on a standardized achievement test);

r1,..., rn = the amounts of school resources 1 through n, respectively, that he received (e.g., resource 1 might be the ability of his teacher, resource 2 the size of his class, and so on);

f1,..., fm = the amounts of background factors 1 through m, respectively, that the student has been exposed to (e.g., f1 might denote his family's income, f2 his father's occupation, and so on);

p1,... pk = the amounts of peer-group influences 1 through k, respectively, that the student has been exposed to (e.g., p1 might denote the proportion of his classmantes that intend to go to college, p2 the proportion of his classmates that are members of minority groups, and so on).

1/ Some researchers prefer the term "student body effects."

The educational production function is expressed in this most general form in the above equation, which merely states that for any particular student, described in terms of his background factors, the amounts of school resources he receives and the influences of his peers determine his outcome. In order to make a quantitative estimation of the impact of any particular resource upon outcomes, the precise relationship between inputs--resources, factors, and peer-group influences--and outcomes must be specified. Conceptually, any one of an infinitely large set of possible relationships can be specified. In practice, however, only one functional form--the linear one--has thus far been employed in educational production-function studies. But this is more a reflection of the limitations of current statistical techniques than the result of any consenus about the underlying nature of the educational process.

The objective of the research is to estimate coefficients for independent variables. If we knew these values, we could predict the impact of providing students with more or less of any particular school resource. This would allow us to determine whether increasing (or decreasing) the amount of any one school resource would affect students' outcomes more or less than increasing (or decreasing) the amount of any other school input. Taking account of the relative prices of the various school resources, we could then determine how much of each school resource should be purchased to attain any particular goal for student outcomes at minimum cost. In short, we could formulate optimal educational policies. 1/ However, the costs of obtaining school resources have not yet been incorporated into empirical analyses. Estimates of educational production functions--the topic to which we devote the remainder of this section--are only the first step toward an educational policy.

Multiple regression analysis is used to estimate the values of the coefficients for the parameters in the equation above. Details of the technique can be found in any statistics text. 2/ A multiple regression analysis provides for tests of the "significance" of the empirical results. These are formal measures of the accuracy of the results in the sense that they indicate how much confidence can be placed in them. In educational production-function studies the analyst is typically concerned with identifying resources or factors that affect student outcomes. He is concerned with identifying inputs where coefficients have non-zero values. To say that the coefficient of a variable is significant means that the test of significance indicates a small probability that that particular coefficient is zero; just how small is referred to as the significance level.

1/ The role of resource prices in the formation of educational policy is often overlooked. Even those researchers who recognize the importance of resource prices in their theoretical discussions do not introduce them into their empirical analyses. But school resources are not free and school systems do not have unlimited budgets. Consequently, the important questions from the viewpoint of the educational policy-maker are not: How much does resource 1 contribute to student outcomes? and so on. Rather, educational policy-makers must ask: How much of resource 1 should be purchased? and so on. For a discussion of this issue see Cain and Watts (1968).

2/ See, for example, Wonnacott and Wonnacott (1970).

The basic assumption underlying all studies in the input-output approach is that the production function is an equally accurate description of the educational process for all students, or at least for some identifiable subgroup of students. In other words, the unit contribution of any given resource, factor, or influence to student outcome is assumed to be approximately the same for all, or some subgroup of, students. This assumption implies that if any particular resource or factor does have a significant impact on student outcomes, the coefficient of that resource or factor should be significant in any study that examines it. Otherwise, every student must be different or respond differently to the same resources.

There is always some possibility that a variable that appears to have a significant impact upon student outcomes may, in fact, be unrelated to outcome. Therefore educational policy cannot be based on the results of any one study. The basic assumption of production-function analysis reinforces this point. We do not emphasize the results yielded by any one study. Rather, our primary concern is to identify results that consistently appear throughout a number of studies.

Findings. Overall, the input-output studies provide very little evidence that school resources in general, even in developing countries, 1/ have a powerful impact upon student outcomes. When we examine the results across studies we find that school resources are not consistently important. The particular resources that seem to be significant in one study do not prove to be significant in other studies that include the same resources in the analysis.

Background factors, on the other hand, are always important. In study after study a student's background has a strong influence on his educational outcome. Furthermore, the results are consistent across studies. The socioeconomic status of a student's family--his parents' income, education, and occupation--invariably prove to be significant predictors of his educational outcome.

The role of peer-group influence is more complex. There is good reason to believe that these variables are, in reality, measures of a student's background or the selection procedures used by the schools. On balance, there is little evidence that a student's classmates exercise a strong, independent influence on his educational outcome.

The results from the input-output approach do not mean that school resources fail, actually or potentially, to affect student outcomes. We

1/ For a review of these and other developing country studies, see Alexander and Simmons (1974). For a full analysis of each of the relevant studies, see Averch et al. (1972), and Jencks et al. (1972). Examples of these studies are Mollen and Melville (1956), Thomas (1962), Coleman (1966), Katzman (1968). See Averch, Appendix A (1972), for the comparative results of these other studies. For European input-output results see Plowden (1965), Husen (1967) and IEA (1973). For examples from developing countries: Tunisia, Simmons (1972); Israel, Ortar (1967); India, Chopra (1967); and Puerto Rico, Farquhar and Christensen (1968).

simply observe that so far these studies have failed to show that school resources do affect student outcomes. In particular, the studies do not show what would happen if the educational system received a massive increase or decrease in resources.

The Process Approach

Method. The approach of the psychologist focuses on a very different aspect of education. Resources are taken as given or predetermined. What matters here are the processes applied to students and the interactions between teachers and students. For example, research may concentrate on the relation between teaching style and student achievement, or on the effect of grouping on achievement.

The review of research covers studies of the educational process undertaken in the classroom as well as studies made in the psychological laboratory that appear to have relevance for the educational process. Laboratory and classroom studies are distinguished in this report not so much on the basis of where the study took place as on the basis of the study objectives, the learning tasks, and the kinds of outcome measures. Classroom studies involve meaningful teaching activities and have the objective of improving our understanding of education in the classroom. Some measure of educational outcome is generally used (achievement tests, grades, and teacher or supervisor ratings). Laboratory studies generally have more theoretical objectives, such as advancing knowledge about psychological phenomena, testing theory, or investigating empirical relationships between psychological variables.

Measures of outcome in laboratory studies are varied and difficult to summarize; they are, however, generally based on the learning or retention of well-defined and highly specific responses. The experimentalist is not primarily concerned with the amount learned but with the way in which the learning takes place and the factors that affect learning or retention. For example, an experimenter might present both auditory and visual stimuli in pairs to children to investigate the different effects of each type of stimulus on learning and retention. The stimulus pairs are presented to each child until he can recall without error the second stimulus in each pair upon presentation of only the first. The measure of learning is the number of presentations necessary before the child has learned the list of stimulus pairs without error. This measure is studied across age groups to determine whether age-related differences exist in the learning of visual or auditory stimuli.

The reader should realize at the outset that classroom and laboratory studies differ greatly in their objectives and approaches. Classroom studies have not generally produced highly definitive results. Laboratory studies, however, have produced many significant and consistent results, but their relevance for classroom learning is often not clear.

Findings. We have divided the results into two parts: those derived from studies of operating classrooms and those derived from the laboratory. For

each set of results, we indicate the focus, the questions being asked, and the answers to the questions. 1/

Looking first at the classroom studies, we find the following:

(1) The research on teaching approaches, teacher differences, class size, and the like shows no consistent effect on student achievement, as measured by standardized cognitive tests.

(2) Work on instructional methods suggests no difference in effectiveness among methods; no innovative method currently appears better than conventional methods. In other words, in terms of differences in achievement, conventional methods appear as effective as, say, teaching by television, although the latter enables one to reach far greater numbers of students.

We consider the following results from the laboratory studies to be particularly interesting and important:

(1) Work on the presentation of material suggests that it is not so much the medium of instruction that is important as its sequencing and organization. There seem to be interaction effects; individual methods of presentation appear superior for some students, but it is still hard to match student characteristics, tasks, and type of instruction.

(2) The work on concept attainment, retention, and learning rewards provides a number of positive findings, but the tasks in the laboratory are so unlike classroom learning that there is a difficult problem of translation. For example, the more meaningful the material, the faster it is learned and the more it is retained. But the definition of "meaningful" is a laboratory one, relating, say, to the difference between nonsense sentences or syllables and those that make sense.

(3) What are termed interaction effects seem to exist among various types of personality, methods of reward, ability to grasp meaningful material, and so on; but these interactions have not yet been studied in detail.

In sum, the process approach has not identified the very specific student relations involved in learning and education. There seem to be interactions between students and teachers, between students and methods, between teachers and methods, and (most complex of all) among students, teachers, and methods. The complex three-way interactions have not yet been studied carefully.

The Organizational Approach

The work on educational organizations views schools as institutions that have to satisfy multiple goals and demands from internal bureaucracies,

---

1/ For examples of the studies see Rosenshine and Furst (1971), Getzel and Jackson (1963), Rosenthal and Jackson (1968), and Westbury (1970).

from the community, from parents, and from students. The allocation of resources and the choice of processes in schools are seen not as the result of a rational decision-making procedure but as the outcome of history, of interactions with constituents and with government, and of simple trial and error.

The focus on the output side is on determinants of innovativeness and responsiveness, and the focus on the input side is on rules, incentives, procedures, and leverage. The question being asked is: How can we make the schools more innovative, adaptive, and flexible as social demands increase and as the composition of the student body changes?

Findings. Most of the work in this approach consists of case studies, and the rules for internal and external validity are weak at best. 1/ Furthermore, there have been few attempts to extract important organizational propositions from the literature. The case studies provide some evidence for the following.

(1) There is a positive correlation between system size and centralization.

(2) The larger the educational bureaucracy and the greater the centralization, the less innovation and adaptation there is likely to be.

(3) Rigidities in the schools can be overcome partly by choice of teachers and principals; however, teacher qualities that are purchased--say, training or experience--tend not to result in innovative teaching.

(4) Real innovation depends on the leverage that can be exerted from outside the system--by the government or by citizens.

The Evaluation Approach

Method. This approach to educational research consists of ex post facto analyses of comprehensive interventions in existing school systems. These studies are characterized by a macroview of educational interventions in which treatements are devoted to groups of children in "diverse programs taken as a whole." In short, these studies ask whether large-scale interventions have had an effect in general, rather than what has been the effect of any particular intervention.

In such interventions the resources devoted to each child are normally increased substantially. This can take the form of smaller class sizes, additional instructional personnel (often specialists or paraprofessionals), more individualized instruction, or more intensive use of audio-visual equipment. Since any number of educational inputs are changed at the same

---

1/ For examples, see Anderson (1968), Havighurst (1964), and Rogers (1968).

time, it is difficult to tell precisely which program features are responsible even when there is demonstrated success. Control-group perfection has naturally been sacrificed to the more pragmatic goal of educating the children who need it most. Usually, the emphasis has been on achieving program goals and not upon the needs of careful research and evaluation. Because of this the research designs of such broad programs of educational intervention are often much less precise than, and the available research materials concerning them are considerably inferior to, those used in the process approach discussed above.

Findings. Virtually without exception, all the surveys made of large, national compensatory education programs have shown them to have no beneficial results on average. 1/ The evaluations on which the surveys report, however, are often based upon suspect research designs.

Two or three smaller surveys show modest positive effects of compensatory education programs in the short run: and a number of quite carefully designed interventions display gains in pupil cognitive performance--again, in the sort run. In particular, pupils from disadvantaged socioeconomic backgrounds tend to show greater progress in more highly structured programs. There is considerable evidence, however, that many of the short-run gains from educational interventions fade away after two or three years if they are not reinforced. Also, this "fade-out" is much greater for the more highly structured programs--the ones which are most unlike regular public school practice.

The Experiential Approach

Method. The experiential approach is represented by the literature of education reform. 2/ The observer, either as researcher or participant, describes the way that the experience of schooling affects the student in relation to himself, his peers, authority, and social institutions. The measure, for these writers, is not educational outcome as indicated by standardized tests but rather the effect of the school experience on people's lives--which is not measurable by cognitive testing.

The approach states, in effect, that the most important thing about schooling is the way in which school experiences affect students' lives and self-concepts, both while they are students and for the rest of their lives. Therefore, to these authors, the other approaches discussed in this report--input-output, process, organizational, evaluation--are all essentially irrelevent unless they affect: (1) the student's concept about himself as an

---

1/ For examples, see Bissell (1970), Wolff and Stein (1967), and Wrightstone *et al.* (1967).

2/ For examples, see Friedenberg (1963), Henry (1963), Holt (1964), Coodman (1965 and 1970), Herndon (1968 and 1971), Kohl (1968), Kozol (1968), Denison (1969), Silberman (1970), Postman and Weingartner (1969), Illich (1971).

individual and as a member of the society (classroom, school, community, and so on) that impinges on him; (2) the style that the student develops to deal with school experiences (notably teacher-student and student-student transactions); (3) the attitudes toward social institutions that the student develops as a consequence of his first major experience with one such institution--in this case, the school system.

The reform writers don't believe that cognitive skills are unimportant. Rather, what they generally do believe is that the nature of the school experience is a dominant factor that determines not only how well cognitive skills are acquired, but also how effectively they can be used after school. Many of the reform writers also raise serious questions about whether the cognitive skills that the schools actually do transmit are helpful or harmful to individual development for example, Henry, 1963, pp. 287-288; Silverman, 1970, Part II). A number of research studies have come to similar conclusions concerning the value of cognitive skill using other approaches, as shown in the review by Gintis (1971).

Findings. Because this literature is one of social reform, it is not subject to the same tests of internal consistency as the approaches discussed above. In effect, there are two elements in this literature, description and prescription. 1/ The description of the schools as constituted at the present time almost invariably emphasizes a set of common themes:

(1) Schools are authoritarian toward students.

(2) Schools make little or no allowance for individual differences in learning styles and needs.

(3) Schools focus on methods that stress rightness and wrongness in learning, thereby destroying independence and creativity, as well as equipping children poorly for the complexities and ambiguities of the real world.

(4) Schools impose a certain set of social, cultural, and ethical views on their students, thereby arousing feelings of inadequacy and resentment in those who share neither those views nor the traditions they imply.

(5) Schools as institutions often fail, in any operationally useful way, to question either the assumption upon which they operate or the relevance of their approach to children's needs.

1/ The writer's triple role as observer, participant, and social critic necessarily places heavy pressure on his objectivity in describing the phenomena he observes. Nonetheless, there seems to be considerable agreement among writers with respect to description.

(6) These themes and school outcomes are consistant with society's desire to use schools to socialize students to future work roles.

The prescriptions are far more varied than the descriptive research. They range from recommendations for moderate reform within the system (Silberman) to abolition of the schools (Illich). In some cases, the value systems leading to the prescriptions are made explicit, in others not. In general, however the experiential literature agrees on the merits of educational systems that are less rigid, more responsive to individual diversity, and more decentralized than most current systems. These changes, however would not be consistent with the school's role of socialization for future work.

Non-Cognitive Outcomes

It is clear why minimum levels of cognitive achievement are needed in most occupations, but why should non-cognitive outcomes be a major concern? Common sense, supported by the research of industrial psychology, suggests that job performance is a function of non-cognitive traits--such as discipline, persistence, docility, and flexibility--as well as cognitive abilities. Recent educational research suggests that these same traits contribute to high grades in school. Furthermore, schooling reinforces and elicits the non-cognitive traits. Since non-cognitive traits are school outcomes that are encouraged by the schools and important to cognitive achievement, they deserve major concern. That these school-developed traits also contribute to job promotion and lifetime earnings is equally important but outside the scope of this review. 1/

We shall focus on two aspects of schooling central to the pattern of personality development. 2/ First, we will discuss the structure of social relations in education, including sources of motivation, authority and control and types of sanctioned interpersonal relations. The social relations of education require students to function routinely and over long periods of time in role situations. These situations involve specific expectations on the part of the teacher, other students, and administrators which tend to elicit uniformity of response and are then codified in the individual personality.

Two studies are useful for illustrating the development of non-cognitive traits in school. The first is an analysis of American data colledted on 649 upper-ability senior high school males (Holland, 1963). Student grade-point averages (GPA) estimated by teachers were predicted on a basis of scores on achievement tests that were not graded by the teachers. IQ was controlled. These tests were the College Entrance Examination Boards: math aptitude, verbal aptitude, scientific performance, humanities comprehension, and scientific comprehension. The results were surprising. Despite significant variance in the achievement test measures and in the grades, the achievement

---

1/ For further information see Kohn (1964) and Gintis (1969).

2/ The rest of this section is an edited version of Gintis (1971), pp. 266-76.

measures do not have significant effect in predicting grades. Some 65 additional personality variables were added to the regression to predict grades. Two of them--"citizenship" (CitT), a teachers' rating, and "drive to achieve" (DrA), a student self-rating--had the greatest power to predict grade-point average.

This example illustrates that: (1) teachers grade independently on the basis of personality, since these personality traits are not rewarded through their contribution to the achievement scores; (2) subjective motivation is taken into consideration in grading, since drive to achieve is rewarded; (3) grading reinforces the student's personality development to enlarge participation in the particular structure of social relations in schools, since citizenship is positively rewarded and can be interpreted as conforming to the dominant role structure of the school; and (4) while grades depend on achievement in general, when ability (IQ) is controlled as we have done with the achievement tests, little additional effect of achievement can be detected; so the subjective experience of the individual student, who, of course, cannot control his intelligence, is that grades depend primarily on affective behavior.

In this study the pattern of reward is no less reflected in the remaining personality traits that are discussed in the tables below. Thus, table 3.1 shows that students are uniformly penalized for creativity, autonomy, initiative, tolerance for ambiguity, and independence, even after correcting for academic achievement, citizenship, and drive to achieve. They are rewarded for perserverance, good student values, and other traits that are indicative of docility, industry, and ego-control.

The content of the citizenship rating by teachers is exhibited in the correlations in table 3.2, showing a similar pattern of behavior evaluation on the part of the teacher, especially in the penalized traits. There is little doubt that this evaluation is a result of the needs of "classroom control" in the typically structured school, rather than the teacher's own personal value preferences.

Table 3.3 shows that the drive for achievement is associated with the same pattern of penalized traits, while the rewarded traits exhibit two separate dimensions: on the one hand, high drive for achievement may involve conformity with classroom norms; and on the other, it may involve the rejection of norms in favor of autonomous personal development--hence the association of artistic performance, creative activities, self-confidence, initiative, self-assurance, and breadth of interests with drive for achievement. 1/

1/ See Gintis (1969) for definition of the variables.

III. Table 1.

Personality Variables Correlated with GPA Corrected for Achievement CitT, and DrA (Significance Levels in Parentheses)

Positively Rewarded:

Perseverance (1%)
Good Student (1%)
Self-Evaluation (5%)
Popular (5%)
Acceleration of Development (5%)
Mastery (5%)
Control (6%)
Status (11%)
Popularity (TR) (13%)
Suppression of Aggression (14%)
SAT-Math (15%)
Scientific Comprehension (15%)

Negatively Rewarded (Penalized):

Independence-Self-Reliance (1%)
Initiative (5%)
Complexity of Thought (5%)
Originality (Barron) (6%)
Originality (11%)
Independence of Judgment (13%)
Creative Acitvities (13%)
Curious (14%)

Table III.2

Correlations of Various Personality Traits with CitT

Positively Rewarded:

Deferred Gratification (1%)
Perseverance (1%)
Control (1%)
Popularity (1%)
Social Leadership (1%)
Good Student (Parent Value) (1%)
Self-Evaluation (1%)
Scientific Comprehension (1%)
Intellectuality (1%)
Aesthetic Sensitivity (5%)
Suppression of Aggression (5%)
Comradeship-Sharing (5%)
SAT-Math (10%)
Artistic Performance (10%)
Mastery (15%)
Initiative (15%)

Negatively Rewarded (Penalized):

Cognitive Flexibility (1%)
Complexity of Thought (1%)
Originality (Barron) (1%)
Sense of Destiny (1%)
DRS-Creativity (1%)
Independence of Judgment (5%)
Independence-Self-Reliance (10%)
Curious (15%)
Self-Confidence (15%)
Verbal Activity (15%)

Table III.3

Correlations of Various Personality Traits with DrA

Positively Rewarded	Negatively Rewarded (Penalized):
Self-Evaluation (1%)	Cognitive Flexibility (1%)
Perseverence (1%)	Complexity of Thought (1%)
Deferred Gratification (1%)	Originality (1%)
Originality (1%)	Independence of Judgment (1%)
Independence (1%)	SAT-Verbal (5%)
Responsibility (1%)	
Control (1%)	
Artistic Performance (1%)	
Creative Activities (1%)	
Sense of Destiny (1%)	
Popularity (1%)	
Social Leadership (1%)	
Good Student (1%)	
Initiative (1%)	
Status (1%)	
Breadth of Interest (5%)	
SAT-Math (5%)	
Scientific Performance (5%)	
Verbal Activity (5%)	
Conformity (10%)	

Source: Holland (1963)

The above study is weak in two respects: it deals with only one ability grouping, and it aggregates over diverse study areas--for example, natural science and social sciences. A similar result can be derived, however, from a path analysis that Gintis fitted to American data supplied by Cline (1963) covering 114 high-school seniors of varying ability in a specific area of natural science performance. This data includes a measure of intelligence, three creativity measures, achievement level in science, a teacher rating of the student's science potential, and average science grades over the three years of high school.

A path analysis using multiple least squares regression equations (Gintis, 1969) indicates that over the broader ability spectrum of students: (1) teacher sttitudes are the major determinants of grades; (2) academic achievement is only one determinant of teacher attitudes and hence of grades received; (3) intelligence is directly rewarded in terms of grades beyond its contribution to actual achievement, whereas many equally important determinants of achievement--creativity, for example--are in no way rewarded; (4) the direct effect of actual achievement on teachers' attitudes is statistically insignificant.

The bulk of existing studies are compatible with these results. Furthermore, these studies show that both structure and pattern of reward in schooling conform to the requisites of adequate job performance in bureaucratically structured and hierarchically organized enterprise (Kubiniec, 1970; Weber 158a and 1958b).

These and other studies 1/ are strong evidence that the structure of social relations in schools have reproduced rather faithfully the work environment. Learning (the activity) is not undertaken through the student's intrinsic interest in the process of learning, nor is he motivated by the goal of the educational process (possession of knowledge). Thus, the student learns to operate efficiently in an educational environment, unmotivated by either the process or product of his activities--in short, in a alienated educational environment in which rewards (grades, class standing, and threat of failure) are in all cases external to the individual. The development of such forms of external reward is a prime outcome of educational socialization, and an important contribution to productive worker characteristics (Dreeben, 1968).

Economists have long noted the relationship between the level of schooling in workers and their earnings. Almost no attempt has been made, however, to determine the mechanism by which education affects earnings or productivity. In the absence of any direct evidence, it is commonly assumed that the main effect of schooling is to raise the level of cognitive development of students in that it is this increase which explains the relationship between schooling and earnings. Since our main concerns in this section have been the outcomes of schooling rather than the contribution of schooling to earnings, we do not further develop this interpretation; but the evidence presented above suggests that it needs to be fundamentally questioned. 2/

1/ Notably those showing the non-cognitive traits rewarded in the work place.

2/ See Gintis (1971) for a full discussion and evidence. See also Bowles and Gintis (1972) for the effect of IQ on lifetime earnings, and Simmons (1975) for evidence from developing countries.

IV. POLICY QUESTIONS

To Reduce Cognitive Inequality

Jencks et at. (1972) used the same input-output studies as the Averch group (1972) and Kieserliog (1971), plus other research especially from the literature on occupational mobility, learning and intelligence. 1/ They used their review of the literature to answer the question: To what extent could you equalize the cognitive ability across individuals, if various inputs were equalized?

Before discussing the results, we must point out that the model used for IQ determination is a simple additive relationship between the effect of genes and the effect of environment. While there are good logical reasons for rejecting additive models, there is not much empirical experience that any other models work better. 2/

The Jencks group review of the U.S. data suggests that: 3/

(1) If we could equalize everyone's genes, inequality in test scores would probably fall by 33 to 50 percent.

(2) If we could equalize everyone's total environment, test score inequality would fall by 25 to 40 percent.

(3) If we merely equalize everyone's economic status, test score inequality would fall by 6 percent or less.

(4) Equalizing the amount of schooling people get might reduce cognitive inequality among adults by 5 to 15 percent, although this estimate is very rough.

(5) Equalizing the quality of elementary schools would reduce cognitive inequality by 5 percent or less.

(6) Equalizing the quality of high schools would reduce cognitive inequality by 1 percent or less.

(7) Eliminating racial and socioeconomic segregation in the schools might reduce the test-score gap between black and white children and between rich and poor children by 10 to 20 percent.

1/ For a discussion of IQ, it is determinants and consequences, see the Jencks et al. (1972), Appendix A, "Estimating the Heritability of IQ Scores," and Bowles and Gintis (1972).

2/ See Jencks et al., pp. 114, footnote 31, and Appendix A.

3/ The Jencks results are a citation from the text, with minor editing pp. 109-10.

(8) Additional school expenditures are unlikely to increase achievement, and redistributing resources will not reduce test-score inequality. 1/

Jencks discusses the implication of these findings. Most differences in adult test scores are due to factors that schools do not control. It does not follow, however, that schools could not equalize people's test scores if they tried. They probably could. If, for example, it were decided that everyone's reading scores should approximate the present U.S. average, we would provide only one or two years of schooling to very bright youngsters, six years to youngsters who were a bit above average, twelve years to those who were a bit below average, and eighteen or more years to the very slow learners. This would, we suspect, greatly reduce inequality of reading scores. Jencks does not, however, favor such a solution. In sum, we tend to think of "equal opportunity" as implying that everyone should get as much schooling as he wants. Equal opportunity, in this sense, guarantees unequal results. 2/

While many observers reject the idea that schools should try to eliminate all variation in cognitive skill, it does not follow that schools need accept the present degree of cognitive inequality as inevitable. Jencks argues that if people's cognitive skills are far below national norms, they are likely to be at a significant disadvantage, not only economically but socially and psychologically as well. At least in a highly competitive society like the United States, an individual who cannot read even simple instructions, or who cannot do enough arithmetic to tell whether he has been shortchanged, is likely to be exploited in a variety of ways. Relatively few American students leave school in this condition--roughly 5 percent of all school-leavers. But in developing countries it is roughly 70 percent. (See Section V).

Unfortunately, few discussions of schooling and inequality focus on these extreme cases. When people talk about the schools' failure to prepare disadvantaged students for modern economic life, they are not usually talking about the handful of illiterates and innumerates, but about the much large number of students who leave high school at the twelfth grade but read at the eighth- or ninth-grade level. These students are by no means unemployable,

---

1/ See the Jencks et al. reference in footnote 157, p. 130, for a derivation of the numbers.

2/ If unequal performance on standardized tests were a principal cause of inequality in other realms, this traditional dcotrine might need reexamination. The relation between cognitive ability and economic success is not discussed in this paper. Such a discussion, based on the reviews of U.S. data, would conclude that there is nearly as much economic inequality among individuals with identical test scores as in the general popualtion. Thus, we can hardly suppose that making everyone's scores equal would appreciably reduce economic inequality in the general population.

nor are they automatically excluded from the mainstream of American life. They are not likely to become physicians or physicists, but this would be true even if they were reading at the twelfth-grade level. At least in economic terms, the cost of reading at eighth-grade rather than twelfth-grade level is quite small.

To Reduce Inequality in School Attainment

Jencks et al. contend that schools have rather modest effects on the degree of cognitive inequality among adults. 1/ Most people find this argument difficult to accept. Highly educated people differ from uneducated people in many important ways, and most people assume that schools must cause many of these differences. In response, Jencks et al. argue that people who stay in school and attend college would differ from people who now drop out, even if they all had exactly the same amount of schooling. They argue, in other words, that schools, including universities, serve primarily as selection and certification agencies, whose job is to measure and label people, and only secondarily as socialization agencies, whose job is to change people. This contention implies that schools serve primarily to legitimate inequality, not to create it. In order to review policies to reduce inequality in educational attainment, we first need to examine the credentialing process.

Who gets educational credentials and why? Schools do not have to be certification agencies, nor does certification have to depend on time spent in school. Certification could be done strictly by examinations. Such a system exists to some extent in many European countries, where schools prepare students for national examinations, and the results of these examinations determine certification. A student can often take the examinations without having attended school. Conversely, merely attending school is no guarantee that he will pass the examinations.

A second reason for schools having become certification agencies is that by doing so they service the interests of a society that wants people sorted and graded, but does not know precisely what standards it wants to use. If high-school diplomas or other certificates of competence were given solely for passing examinations, there would have to be political agreement on what the examination should cover. This would be hard to get. Delegating the problem to the schools is a way of sweeping it under the rug.

A third advantage of relying on schools to certify students is that employers are at least as interested in whether their workers behave properly and do what they are told as in the workers' cognitive skills. Consequently, employers need a certification system that includes some direct observation of an individual "at work." Schools can provide this with their system of grading and teacher evaluation. Examination boards cannot.

1/ This section is an edited version of Jencks et al., pp. 135-37, 158-60.

We now turn to examining the effects of turning certification over to the schools. What kinds of people do schools certify, and what kinds of people do they fail to certify? This is obviously a complex question. Schools and colleges in the United States issue an enormous variety of diplomas and degrees. In order to get an overall picture of the relationship between educational certification and other kinds of inequality, we will have to simplify the question by assuming that the value of any given credential depends solely on how long it takes to acquire. We will assume, in other words, that the value of an individual's credentials is proportional to the highest grade of school he has completed. We will call this "educational attainment."

This approach has at least two limitations. First, it treats each extra year of school or college as if it were excatly as valuable as the next. This is clearly an oversimplification. An extra year of college increases a man's earning power more than an extra year of high school. Completing the last year of either high school or college also brings more economic benefits than completing any of the three preceding years. Nonetheless, the differences in year value are relatively modest, and if we go ahead and assume that the effects of educational attainment are linear--i.e., that one year of school or college is just as valuable as the next--we reduce our ability to explain adult economic success by only 3 to 5 per cent and at the same time greatly simplify the analysis.

A second major limitation of the educational attainment approach is that it makes no qualitative distinctions between types of certification or types of institutions. A master's degree in Engineering is harder to get than a master's degree in Education, and it is more valuable economically. Similarly, a B.A. in English from an Ivy League college is harder to get than a B.A. in English from a state college, and the economic benefits are greater. Once again, however, the effect of these distinctions among types of B.A.'s and higher degrees is rather modest. 1/

Why do some people end up with more impressive educational credentials than others? Jencks et al. (1972) suggest that the most important determinant of education attainment is family background. The impact of family background is accounted for partly by measurable economic differences between families, and partly by more elusive non-economic differences, like parental behavior. The other important determinants of educational attaiment are both cognitive and affective skills. The precise effect of cognitive skills is hard to determine, however, since we do not know to what extent test scores are a proxy for unmeasured, non-cognitive differences between home environments.

Qualitative differences between high schools seem to explain about 2 percent of the variation in students' educational attainment. Unfortunately,

---

1/ See Jencks et al (1972) footnote two, pp. 160-61.

we cannot say what qualities of a high school boost its college entrance rates and what qualities lower them. School resources do not appear to influence students' educational attainments at all. Attending high school with bright, highly motivated classmates seems to have both positive and negative effects on a student's chances of attending college. The curriculum to which a student is assigned is the one measurable factor that influences attainment, and it explains differences within rather than between schools.

Policy Implications. The answer is not that "nothing matters." It is true that making high schools more alike would have a negligible effect on the distribution of credentials. But if colleges altered their criteria for admitting students and awarding degrees, some groups would gain and some would lose. This is why there is now so much political pressure to alter admissions procedures in various ways.

Suppose, for example, that America adopted a certification system based entirely on standardized tests of the kind now used in college admissions. Such a system would benefit white working-class children significantly because they do better on the tests than they do in their grades. But while such a system would benefit poor whites, it would not benefit blacks. Blacks now seem to get slightly more education than whites with comparable test scores. They would lose this advantage if test scores were the sole basis for awarding diplomas and degrees.

A system in which credentials were distributed entirely on the basis of grades, and in which standardized tests played no part, would improve the position of working-class students and reduce the advantage of the middle classes even more than a system based entirely on test scores. This may be one reason why admission to public colleges has traditionally depended largely on high-school grades, while private colleges have ususally weighted grades and test scores about equally.

A system in which all students end up with as much education as they think they can stand would have less predictable effects. Suppose all education were free and no institution had admission or graduation requirements. If we judge by the amount of schooling people now say they would like, the relative advantage of middle-class over working-class students would not decline at all. A system that gave everyone as much education as he wanted would, however, slightly reduce the correlation between educational attainment and cognitive skill. If aspirations remained unchanged, then, and credentials remained equally valuable, such a system would create an elite, no less hereditary but slightly less clever, than the present elite. If aspirations did change, the consequences might be more appealing.

Without some evidence about the way in which society would operate if credentials were distributed on a different basis, we cannot really choose among these alternatives. If cognitive skills are important for on-the-job success, for example, a credentialing system that ignores cognitive skills will not work. If job competence depends mainly on non-cognitive traits, then a drastic change in the present system may be more feasible.

V. International Comparisons

While there have been some references to studies in other countries, the preceding sections have concentrated on describing the results from studies made in the United States. Now we turn to a comparative study using the input-output method that includes the United States as well as other developed and developing countries.

Some readers may well ask why all the preamble about the United States when the studies in this section seem to be the relevant data for consideration by policy-makers in developing countries. In contrast to the United States where similar studies have been replicated a number of times, there are few such studies in developing countries. In social, as in physical science research, the replication of studies and experiments is important for verification of the results. As we found above in discussing the United States, some studies reached quite different conclusions than the majority of studies. The consistency which appears across countries in the comparative study about to be discussed is evidence to suggest that further studies will tend to support rather than refute the findings.

With the successful termination of the International Study on Mantematics Achievement by Thorsten Husen in 1965, planning began for a study of achievement in other subjects. After seven years of effort and $5 million of expenditures, researchers in twenth-three countries have now produced an unusual set of research findings. They are based on data for 258,000 students, 5,900 teachers, and 9,700 schools. The work, orgnized by the International Association for the Evaluation of Education and Achievement (IEA), was reviewed by a conference of international experts at the Harvard School of Education in November 1973. Preliminary findings were presented in 1972 to the staff of the International Bank for Reconstruction and Development.

Because of the potential interest of the Bank policy-makers in the results, we depart from the method used in the section describing the U.S. data. There, we limited our discussion to a summary of the results because there were a large number of similar studies. Below, we will present the findings in more detailed form since they are often the only data on the country. Given the amount of data available, the results can still be considered only a summary.

The conclusions parallel the results of the Coleman study in the U.S., the Plowden Report in the U.K., and other evidence reported in Section

III. The text below is extracted from a paper by Neville Postlethwaite, a director of the study. 1/

Data and Methods

The data for the IEA study cover three populations in each of the countries participating in the study.

The definitions of the three populations are as follows:

Population I--all students in full-time schooling aged 10:00-10:11 at the time of testing.

Population II--all students in full-time schooling aged 14:00-14:11. (There is no population III.)

Population IV--all students in the terminal year in full-time secondary education programs that were either pre-university programs or programs of the same length. (The interpretation of this definition of this group varied as well as the percentage of an age-group in this target population.)

The countries are listed below. Some of them did not participate in all the tests. 2/ Note that there are four developing countries: Chile, India, Iran, and Israel.

Australia	Ireland
Belgium (Flemish)	Israel
Belgium (French)	Italy
Chile	Japan
England	Netherlands
Germany (FRG)	New Zealand
Finland	Poland
France	Rumania
Hungary	Scotland
India	Sweden
Iran	Thailand
	U.S.A.

1/ T. Neville Postlethwaite, "A Selection from the Overall Findings of the IEA Study in Science, Reading Comprehension, Literature, French as a Foreign Language, English as a Foreign Language, and Civil Education," Conference on Educational Achievement, Harvard University, November 1973. The edited extracts continue to the section "Why do some students read better than others?"

2/ See Annex Table 1, p. 70, for a full description.

The above definitions held true for Science, Reading Comprehension, Literature, and Civic Education. In the case of English and French as foreign languages, a further condition was that the students should be currently studying the language and have studied it for at least two years. Indeed, when examining the home background distributions of the students taking English and French, it was clear that where English was part of the curriculum it was studied by more or less the whole of the students in full-time schooling, whereas French tended to be restricted to smaller segments of the population.

Probability samples of schools and students within schools were drawn for each level for each subject (or group of subjects) within each country. The way in which the complex samples were drawn, together with the resultant standard errors of sampling and design effects for selected variables, has been shwon in detail elsewhere (Peaker, 1974).

Three to five years of development work were required to construct the outcome (i.e., performance) measures in each subject area at each level. The resultant tests had good reliabilities and were judged by the subject matter panels in the various countries to be appropriate for testing what was, in general, meant to have been learned in school up to that point of testing. The test formats used included multiple choice, open-ended, and file-in formats. For foreign languages, tape recorders were used for Listening Comprehension and Speaking. Both mark-sense cards and punched cards were used for data recording.

A series of questions was also given to the students, teachers, and school principals in questionnaire booklets. The variables (representing inputs to, and processes in, the school) for any one subject at any one level numbered approximately 200 to 500. The sheer amount of information collected (over 150 million pieces) makes this study one of the largest ever undertaken in the field of Social Science. As an example, Annex Table 2 lists the student, teacher, and school variables associated with French. These are the variables used in every country. Each country, however, had the opportunity to ask further questions that were of specific interest for its own system. Hence, although this review will present the findings of the international analyses, each national panel will be publishing a national report using a fuller set of data. Three examples of national reports are those of Australia (Rosier, 1973), Japan (Shimaoda et al., 1973), and Sweden (Husen et al., 1973). A recent issue of the Comparative Education Review contains a series of articles from most National Centers, each concerning a special national analysis undertaken on a current national problem in education (Postlethwaite, 1974).

The aims of the research were to identify those factors accounting for differences between countries, between schools, and between students. The technique used was a cross-sectional survey, albeit at three different levels. The survey was, therefore, a description of education as it is and not as it might be. Furthermore, there were no direct measures of past events; surrogate measures had to be used. A logitudinal study would have produced such measures, but for various reasons the IEA felt that the time and money required for such a study were beyond its reach. The input and process variables data were collected from students and school personnel. The

limitations of questionnaire procedures are that the responses, on the one hand, may not be accurate and, on the other hand, may provide only an incomplete and often indirect indication of the phenomenon in which one is interested. These constraints are important to note and bear in mind when interpreting the results.

Before passing to the results, it should be mentioned that the fact that this study has been conducted in twenty-three nations with fifteen different languages of instruction is in itself something of a tour de force. In spite of translation problems and the difficulties of planning and executing such a project, not to mention the research skills required by so many people in so many countries, it has been possible to measure outcomes in a reliable way and to make meaningful comparisons. There was a certain variation in the quality of the research competencies possessed by the participating centers, and the standard form of research proved to be very effective in raising the standards to the highest common factor. It also allowed each nation's educational system to be seen in an international context.

Findings

1. Between Countries

With the limited number of observations of country mean scores it is not possible to undertake between-country multivariate analysis. Differences in means and distributions on single variables are as interesting, however, as bivariate relationships. IEA has always stressed that it is not conducting a "cognitive olympics" and that great care should be exercised when comparing country mean scores. First, although international tests have been constructed and each country was happy to use them, they were not 100 percent valid for each country and some countries may well stress some objectives that were not tested by the tests. Secondly, the variation from country to country in the proportion of an age group in school, even at the same population level, makes comparison difficult. Keeping these cautions in mind, however, some of the test scores between countries are of interest. In what follows, four different modes of presentation have been used. For Reading Comprehension, the pass-fail rate on particular items is given. For Literature, a profile of four different aspects of literature is given. In Science, the increment in General Science performance from one population level to another is presented. For the other subjects, means and standard deviations are presented.

(a) Reading. Perhaps the most dramatic finding is the very large difference in performance between developed and developing nations. If we take Reading, we find that differences among developed countries are fairly modest. With complete consistency, however, the four are listed on P. 42 developing countries fall far below those that have a relatively high level of economic development and a long standing tradition of universal education. The differences are so large that by the standards of the developed countries, fourteen-year-olds in the developing countries seem almost illiterate.

To exemplify this somewhat, consider one selected passage from the battery of reading tests for ten-year-olds:

Population I

Passage and Items

One of the most interesting birds I have seen is the Indian Tailor Bird. It is a small olive green bird that does not look at all unusual, yet it has a most unusual way of making its nest. The birds work together in pairs. First they find a leaf, the right size, and make holes along the edges with their beaks. Through these holes they thread grass. One bird pushes the thread from the outside, while the other bird sits in the nest and pushes it back until the edges of the leaf are sewn together to make a kind of bag, still hanging on the tree, in which the Tailor Bird lays its eggs.

1. What does the Tailor Bird use in place of thread?

   A. Grass
   B. String
   C. Spider web
   D. Thorns

2. The Tailor Birds are interesting because they

   A. are small and olive green in color
   B. live in pairs
   C. make their nests in a special way
   D. fly very fast

3. The Tailor Bird got the name because it

   A. is a small bird
   B. looks unusual
   C. can sew
   D. has a beak shaped like a needle

4. The Tailor Birds make their nests

   A. from leaves
   B. in a hole in a tree
   C. in the tall grass
   D. with a lining of grass

5. The person who wrote about Tailor Birds was trying to

   A. give you some new information
   B. tell you a story
   C. get you to share his feelings
   D. keep you guessing on how the story will come out.

Source: Postlethwaite, 1973

Clearly, the paragraph shown here is at a level that presents difficulty to a substantial fraction of ten-year-olds in every country. Considering all 14 countries and all five items (test questions), the typical failure percentage runs between 35% and 40%. The median percentage goes as low as 26% in Finland, but is 48% in Chile, 58% in India, and 65% in Iran. When one considers that random marking would be expected to give only 75% of error on these four-choice items, it becomes clear that even this passage is pushing the limit of competence of most ten-year-olds in these three countries.

Similar results were obtained for Populations II and IV. In one passage for Population II, the fourteen-year-olds, the median error rate was approximately 30% but in Chile, India, and Iran the rate was 47%, 63%, and 61%, respectively. For Population IV for one passage the median error rate was 25%, but for the above-mentioned developing countries it was 44%, 66%, and 68%, respectively (Thorndike, 1973, pp. 133-39).

A further indication of the base level of competence in the developing countries is provided by error rate on the first nine items of the test of reading speed. These are items such as:

Peter has a little dog. The dog is black with a white spot on his back and one white leg. The color of Peter's dog is mostly

black brown grey.

The percentage of wrong answers on such items are shown in Table V.1 is as follows:

Table V.1

Reading Test Errors

Country	Ten-Year-Olds	Fourteen-Year-Olds
Belgium (Fl.)	8	3
Belgium (Fr.)	9	2
Chile	26	16
England	9	4
Finland	11	8
Hungary	20	8
India	36	33
Iran	52	20
Israel	17	9
Italy	11	9
Netherlands	8	4
New Zealand	(not taken)	4
Scotland	10	3
Sweden	7	2
United States	11	4

In the European countries, a typical error rate on these items is about 10% for ten-year-olds and 4% for fourteen-year-olds. With these values, one must contrast percentages of 26,36, and 52 for the ten-year-olds of the developing countries (Chile, India, Iran) and 16%, 33% and 20% for the fourteen-year-olds. Admittedly, this material was given as a speed test. But it was also given as a reading test! If a substantial proportion of the students in a school system have real difficulty in reading these materials, one must question whether any more than a minimal level of literacy has been achieved in that school system.

More advanced tests were designed for students in the final year of secondary school prior to entering the university. Diagram V.1, below, measures the results from two perspectives. First, the mean of the scores of all students within one country was estimated, and then the country means' ranked. The results show the real relative weakness in scores of the developing countries compared to the developed. Second, the mean scores of the top 9% of the students in each country were then compared. The developing countries do not rank any better, but there is more dispersion among the scores of the developed countries. In the United States, where mass education is most extensive and automatic promotion from one year to the next virtually universal, the top 9% did better than their peers in the other 14 countries. Opening education to the mass of the population did not seem to affect the quality of the best students. Selective systems appear to pay a high price in lost talent and social dislocation. 1/

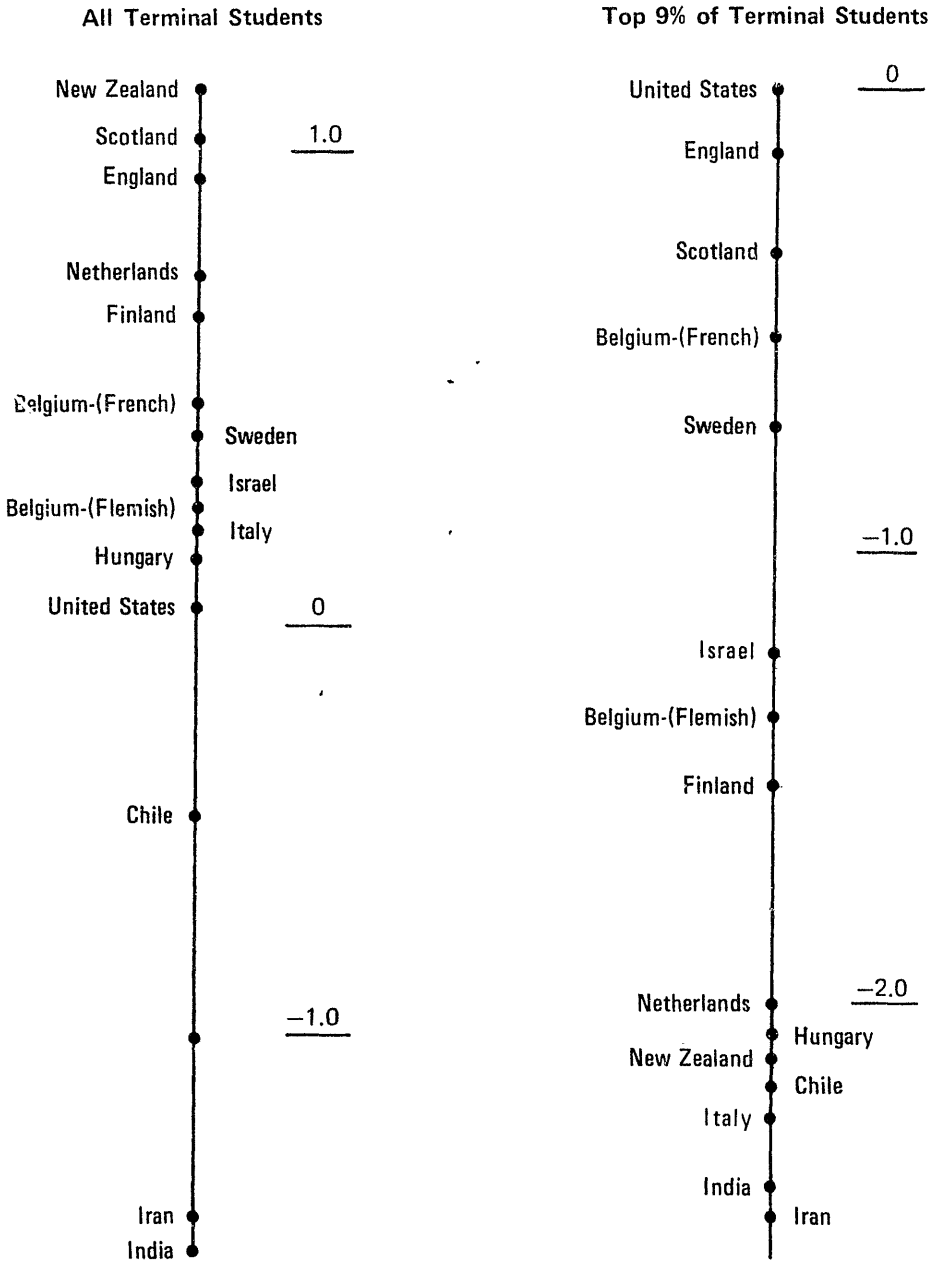
(b) Science

Because there were anchor items among the tests for the various populations, it was possible to bring the population test scores onto a common scale and to examine the increment from one population level to another in Science as measured by the IEA tests. Diagram V.2, below, presents the increases. Again, the developing countries are presented separately: note the different position of the zero point. The "retentivity" figure given next to each country's name is an estimate of the percentage of an age group still in school in the pre-university year grade. Note that Australia and New Zealand did not test ten-year-olds; nor did Japan, since it only tested one population--namely, "la classe terminale"--and hence no increment measure was available. The left-hand side of the striped bar represents the scaled mean score at age ten, the right-hand side the scaled mean score at age fourteen, and the right-hand side of the block bar represents the scaled mean score of the pre-university year group. For example, the mean science score for Iran's population IV

1/ For example, West Germany, which screens out "non-academic" children at the age of ten, shows the highest index of social bias. This means that the top-ranking students come from highest income groups. At age eighteen only 1% of children from lower income homes were found in school, compared to 14% in the United States.

Diagram 1

CROSS-SECTIONAL COMPARISON OF READING COMPREHENSION (RANKING OF READING ABILITY OF STUDENTS IN THE TERMINAL YEAR OF SECONDARY SCHOOL)

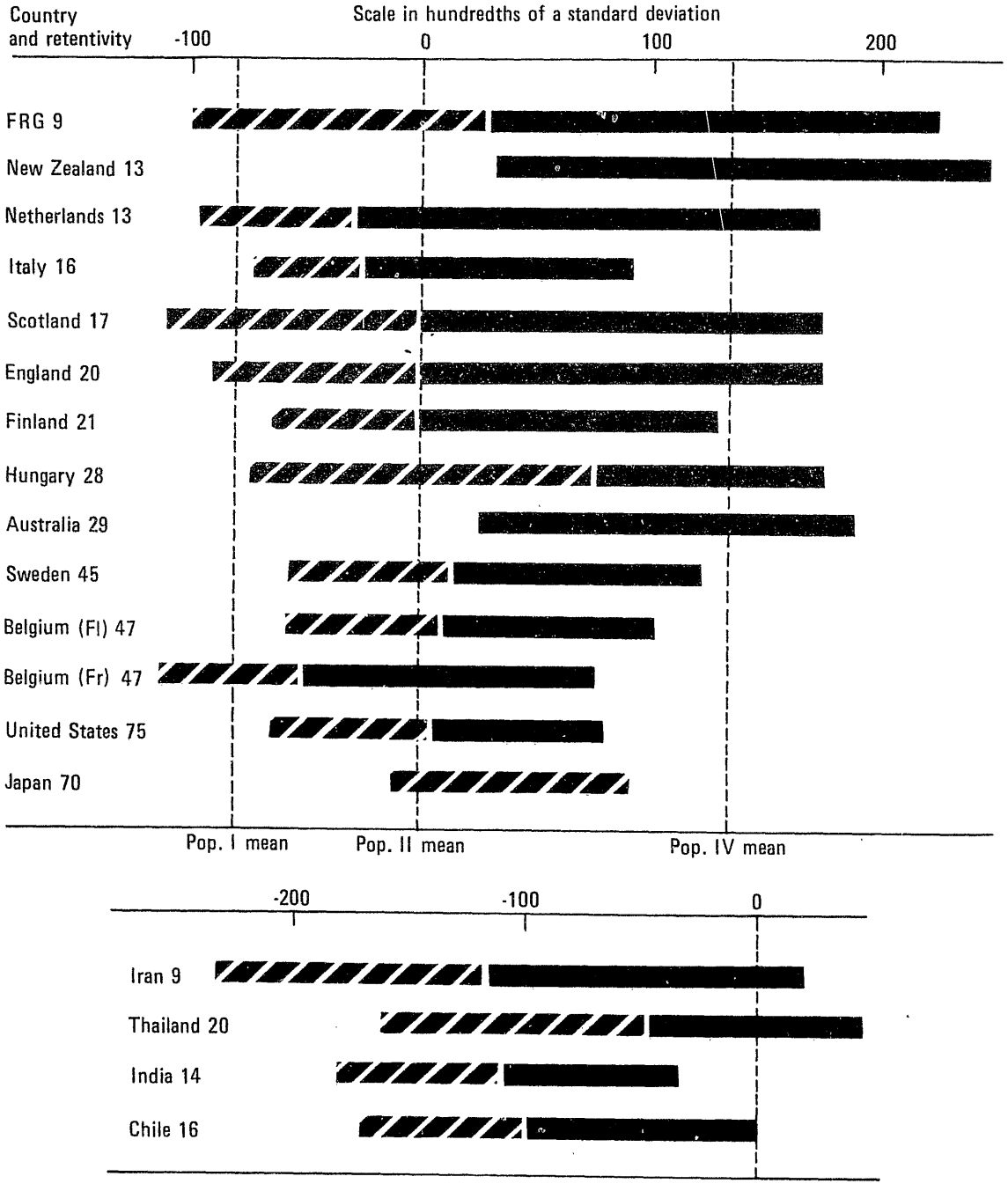


Source: Based on data from Thorndike, 1973.

World Bank—8611

Diagram 2

INCREASE IN LEVEL OF PERFORMANCE IN SCIENCE FROM THE 10-YEAR-OLD LEVEL TO THE TERMINAL SECONDARY SCHOOL STAGE



Source: Postlethwaite, 1973

World Bank—8612

is lower than the score of Australia's population II. The length of the bar represents the amount learned between the different populations.

In the Science report, the authors have broken down the total Science score into a series of sub-scores: Earth Science, Physics, Chemistry, Biology, and Practical Science, on the one hand; and on the other, a classification of educational objectives such as functional knowledge, comprehension of a scientific principle, application of knowledge for problem-solving and analysis, synthesis, and evaluation (Bloom, 1956). The country-by-country profiles of such sub-scores help to identify the relative strengths and weaknesses of the students' performance. The item-by-item analyses of the Science items show not only the percentage of the population answering the item correctly (which can in certain sense be construed as a criterion-referenced test) but also the types of wrong answers typically produced. Both the profiles and item analyses are extremely useful feedbacks to the curriculum developers as a check to what extent performance matches the global and detailed objectives set for Science education in a particular country.

Each participating research center will be carrying out a set of special national analyses. Where different curricula exist for different school types or regions within a country, such profile and item analyses will be important as feedback to the curriculum developers. In a centralized system of education the feedback links are, of course, easier to forge or maintain than in a decentralized system, but it is hoped that all research centers will create the necessary links, however complex this may be.

Experience suggests that boys show a greater "interest in Science" than girls and perform better on Science tests. These impressions receive powerful support from the IEA study. These differences exist across all countries and the differences increase as the students grow older. Thus, at ten years of age, boys performed on average about one quarter of a standard deviation higher than girls, half a standard deviation at fourteen years of age, and three quarters of a standard deviation higher in the terminal year of secondary education. Boys are more strongly attracted to the Physical Sciences and girls to the Biological Sciences. The score differences in Biology were less than in the other branches. This discrepancy in Science performance is clearly a problem that deserves earnest attention in the near future in all countries.

(c) English as a Foreign Language

Table V.2, below, presents the means, standard deviations, and numbers of students for the participating countries in the four major skills of Reading Comprehension, Listening Comprehension, Writing, and Speaking. Blanks denote non-participation. The large standard deviations for Population II Reading are noteworthy as are the differences in the Population IV mean reading scores.

(d) Civic Education

There were many outcomes of Civic Education including attitude scores in addition to the Cognitive Test Total and subscores. In Table V.3,

TABLE V.2

Means, Standard Deviations and Numbers of Students for Scores in Four Major Skills of English

(x is mean score, G is Standard Deviation, and N is number of Observations)

	Reading			Listening			Writing			Speaking		
	$\bar{X}$	G	N	$\bar{X}$	G	N	$\bar{X}$	G	N	$\bar{X}$	G	N
Pop II												
Belgium (Fr)	22.8	15.2	687	15.4	8.4	671	12.5	6.8	697	20.9	7.4	162
F.R.G.	29.9	16.2	1074	-	-	-	-	-	-	-	-	-
Finland	16.7	17.8	2100	12.1	9.9	2118	9.7	8.4	2087	19.8	9.7	200
Israel	26.3	18.7	1065	-	-	-	-	-	-	-	-	-
Italy	18.6	17.3	791	9.1	6.0	694	11.3	7.0	731	19.8	7.7	71
Netherlands	37.4	16.4	2090	-	-	-	-	-	-	-	-	-
Thailand	18.4	13.4	1951	-	-	-	-	-	-	-	-	-
Pop IV												
Belgium (Fr)	24.2	10.8	1440	16.9	7.0	1394	20.5	6.0	1391	24.2	6.9	192
Chile	10.6	9.8	2118	7.1	4.5	231	7.2	4.8	183	14.4	6.5	94
F.R.G.	41.3	6.1	1374	-	-	-	-	-	-	-	-	-
Finland	85.6	8.3	2310	23.2	6.1	2296	26.6	3.2	2238	29.9	8.7	279
Hungary	20.1	11.2	1063	16.3	7.3	1063	18.2	7.2	1063	30.3	12.3	119
Israel	27.4	12.9	604	-	-	-	-	-	-	-	-	-
Italy	20.2	12.6	324	14.7	7.6	314	18.9	6.0	322	27.0	7.2	28
Netherlands	42.9	6.5	1561	-	-	-	-	-	-	-	-	-
Sweden	39.5	9.0	1626	26.2	5.4	1602	25.3	4.6	1544	34.0	8.8	197
Thailand	19.7	9.1	937	-	-	-	-	-	-	-	-	-

TABLE V.3

Means, Standard Deviations and Numbers of Students for Scores in Three Outcomes in Civic Education

	Total Cognitive score			Anti-authoritarian (attitude)		
	$\bar{X}$	G	N	$\bar{X}$	G	N
Pop. I						
F.R.G.	13.9	8.5	1070	3.4	.8	1051
Israel	-	-	-	3.3	.6	402
Italy	18.6	11.1	2390	3.5	.6	2325
Netherlands	15.6	7.6	1746	3.4	.6	1730
Pop. II						
F.R.G.	26.0	8.7	1313	4.0	.5	1275
Finland	24.5	9.8	2370	3.9	.6	2356
Iran	9.7	5.9	2204	3.2	.5	2032
Ireland	20.8	10.4	834	3.9	.6	817
Israel	25.6	9.6	1039	3.7	.6	952
Italy	22.9	9.4	930	3.9	.6	918
Netherlands	27.3	7.9	1685	4.1	.5	1645
New Zealand	24.3	9.6	1983	4.1	.5	1969
U.S.A.	24.7	9.9	3186	4.0	.5	3119
Pop. IV						
F.R.G.	28.2	5.8	1163	4.6	.4	1176
Finland	26.1	6.6	2315	4.3	.4	2282
Iran	6.8	4.9	2159	4.3	.5	2028
Ireland	16.9	8.4	786	4.2	.4	782
Netherlands	25.5	6.5	1206	4.3	.5	1203
New Zealand	28.4	7.1	1665	4.3	.4	1668
Sweden	27.0	7.8	1723	4.5	.5	1636
U.S.A.	21.4	9.7	3016	4.1	.5	2928

below, means for Civics Cognitive Test Total are compared for different countries within population. (Comparisons between age population groups are not justified because the cognitive tests were composed of different items for each population.) The phenomenon noted in the other subjects--namely, considerably lower scores for less developed nations--holds true in Civics for Iran. The other countries' mean scores are fairly closely clustered; it is interesting to note, however, that at both the Population II and Population IV levels, students in Ireland achieve relatively low mean scores.

The anti-authoritarianism scale, which is derived from a series of attitude items concerning democratic values that are administered in the same form at each age level, can be compared across population. Anti-authoritarianism is a good representation of the first major attitudinal factor derived in factor analysis on the fourteen-year-old (population II) students and was used as a criterion in the regression analysis. Although the between-country and between-population differences are interesting for speculation, of more interest are patterns of attitude scale difference which are too complex to be examined here. for example, in some countries high anti-authoritarianism is combined with a low sense of citizen efficacy; in other countries both types of attitudes are low.

(e) Other Between-Country Differences

It is not possible to carry out a multivariate analysis between countries (with so few degrees of freedom) in order to identify the factors associated with differences between national mean scores. Some differences in certain factors are striking, however, and provide interesting leads. Two examples are given below:

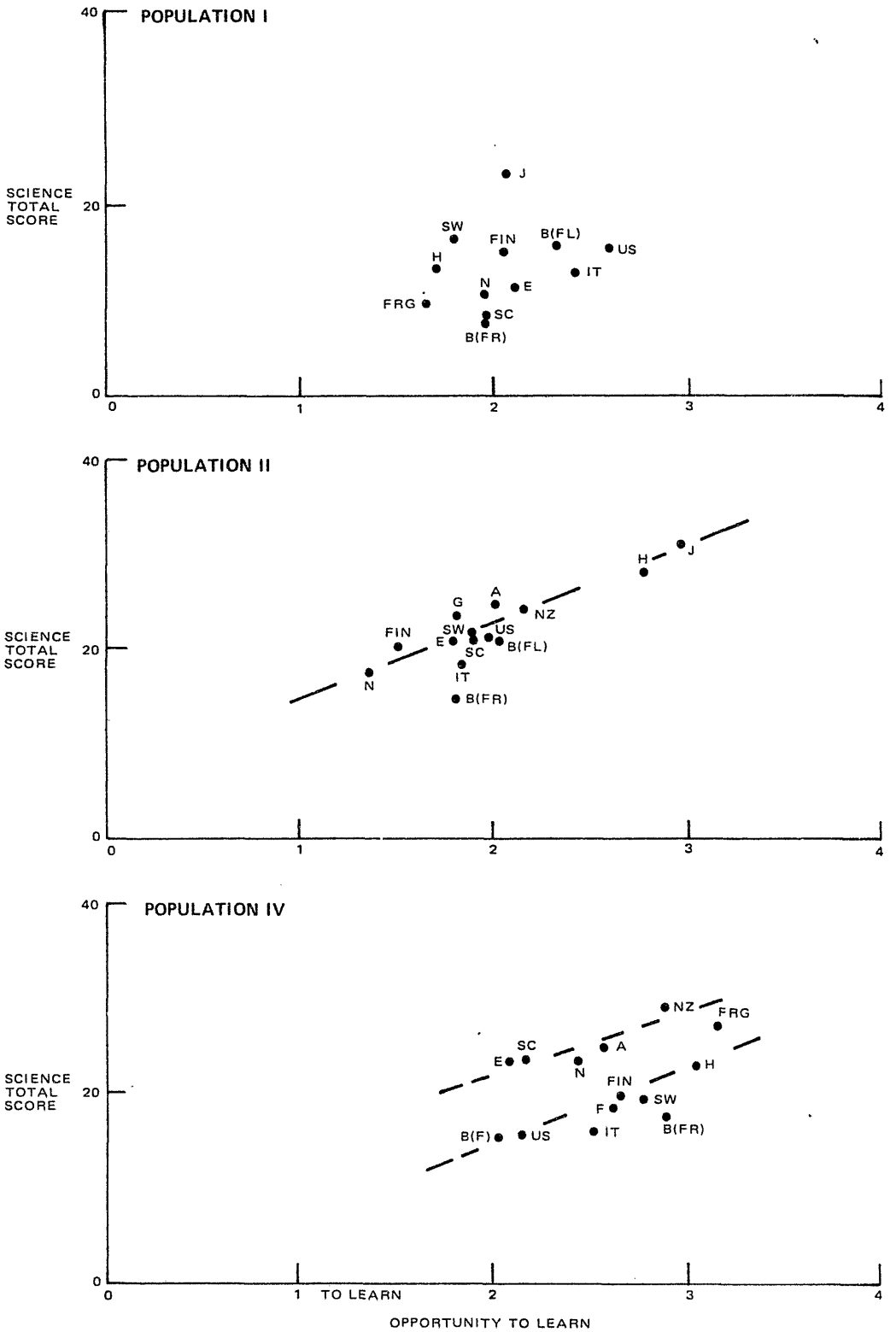
(i) Opportunity to Learn in Science. In each school in which Science was tested, the Science teachers in that school were asked collectively to rate each item in the Science tests as to the appropriate percentage of students in the target population who had had "the opportunity to learn" the substance tested in the item. The results are in Diagram V.3, below. The scale was:

A.	All students	4
B.	More than 75% of the students	3
C.	Between 25% and 75% of the students	2
D.	Less than 25% of the students	1
E.	None of the students	0

Despite the crudity of the measure, it shows a considerable difference between countries when summed to a national level. Diagram V.3 presents graphs of the opportunity to learn against Science total test scores for Populations I, II, and IV. While the relationships are striking, one cannot help wondering if, for example, in Population II, the tests are more appropriate for Japan than for the Netherlands and if the Netherlands' students are learning other things in Science not measured by the Science tests. Such questions are for the curriculum-developers to decide, but it should be pointed out that the

Diagram 3

GRAPHS OF OPPORTUNITY TO LEARN AGAINST SCIENCE TOTAL TEST SCORES FOR POPULATIONS I, II AND IV



World Bank 8613

test construction was a lengthy and detailed expercise involving Science educators in all countries. All countries voiced agreement that the tests were a reasonable general measure of Science as taught at the particular level. The lack of relationship for Population I may be attributed to the unstructured curriculum in Science in most countries at that stage of schooling.

(ii) Social Bias. There has been much discussion recently concerning the equality of participation, in the whole school system, of students from all social classes. Each research center utilized its own social class set of categories. (This design differed from the IEA mathematics study 1962-1967, where one international set of categories was used.) Although, consequently, the categories are not exactly the same from country to country and it is therefore difficult to make comparisons between countries, comparison within countries is possible. Table V.4, below, presents the percentage of students for each population in the fathers' occupational categories in seven countries.

Of interest is the progression from population. Furthermore, the difference between England, the Federal Republic of Germany, and the Netherlands, on the one hand, and Finland, Hungary, Sweden, and the United States of America, on the other hand, is striking. It should be recalled that the Federal Republic of Germany tested only students in the Oberprimaria in the Gymnasium.

(iii) Does More Mean Worse? It is often argued that "more means worse": the higher the proportion of an age-group allowed into the final year of schooling the lower will be not only the average achievement but the achievement of the elite students as well.

Taking Science as an example, Diagram V.4 presents the average score of those in full-time schooling in the pre-university grade (on the bottom line), the average of the top 9% of an age-group since the Federal Republic of Germany only has 9% of an age-group in the Gymnasium (on the line second from the bottom), the average of the top 5% (on the second line from the top), and the top 1% (on the top line). (The vertical axis gives the score; the figure under the bottom line for each country represents the estimated percentage of an age-group in school, thus, Sweden with 45% of an age-group in school has a higher average of that group than France with only 29% of an age-group in school.) The comparisons between countries are fascinating, but it is clear that the differences in score shown by the top 1%, or 5%, or 9% are not related to the differences in percentages retained in school: "more does not mean worse."

(iv) Single Variables. Means, standard deviations, and frequency distributions were produced for all questionnaire items (student, teacher, and school questionnaires) and for every population in every country. Such extensive information is typically not collected by statistics divisions of ministries of education. For each variable there is a known standard error of sampling. The data collected on the hundreds of variables used in the IEA survey constitute a mine of information not only for national ministries of education but also for international agencies such as UNESCO and OECD.

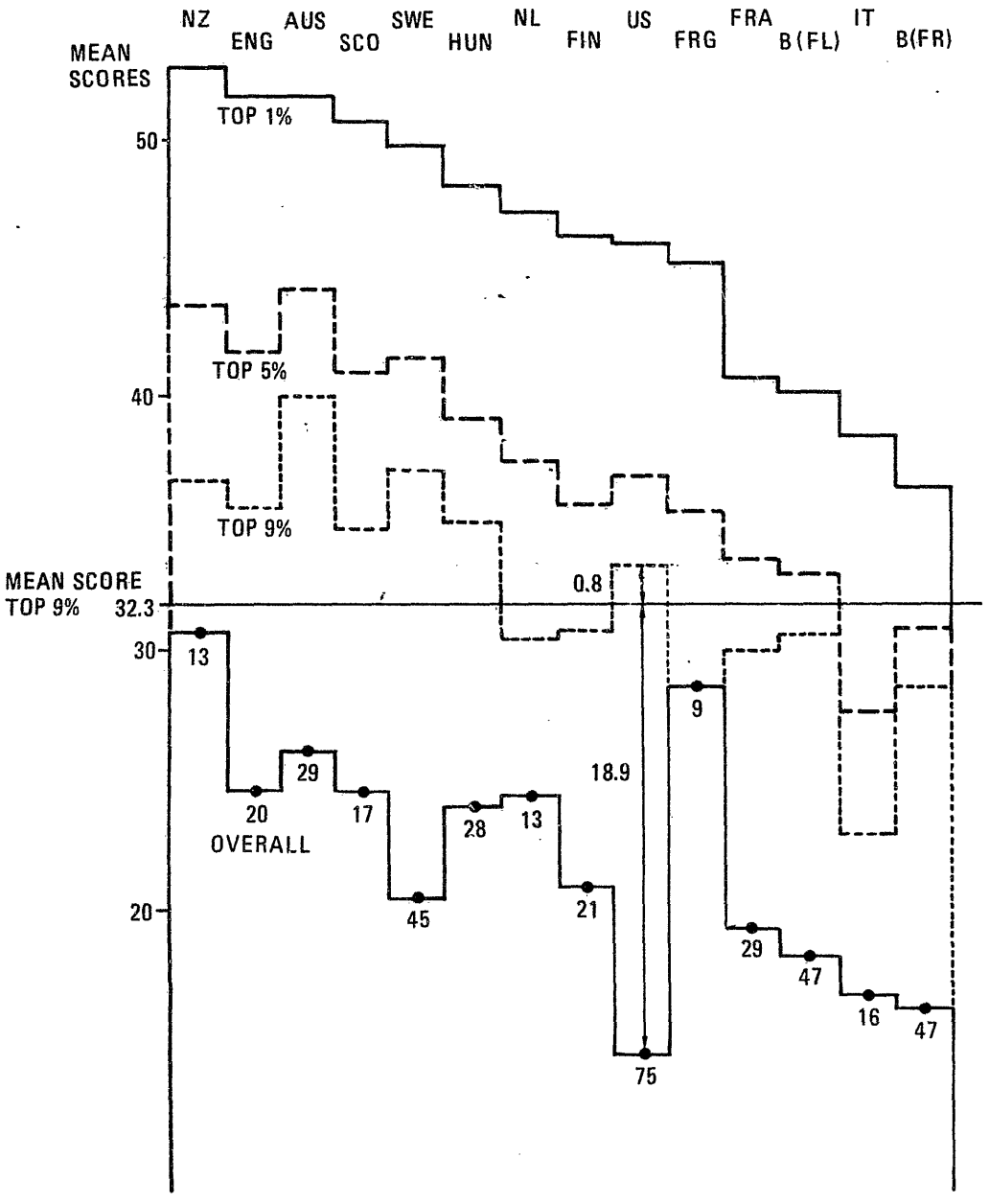
TABLE V.4

Percentage of Students for each Population in Selected Categories of Father's Occupation

(Science, Reading Comprehension and Literature)

	POPULATION I		POPULATION II		POPULATION IV	
	Professional & Managerial	Unskilled & semi-skilled	Professional & Managerial	Unskilled & semi-skilled	Professional & Managerial	Unskilled & semi-skilled
England	16	21	14	14	38	5
Fed. Rep. of Germany	13	7	14	8	49	1
Finland	9	35	10	34	20	15
Hungary	15	43	20	36	38	18
Netherlands	26	12	20	12	55	5
Sweden	23	31	26	27	35	15
U.S.A.	24	18	31	16	34	14

SCIENCE MEAN SCORES OF TOP 1%, TOP 5%, TOP 9% OF AN AGE–GROUP AND OVERALL GROUP



World Bank 8614

Furthermore, a comparison of the national means of a particular variable is of interest. As an example, Table V.5, below, presents the means of the student-teacher ratios for each of the populations tested.

2. Within Countries

Within countries the student score in the various subject-matters was used as the school outcome (dependent variable) and the input and process factors as the independent variables in a regression analysis; the purpose of the study was to identify the importance of groups of variables in accounting for variance among students in their performance. Certain groups of variables were formed. The details of how these groups were formed, both through the sieving and compositing of variables, is reported in detail in the various reports and particularly in the technical volume (Peaker, 1974). For each subject area there were, as has been mentioned above, some 500 independent variables describing the students' home background, their previous schooling, and current schooling, the teachers and their teaching, the school principal, the organization of the school, its size and budget, etc.

Clustering Variables

The major groups (or blocks) of variables used were:

1. (a) Home Background. A weighted composite of father's education, mother's education, father's occupation, number of books in the home, use of dictionary in the home, and size of family. (There were minor deviations in this composite between subject areas; and in one country, Sweden, the first two variables had to be omitted because of ambiguity in the Swedish translation of the questions.)

(b) Sex, Age (age was entered in Block 3 for French and English).

2. Type of School and/or Program in which the student was then enrolled. (The elements of each variable were criterion scaled against Reading, Word Knowledge, or Science in order to form the variable type of school and/or type of program. This variable was intended as a surrogate--perhaps a weak one--for previous schooling.)

3. (a) School and Teacher Variables (for Science, Reading Comprehension, Literature, and Civics).

(b) Time Variables, e.g., years studied subject, current grade, grade beginning study of subject, and age (for French and English).

4. Kindred Variables, e.g., expected occupation, education, attitudes, etc. (for Science, Reading Comprehension, Literature, and Civics).

5. School and Teacher Variables (for French and English).

(In French and English the kindred variables formed a fifth block.)

TABLE V.5

Student-Teacher Ratios by Country and by Population

(Science, Reading Comprehension and Literature)

	Student-Teacher Ratio 1/ Population		
	I	II	IV
Australia	-	19	20
Belgium (Fl.)	22	15	14
Belgium (Fr.)	18	12	14
Chile	40	39	41
England	28	18	17
Federal Rep. of Germany	33	27	20
Finland	19	20	22
France	-	-	16
Hungary	20	20	17
India	36	27	26
Iran	46	74	72
Israel	20	15	14
Italy	20	16	17
Japan	27	22	-
Netherlands	31	18	17
New Zealand	-	22	22
Scotland	27	17	17
Sweden	17	14	15
Thailand	24	24	24
United States	26	20	20

1/ Data reported by schools. This ratio comprises the total number of students in a shool divided by the total number of full-time teachers (2 half-time equal one full-time, etc.) in that school. It does not include teacher aides or technical assistants.

Typically, two further groups of variables were used but not consistently from subject to subject. The Word Knowledge scores (on an antonym-synonym test) were used and the Reading scores where Reading was not the criterion.

Within countries two major types of analyses were undertaken: a between-student analysis and a between-school analysis. The between-school analysis was undertaken for Science, Reading Comprehension, and Literature only. In each case the sieving of independent variables (particularly the instructional, school organizational, and kindred variables) was based on the partial correlation, after home background and/or type of school, type of program had been partialled out, exceeding twice the standard error of sampling and the importance attached to a particular variable by the subject committee. In this way, the effect of school and teacher variables could be judged after the calibre of input of students to the school had been taken into account. Cases of detected high multi-collinearity were "solved" by compositing the variables in question.

Where the school and teacher variables were based on fewer than 50 observations (i.e., 50 schools), there is a problem of stability of estimates. The populations in this category are the two Belgiums (Populations II and IV), Iran (all Populations), and the Netherlands and Hungary (Population IV). It should also be pointed out that, where a variable had 20% or more missing data, it was dropped from the analysis. This was the case for all data relating to the school budget. In countries where these data are complete, however, further analyses should be run in order to examine the effectiveness of differential school-budget aspects after at least home background has been partialled out. Some variables (e.g., class size, laboratory class size) were omitted because they were curvilinear. The separate analyses of these were never undertaken because of temporal and financial constraints, and clearly such analyses have high priority.

Table V.6, below, presents as an example the surviving variables by block for New Zealand Science-Population II between-school analysis. There are forty single variables which with compositing form thirty variables. This was a typical number of variables for the final regression analyses.

TABLE V.6

Variables Surviving to Final Between-School Regression Analyses

in New Zealand Science--Population II

BLOCK 1	Father's occupation)	
	Father's education)	
	Mother's education)	
	Use of dictionary)	Home background
	Books in home)	
	Family size)	
	Age	
	Sex	
BLOCK 2	Type of school	
	Type of program	
BLOCK 3	Percent male teachers in school	
	No. of laboratory assistants	
	Sex of teacher	
	"Opportunity to learn"	
	School environment	
	Total homework per week)	Homework composite
	Total Science homework per week)	
	Taking Science)	Study of Science composite
	Total years study of Science)	
	Total hours study of Science)	
	Principal teaching experience	
	Total enrollment	
	Decision-making--syllabus	
	Admission criteria--residence	
	--exams	
	Economy of school region	
	Methods--audio-visual	
	--field trips	
	Science expeditions	
	In-service Biology course	
	Total in-service training	
	Grade in school	
	Science teacher training	
BLOCK 4	Interest in Science)	
	Science in work)	
	Importance of Maths)	Attitude composite
	Reading Science/technical	
	Reading Science fiction	
	Reading Science articles	
	Watching TV--Science related programs	

Incremental Variance

Annex Tables 3 to 6 present the incremental variance for various blocks of variables, as well as the total variance accounted for in each subject area in each population for the between-student analyses. Table V.7, below, presents a summary of the block variances and total variances account.

The percentage of total variance accounted for varies from a low of 11% in Iran for Population IV--Science to a high of 79% in Germany for Population II--English. The average percent total variance accounted for was 38.9%, indicating that there is a great deal of unaccounted variance that educators must work on identifying and measuring other factors not yet included in the IEA analyses. The total variance accounted for is notably lower in developing countries.

The home background variables vary from 1% to 30%, with an average of 11.5%, which raises the question of the extent to which the appropriate home variables are being tapped in some countries.

Learning conditions vary from a low to 1% to a high to 52%. The average is 10% across all populations in all subjects in all countries. The percentages are clearly lower for Reading Comprehension, Literature, and Civic Education than for Science, French,and English. Furthermore, the percentage in these last three subjects rises with the school population being examined. Let us be clear on what these percentages are. In a cross-sectional survey of this kind we are saying that after the home background of children has been taken into account the differences between schools (as they now exist) in the learning conditions they provide are associated to a considerable degree with differences in performance between students on the criteria under consideration. In some cases, the learning conditions are two-thirds of the total variance accounted for.

But then why should these differences between learning conditions be more important for Science, French, and English than for Reading Comprehension, Literature, and Civics? Is it that the first three subjects, which require learning specialized material for which there exists little knowledge in the home and therefore little backing, are necessarily more school-oriented? Or is it that Reading Comprehension is not systematically taught in schools once the decoding and encoding of the mechanics of reading have been accomplished?

Why Do Some Students Read Better than Others?

Although there are appreciable variations from country to country, 1/ certain general trends stand out. In the first place, it is clear that information about characteristics of the home and community environment in

---

1/ This section is an edited extract from Thorndike (1973), pp. 98-99.

TABLE V.7

Averages and ranges of all countries of contributions by selected blocks to total variance in the between-student analyses for each subject

Home background (Block 1)

	Pop. I		Pop. II		Pop IV	
	Average	Range	Average	Range	Average	Range
Science	11	1 - 27	16	3 - 29	13	3 - 25
Reading Comprehension	14	1 - 25	16	1 - 27	8	1 - 18
Literature	-	-	15	5 - 25	9	4 - 17
Civics (cog.)	-	-	16	9 - 20	10	5 - 18
French (reading)	-	-	15	3 - 15	4	1 - 16
English (reading)	-	-	14	1 - 26	7	1 - 16

Previous schooling (Block 2)

	Pop. I		Pop. II		Pop. IV	
	Average	Range	Average	Range	Average	Range
Science	1	0 - 14	6	0 - 17	6	0 - 26
Reading Comprehension	1	0 - 9	10	0 - 21	7	0 - 19
Literature	-	-	8	0 - 14	3	0 - 7
Civics	-	-	11	5 - 20	5	0 - 15
French	-	-	22	1 - 22	3	0 - 9
English	-	-	20	0 - 52	9	0 - 31

Learning conditions (Block 3 for Sc, R.C., Lit & Civics, Blocks 3 & 4 for Fr and Eng)

	Pop I		Pop II		Pop IV	
	Average	Range	Average	Range	Average	Range
Science	8	1 - 21	9	4 - 23	15	4 - 41
Reading comprehension	6	2 - 18	6	3 - 10	5	2 - 15
Literature	-	-	7	3 - 12	4	1 - 6
Civics	-	-	11	3 - 18	5	1 - 9
French 1/	-	-	17	12 - 24	19	10 - 52
English	-	-	16	8 - 27	17	4 - 29

Total variance explained 2/ (The total variance is always unity or in this case unity x 100)

	Pop I		Pop II		Pop IV	
	Average	Range	Average	Range	Average	Range
Science	27	14 - 36	36	17 - 55	39	11 - 63
Reading Comprehension	28	18 - 39	39	20 - 51	25	13 - 47
Literature	-	-	55	36 - 66	37	28 - 53
Civics	-	-	59	55 - 62	39	28 - 57
French*	-	-	45	28 - 65	43	27 - 61
English	-	-	60	47 - 79	44	22 - 75

1/ This is with Reading as the criterion. For Listening the percentages accounted for are greater.

2/ Both in this table and in Appendices 3 to 8, it will be noted that the sum of the percentages of the three blocks for any one subject do not total to the percentage in the "total variance" block. This is because the "total variance" in each subject is the sum of the first three blocks plus a kindred variables block and typically one or two other blocks which vary from subject to subject but which usually include a Word Knowledge score which might be regarded as a partial surrogate for "intelligance" and "previous experiences" not measurèd in Blocks 1 and 2.

which a youngster has grown up permits a fairly good prediction of his achievement in reading at age ten and at age fourteen. The prediction is less effective at the end of secondary education, but this can be understood in most countries as stemming from the fact that a good deal of selection has taken place by the end of secondary education on the basis of both of academic ability and economic level, so that those who remain in school are a select group, not representative of the total population of the country.

In the ten-year-old group, placement in type of school seems to be of no great importance as a predictor. With fourteen-year-olds in many countries, however, a very substantial part of the variation in reading achievement is represented by the type of school or type of program in which the individual is placed. It seems reasonable to think of these variables as representing primarily classification variables rather than treatment variables. That is, assignment to one or another type of school or program is likely to take place on the basis of previously demonstrated ability, so that the differences in achievement of students in the different categories of schools should probably be thought of primarily as an indication of the type of classification that has taken place and only secondarily as a result of the treatment that has been in effect in one school as compared to another. Among those in the final year of secondary education, type of school or program seems to be of less importance in many countries that it was for fourteen-year-old--again because the continuing programs are more uniform by the pre-university year, and because a high degree of selection has taken place.

The third general category of variables, and one in which the study was particularly interested, is composed of variables describing schools in terms of their organization, facilities, and treatment of students. The effects of these differences can be seen clearly only after allowance is made for differences in the input of students into different categories of schools. Hence, it becomes necessary to partial out the effects of the previous two kinds of variables; to wit, family and background variables, and the classification effects resulting from assigning students in terms of their previously demonstrated ability. After these effects have been taken care of, very little consistent pattern is found in the school variables associated with differences in achievement. Though certain variables do emerge in single countries, the same variable may appear with a reversed significance in some other country. See annex 2 for the school bariables that were not significant. The number of schools studied in each country is not large enough for one to be confident that the differences from country to country are really stable and meaningful. At this point, about the only conclusion that it seems possible to draw is that from among the kinds of school-treatment variables on which it was possible to obtain information in a large survey type of study, little emerges that is useful in understanding the progress of students toward higher levels of Reading Comprehension. The same conclusion seems to apply in Science.

Finally, there are a number of other current characteristics of students that tend to be associated with higher levels of reading ability.

These are characteristics that should probably be thought of as concomitants rather than as either causes or results. They are part of a syndrome or constellation of characteristics that describe the better reader, on the one hand, as compared with the poorer, on the other. Thus, the better reader reads more, reads more in such areas as current events, science, philosophy, and even humor, and reads less in the fields of romance and adventure. The better reader plans to continue his education and aspires to a higher level of occupation in the future. He has certain attitudes that differentiate him: the one that is the most clear-cut in the available evidence is a belief in the importance and value of science for human affairs.

It must be confessed that the results of the IEA study provide little guidance for the improvement of the educational enterprise. They point out the very decided importance of the input into any school system in determining its outcomes; but, as in the massive study of schools in the United States included in the Coleman report (1966), the IEA results do little to accentuate the importance of differences between schools in their effects upon students. This is not to s y the schooling in unimportant. It may merely be that schooling is relatively standardized, so that extreme variations in quality tend not to occur. These results are consistent with the view that more extreme variations occur in home and familial backgrounds, and that the school is an aspect of society that provides more nearly standard experiences and opportunities.

Why Do Some Students Learn More Science than Others?

At the fourteen-year-old level, a number of the variables emerging with a consistent pattern of relationships from the regression analyses for the seventeen countries testing in Science were concerned with the exposure of the students to learning Science. 1/ Specifically the grade level of the students, their total study of Science over past years, whether they were currently studying Science or not, the time spent per week on homework, and the opportunity that they had to learn the items tested, contributed to different extents in different countries to variation in Science achievement test scores. In addition, some variable relating to characteristics of the Science teachers in a school were found to be important. The sex of the teachers, whether or not they were teaching Science, were employed full-time, or were members of a Science teachers' association helped to explain differences between students in achievement in Science. Although there were sometimes exceptions to these general findings, these factors of Science exposure and Science teacher characteristics were linked with the student Science performance in the schools.

Of the many school organizational measures assessed in this study, only size of school (total enrollment) was found to be important and then only in three countries. There are many other factors whose influence is legendary in the school systems of different countries, but no evidence of their effects

---

1/ This section is an excerpt from Coomber and Keeves (1973), pp. 266-67.

was found in this study. Perhaps by now in many countries major differences between schools arising from these traditional factors have been eliminated. Alternatively, the effects of such variables as class size may be unique to a country and dependent on other forces operating. In addition, it is possible that the methods of measurement used were too prone to error for significant relationships to be observed. One must, however, report an absence of consistent evidence to support the influence of variables other than those referred to in the preceding discussion on student Science achievement. See Annex Table 2 2 for a list of school variables that were not significant.

With all variables included in the between-student regression analyses, the average percentage of variation in the Science achievement test scores accounted for was 36%, and ranged from 17% in Iran to 55% in Scotland. The remaining rather substantial proportions of unexplained variance are attributable to factors that were not taken into account and to errors of measurement.

At the terminal secondary school level, different patterns of results emerged about factors contributing to achievement in Science from the analyses carried out when compared with the results obtained at the ten-year-old and fourteen-year-old levels. 1/ Where a school system becomes highly selective at the upper secondary school level, the home circumstances of the students cease to account for the differences in performance in Science of those students who remain. This result is consistent with the evidence obtained in other ways that a more homogeneous student group is generally formed with students from more advantaged home backgrounds.

In addition, at the pre-university level the sex of the student becomes an even more powerful factor, with male students performing markedly better than female students on the Science tests. The development of Science as a subject in which boys exceed girls in performance as they move from the ten-year-old to the fourteen-year-old and through to the terminal secondary school level is common to all countries taking part in this investigation, even though these countries differ in the proportions of boys and girls remaining at school. The IEA report shows something of the emergence of this trend but does not effectively expose the causes involved.

Perhaps the most powerful factors influencing achievement in Science at the pre-university level are those related to exposure to learning the subject. Clearly related to a higher level of achievement in Science were the following factors: the extent to which Science had been studied in earlier years, the time given currently to the study of Science both formally in classes and informally at home, and grade level. Some characteristics of the teachers in the schools were also found to be related to student performance in Science. In general, where the teachers had received more post-secondary school training, were members of a subject association, spent more time in preparation either

---

1/ The remaining paragraphs in this section are excerpted from Coomber and Keeves (1973), pp. 279-280.

in or outside school hours, and had taken part in Science curriculum reform activities, their students tended to do better on the Science tests. It must be repeated, however, that although the evidence is only slight, it gains some support by being common to several countries.

While there was no clear evidence to show the extent to which the student's practical experience was related to a higher level of performance in Science, there was some suggestion that if students were given greater opportunities for planning and carrying out limited scientific investigations they tended to do better. This is an area of cinsiderable importance in the planning of secondary school Science course, particularly in the later years, and it is disappointing not be able to report a more definite result.

Final Remarks

The worthwhileness of the results will depend to a great extent on the way in which policy-makers, at whatever level, understand both the strengths and weaknesses of the results. In many cases, this will depend to a large extent on the systematic links already forged or in the process of being forged between the IEA researchers at the national level and policy-makers in each nation. The univariates descriptive statistics on so many input, process, and output variables for various levels in each school system are a mine of information for each national system of education. Some of the between-nation differences, as can be seen from the selected examples given earlier in this paper, have very pointed messages for some national systems. (See also Passow et al., 1974.)

The within-country analyses, despite the problems of correlations between the variables in the final regression analysis and exogenous variables, have identified aspects of learning conditions (as they now exist) that are strongly associated with differences in student achievement. In a cross-sectional study of this kind, cause and effect cannot be proven, only inferred, and there is a strong case for submitting some of the variables shown to be important to a strict experimentation.

For international organizations one of the problems emerging from the IEA study concerns the types of variables on which it would be useful (and why?) to collect systematically standardized data (and how?). What sorts of pilot work can be undertaken to determine how, at low cost, one can collect reliable data from national samples of children in developing countries where the illiteracy rate is high? What type of interdisciplinary work should be undertaken in developing countries to help identify the types of factors on which data should be collected in an effort to "explain" differences between students and between school?

Although evaluators do conceptualize in their roles as researchers, it is clear that in all countries the art of educational theory (or in some cases, perhaps, the communication of it to the educational researchers and evaluators) is in a poor state, as witnessed by the low total variances accounted for. What types of work can be undertaken to improve this state of affairs?

Finally, it is through evaluation projects of this kind that more "hard" information can be collected and used to improve the educational provisions for children in the coming years. The knowledge explosion in measurement and evaluation techniques has been rapid. The cooperative nature of IEA's work has helped researchers in some countries to advance their competency by twenty years in a five-year period. More such enterprises of this kind are needed.

VI. Limitations of Available Research

Each research approach is subject to substantive and methodological problems peculiar to itself. 1/ These problems were discussed other sections and will not be reviewed here. Some research limitations appear throughout educational research, however, and have, we feel, special importance.

First, educational outcomes are almost exclusively measured by cognitive achievement. But the educational system has many functions and many outputs. Cognitive achievement--in particular that part measured by standardized tests--is only one aspect of student learning. Higher cognitive processes(abstract reasoning, problem solving, and creativity, among others) are obviously important educational outcomes--as is non-cognitive achievement. Thus, of the many and diverse kinds of student learning, almost all the educational research that examines student learning is based on a narrow range of cognitive skills. Therefore, current research cannot lead to conclusive generalizations about educational outcomes, because it cannot measure most of them well.

Second, there is virtually no examination of the cost implications of research results. By and large, educational researchers have concentrated on discovering effective educational practices. Virtually no attention has been paid to the notion of cost-effective educational practices. Research results are thus difficult to translate into policy-relevant statements.

Third, few studies maintain controls over what actually goes on in the classroom as it relates to achievement. Data on classroom transactions are the only source of information on the content of the student-teacher relationship. Studies that omit transactions of the data can hope to identify among variables only those broad associations that hold no matter what might be the nature of the relationship between student and teacher. Thus, researchers' results may well be affected by circumstances unrecognized in their analyses.

Finally, the data used by researchers are, at best, crude measures of what is really happening. Concepts such as a teacher's ability to teach or a student's ability to learn are easily discussed, but objective measures of these abilities are extraordinarily elusive; and empirical analysis is based upon measurement. There is no way of knowing the extent to which inconclusive results stem from the researcher's inability to measure accurately the variables he includes in his analysis.

1/ This and the following section are edited versions of Averch et al. (1972), p. 153-57.

VII. Conclusions and Policy Implications

With the limitations of research clearly in mind, we return to the issue of educational effectiveness. 1/ The first major implication of the research is:

Research has not identified a variant of the existing system that is **consistently related to students' educational outcomes. The term "a variant of the existing system" is used to** describe the broad range of alternative educational practices that have been reviewed above. We specifically include changes in school resources, processes, organizations, and aggregate levels of funding.

We must emphasize that we are not suggesting that nothing makes a difference, or that nothing "works." Rather, we are saying that research has found nothing that consistently and unambiguously makes a difference in student outcomes. The literature contains numerous examples of educational practices that seem to have significantly affected student outcomes. The problem is that there are invariably other studies, similar in approach and method, that find the same educational practice to be ineffective; and we have no clear idea of why a practice that seems to be effective in one case is apparently ineffective in another.

We must also emphasize that we are not saying that school does not affect student outcomes. We have little knowledge of what student outcomes would be were students not to attend school at all. Educational research focuses on variants of the existing system and tells us nothing about where we might be without the system at all.

Furthermore, nothing we have found in the educational research literature proves that current educational systems cannot be substantially improved. But the research results we review above provide little reason to be sanguine. Our general conclusion, so far, is that there are few consistent, positive, policy-relevant findings. That is, the research offers little guidance as to what educational practices should be implemented. This condition can arise because that is the way the world really is, or because researchers have been asking the wrong questions, or because the research methods used are not sufficiently powerful, or because the data are "bad." For whatever reason, we can only say that the educational practices examined thus far are only weakly connected to student achievement.

Finally, the educational practices for which school systems have traditionally been willing to pay a premium do not appear to make a major difference in student outcomes. Teachers' experience and teachers' advanced

---

1/ This sections is based on edited extracts from Averch (1972) and is consistent with Jencks (1972).

degrees, the two basic factors that determine salary, are not clearly related to student achievement. Reduction in class size, a favorite high-priority reform in the eyes of many school systems, seems not be be related to student outcomes. In general, the second major implication of the research (and the most important one for school finance) is:

(2) Increasing the expenditures on traditional educational practices is not likely to improve educational outcomes substantially.

The third major policy implication of the research is:

(3) There seem to be opportunities for significant reduction or redirection of educational expenditures without deterioration in educational outcomes.

Researchers have examined many variants of the existing educational system. As we have indicated, none of these variants has been shown to improve educational outcomes consistently. A fact often overlooked is that few have been shown to lead to significantly worse outcomes, either. Consequently, educational research has provided a long list of equally effective variants of the existing system. If these variants are not all equally expensive, then choosing the least expensive provides opportunities to redirect (or even reduce) costs without also reducing effectiveness. 1/

Educational research consists almost entirely of effectiveness studies. There are very few cost-effectiveness studies. The tremendous volume of "negative" results--negative according to the peculiar bias of educational research, which seeks only improvement on the effectiveness side--must surely contain many "positive" results in the sense of indicating less costly methods of accomplishing as much as is currently attained.

The research contains some evidence supporting a fourth major implication:

(4) Innovation, responsiveness, and adaptation in school systems decrease with size and depend upon exogenous shocks to the system. In other words, large systems are less likely to be innovative, responsive, or adaptive than are small systems. Further, whatever the size of the system, innovation is not apt to come from within the system. Outside pressures, from the community or from the federal government, are likely to be needed. Since relatively little research has been directed toward these issues, this finding must be viewed as tentative.

1/ This conclusion applies only to questions of educational effectiveness as now measured. It cannot be applied to justify situations in which constant or decreasing expenditures would impair the health or safety of children and staff.

The implication of this tentative conclusion, however, is clear. There is currently a good deal of interest in government leverage and in the question of whether aid to the schools should be tied or untied. The literature that we have examined suggests that government influence is important in getting innovation into urban school systems, although the hypothesis has not really been tested rigorously.

Our review of educational research supports a fifth major implication:

(5) Educational research is seriously deficient in terms of the size, scope, and focus of research efforts and in the integration of research results. Beyond these specific limitations, educational research has tended to be small in scale, narrow in scope, diffuse, maldistributed, and lacking in focus. In comparison with other major sectors, the amount of research activity devoted to educational problems is surprisingly small. The amount of resources allocated to agricultural research and development in the United States, for example, is more than four times as much, and to health research more than thirteen times as much as education receives. Moreover, educational research is a relatively recent development. Quantitative research on American education goes back to the work of Joseph Meyer Rice in the 1890's; but significant levels of activity did not begin until the late 1950's when first the National Science Foundation, then the Office of Education, began to fund a wide range of research activities.

A comparison of R & D communities by institutional affiliation shows that educational research is very unlike other R & D sectors in the economy because colleges and universities perform the majority of R & D in the educational sector. The typical education study is not founded on a wealth of previous knowledge and understanding, nor is it directed toward the needs of the educational policy-maker. There are virtually no research-based, problem-solving units in the typical education agency in most countries. 1/ The academic community tends to conduct relatively small studies on a part-time basis and to concentrate on basic research. Furthermore, educational research has tended to be almost exclusively the domain of the psychologist. Only recently has it begun to attract the attention of more than a handful of well-trained researchers in other fields. 2/

Finally, the sixth major implication of our work is:

(6) Research tentatively suggests that improvement in student outcomes, both cognitive and non-cognitive, may require sweeping changes in the organization, structure, and conduct of educational experiences.

1/ In 1968 there were in the United States only 1,300 man-years devoted to research, development, or innovation in the almost 20,000 state and local education agencies; most of that was devoted to testing and to gathering statistics (Levien, 1971).

2/ See Levien (1971) for a discussion of the current state of educational research.

This inference follows from the first four conclusions cited above, as well as from the testimony of the experiential approach. Even the fifth conclusion, which cites the paucity of educational research, tends to reinforce the point that drastic changes in education may be necessary because it implies that marginal changes in research have been and will be inadequate to indicate clearly the directions educational improvement should take.

BIBLIOGRAPHY

Anderson, James G., Bureaucracy in Education, Baltimore, Johns Hopkins Press, 1968.

Averch, Harvey et al., How Effective Is Schooling? A Review of the Research, Santa Monica, Cal., The Rand Corporation, 1972.

------, and Kiesling, Herbert, "The Relationship of School and Environment to Student Performance: Some Simultaneous Models for the Project TALENT High Schools," Santa Monica, The Rand Corporation, 1970 (mimeographed).

Bissell, Joan S., "The Cognitive Effects of Preschool Programs for Disadvantaged Children," Ph.D. dissertation, Harvard University, 1970.

Bloom, B.S. (ed.), Taxonomy of Educational Objectives; The Classification of Educational Goals, Handbook I, Cognitive Domain, New York, Longmans, Green, 1956.

Bowles, Samuel, and Gintis, Herbert, "IQ in the U.S. Class Structure," Social Policy 3(1973): 65-96.

Cain, Glen, and Watts, Harold, Problems in Making Policy Inferences from the Coleman Report, Institute for Research on Poverty, Madison, University of Wisconsin, 1968.

Carroll, John B., French as a Foreign Language in Seven Countries, International Studies in Evaluation, vol. 5, Stockholm, Almqvist & Wiksell, 1974.

Central Advisory Council for Education, Children and Their Primary Schools, The Plowden Report, London, HMSO, 1967.

Chopra, S.L., "Parental Occupation and Academic Achievement of High School Students in India," The Journal of Educational Research 60 (April 1967):359-62.

Cline, Victor B. et al., "Creativity Tests and Achievement in High School Science," Journal of Applied Psychology 47(1963):184-89.

Coleman, James et al., Equality of Educational Opportunity, U.S. Department of Health, Education, and Welfare, Washington, D.C., Government Printing Office, 1966.

Coomber, L.C., and Keeves, John P., Science Education in Nineteen Countries: An Empirical Study, International Studies in Evaulation, vol. 1, Stockholm, Almqvist & Wiksell, 1973.

Dennison, George, The Lives of Children, New York, Random House, 1969.

Dreeben, Robert, On What Is Learned in School, Reading Mass., Addison-Wesley, 1968.

Farnen, R.F., Marklund, S., Oppenheim, A.N., and Torney, Judith V., Civic Education in Ten Countries, International Studies in Evaluation, vol. 6, Stockholm, Almqvist & Wiksell, 1974.

Farquhar, W.M., and Christensen, E.W., Motivational Factors Influencing Academic Achievement of Eleventh Grade Puerto Rican High School Students, East Lansing, Michigan State University Press, 1968.

Friedenberg, Edgar Z., Coming of Age in America, New York, Random House, 1963.

Gintis, Herbert, "Alienation and Power: Towards a Radical Welfare Economics," unpublished Ph.D. dissertation, Harvard University, May 1969.

_____, "Education, Techology, and the Characteristics of Worker Productivity," The American Economic Review 61(May 1971):266-79.

-----, and Bowles, Samuel, "The Ideology of Educational Reform," unpublished paper, Cambridge, Mass., 1971.

Goodman, Paul, Growing Up Absurd, New York, Random House, 1965.

------, The New Reformation, New York, Random House, 1970.

Havighurst, R.J., The Public Schools of Chicago, Chicago, Board of Education, 1964.

Henry, Jules, The Way Its Spozed to Be, New York, Simon and Schuster, 1968.

------, How to Survive in Your Native Land, New York Simon and Schuster, 1971.

Holland, John L., "Creative and Academic Performance Among Talented Adolescents," in Robert E. Grinder (ed.), Studies in Adolescence, New York, Macmillan, 1963.

Holt, John, What Do I Do on Monday?, New York, E.P. Dutton, 1970.

Husen, T. (ed.), International Study of Achievement in Mathematics, 2 vols., New York, John Wiley and Sons. 1967.

----- et al., Naturorienteranda Amnen, Stockholm, Almqvist & Wiksell, 1973.

Illich, Ivan, Deschooling Society, New York, Harper & Row, 1971.

Jencks, Christopher et al., Inequality: A Reassessment of the Effect of Family and Schooling in America, New York, Basic Books, 1972.

Katzman, Martin T., "Distribution and Production in a Big City Elementary School System," Yale Economic Essays 8(Spring 1968):201-56.

Klein, S.P., "The Uses and Limitations of Standardized Tests in Meeting the Demands for Accountability," UCLA Evaluation Comment, Center for the Study of Evaluation, 2, No. 4(January 1971).

Kohn, N., "Social Class and Parental Values," in R. Coser (ed.), The Family, New York, St. Martin's Press, 1964.

Kozol, Jonathan, Death at an Early Age, Boston, Houghton Mifflin, 1968.

Kubiniec, Cathleen M., "The Relative Efficiency of Various Dimensions of the Self-Concept in Predicting Academic Achievement," American Educational Research Journal 7(May 1970):321-36.

Lewis, E. Glyn, English as a Foreign Language in Ten Countries, International Studies in Evaluation, vol. 4, Stockholm, Almqvist & Wiksell, 1974.

Light, Richard J., and Smith, Paul N., "Accumulating Evidence: Procedures for Resolving Contradictions Among Different Research Studies," Harvard Educational Review 41(1971):429-71.

Mollenkopf, William G., and Melville, S. Donald, A Study of Secondary School Characteristics as Related to Test Scores, Research Bulletin RB-56-6, Princeton, Educational Testing Service, 1956.

Ortar, Gina R., "Educational Achievements of Primary School Graduates in Israel as Related to Their Socio-Cultural Background," Comparative Systems, International Studies in Evaluation, vol. 7, Stockholm, Almqvist & Wiksell, 1974.

Peaker, G.F., An Empirical Study of Education in Twenty-One Countries: A Technical Report, International Studies in Evaluation, vol. 8, Stockholm, Almqvist & Wiksell, 1974.

Postlethwaite, T. Neville, "A Selection from the Overall Findings of the IEA Study in Science, Reading Comprehension, Literature, French as a Foreign Language, English as a Foreign Language, and Civil Education," Conference on Educational Achievement, Harvard University, November 1973.

Postman, Neil, and Weingartner, C., Teaching as a Subversive Activity, New York, Dell Publishing Company, 1969.

Purves, Alan C., Literature Education in Ten Countries: An Empirical Study, International Studies in Evaluation, vol. 2, Stockholm, Almqvist & Wiksell, 1973.

Rogers, David, 110 Livingston Street: Politics and Bureaucracy in the New York Schools, New York, Random House, 1968.

Rosenshine, B., and Furst, J., "Current and Future Research on Teacher Performance Criteria," in B.W. Smith (ed.), Research on Teacher Education: A Symposium, Englewood Cliffs, N.J., Prentice-Hall, 1971.

Rosenthal, R., And Jackson, L., Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development, New York, Holt, Rinehart and Winston, 1968.

Rosier, Malcolm J., Science Achievement in Australian Secondary Schools, Melbourne, Australia, Australian Council for Educational Research, 1973.

Shimada, S. et al., International Survey of Science Education: Report of Japanese National Commission of IEA, Results of Japanese Survey, vol. 1, Tokyo, Japan, NIER, March 1973.

Silberman, C.E., Crisis in the Classroom, New York, Random House, 1970.

Simmons, John, "The Report of the Faure Commission: One Step Forward and Two Steps Back," Higher Education 2(November 1973) 475-88.

------, and Alexander, Leigh, "Educational Production Functions in Developing Countries: A Review of the Research," Washington, D.C., IBRD, 1973.

Thomas, James Alan, "Efficiency in Education: A Study of the Relationship Between Selected Inputs and Mean Test Scores in a Sample of Senior High Schools," Ph.D. dissertation, Stanford University Library, 1962 (microfilm).

Thorndike, Robert L., Reading Comprehension Education in Fifteen Countries: An Empirical Study, International Studies in Evaluation, vol. 3, Stockholm, Almqvist & Wiksell, 1973.

U.S. Department of Health, Education, and Welfare, U.S. Office of Education, Do Teachers Make a Difference?, OE-58042, Washington, D.C., 1970.

Weber, Max, "On Bureaucracy," in Hans Gerth and C. Wright Mills (eds.), From Max Weber: Essays in Sociology, New York, Oxford University Press, 1958.

-------, "The Meaning of Discipline," op. cit., pp. 195-96.

Westbury, I., "Curriculum Evaluation," Review of Educational Research 40(1969):239-60.

Wolff, Max, and Stein, Annie, "Head Start Six Months Later," Phi Delta Kappan 48(March 1967):349-50.

Wonnacott, R.S., and Wonnacott, T.H., Econometrics, New York, John Wiley and Sons, 1970.

Wrightston, J. Wayne et al., Evaluation of the Higher Horizons Program for Underprivileged Children, New York, New York City Board of Education, 1964.

Annex

Table 1. Subject areas and populations tested by country

	Science				Reading Comprehension				Literature			French				English			Civic Education			
	I	II	III[a]	IV	I	II	III[a]	IV	I	II	IV	I	II	III[a]	IV	II	III[a]	IV	I	II	III[a]	IV
Australia		x		x																		
Belgium (French)	x	x		x	x	x		x	x		x					x		x				
Belgium (Flemish)	x	x		x	x	x		x	x		x					x		x				
Chile	x	x		x	x	x		x	x		x				x			x				
England	x	x	x	x	x	x	x	x	x	x	x		x	x	x							
FRG	x	x		x	x	x		x								x		x	x	x		x
Finland	x	x	x	x	x	x	x	x	x	x	x					x	x	x		x	x	x
France				x																		
Hungary	x	x		x	x	x		x									x	x				
India	x	x		x	x	x		x														
Iran	x	x		x	x	x		x	x		x									x		x
Ireland																				x		x
Israel					x	x		x								x		x	x	x		
Italy	x	x		x	x	x		x	x		x					x		x	x	x		
Japan	x	x																				
Netherlands	x	x		x	x	x		x					x		x	x		x	x	x		x
New Zealand		x		x		x		x	x		x		x		x					x		x
Poland																						
Rumania													x		x							
Scotland	x	x		x	x	x		x					x		x							
Sweden	x	x	x	x	x	x	x	x	x	x	x			x	x	x	x	x			x	x
Thailand	x	x		x												x	x	x				
USA	x	x		x	x	x		x	x		x	x	x		x					x		x

(a) Population III was not an internationally defined population. Each participation research centre was able to define a population somewhere between Pop II and IV to test with Pop II for national purposes. The analyses were national and not international.

Annex Table 2

Student, Teacher and School Variables in French Study

STUDENT	TEACHER	SCHOOL
Age	French in Classroom	Class size
Grade	Teach other subjects in French	Principal's degree
Father's occupation	Hours instruction in French	Years as principal
Expected occupation	Sex	Years as principal of this school
Father's language	Age	Years teaching experience
Mother's language	Specialist teacher	Type of community
First language	Full-time education	Museum available
Second language	Post-secondary education	Zoo available
Sex	Teaching experience	Public library available
Father's education	Time teaching at this school	Concert hall available
Mother's education	Hours preparation of lessons	Opera available
Expected education	Hours marking papers	Foreign language societies available
Hours homework	Membership of teachers association	Amenities score
Place of homework	Teaching journals	Total enrolment, boys
Fixed time for homework	Subject journals	Total enrolment, girls
Help with homework	Attended conferences	Total enrolment
Parents correct speech	Assessment, standardised tests	Lowest grade in school
Parents check spelling	Assessment, essay tests	Highest grade in school
Dictionary in home	Assessment, objective tests	Mean grade of school
Parents encourage reading	Assessment, homework	Grade range of school
Parents' interest in school	Assessment, projects and papers	Enrolment of Pop.I
Encourage culture	Assessment, variety and extent	Enrolment of Pop.II
Daily newspapers	Importance of student needs	Enrolment of Pop.IV
Books in home	Importance of curriculum	Co-education, Pop.I
Hours pleasure reading	Importance of text books	Co-education, Pop.II
TV and radio	Importance of examinations	Co-education, Pop.IV
Size of family	Importance of next grade	Day or boarding school
Position in family	Methods text books	Beginning French grade
Like/dislike of different subjects	Methods drill materials	Beginning English grade
Interest in French		Beginning Social Studies Grade
French activities out of school		

STUDENT	TEACHER	SCHOOL
Mother studied French	Methods individual materials	Decision making, textbooks
Parents interest	Methods small group work	Decision making, rules
Help with French homework	Methods indivudual tutoring	Decision making, choosing teachers
Utility of French	Methods audio and visual	Decision making, conditions
Aspiration in French skills	Methods Field trips	Decision making, student selection
Grade beginning French	Methods Lectures	Decision making, expenditure
Entry knowledge	Methods questioning	Decision making, tuition fees
Years studied French	Methods discussion	Decision making, administration
Perception spoken French	Methods variety and extent of approaches	Decision making, syllabus
Perception listening French	Within class grouping	Decision making, methods
Perception reading French	Part of time employed	Decision making, syllabus and methods
Perception writing French	Teacher training institution	Inspection
Opportunity to speak French	Foreign language teaching load	Report to authorities
Time in French country	French in other subjects	Advice on school problems
Ease/Difficulty French	Other foreign language	Advise teachers
Perceived industry in French	Grade range of French teaching	Assess teachers
French books in home	Number of groups taught	Role of inspection
Frequency of exposure to French books	Total time teaching French	School operating costs
	Mother tongue	Teacher salaries
	Age beginning French	Non-teaching staff salaries
	Perceived listening skill	Maintenance and repair
	Perceived speaking skill	Books and stationery
	Perceived reading skill	Purchase of equipment
	Perceived writing skill	Other (loan charges)
	Perceived pronunciation skill	Total budget
	Residence French country	Total budget per pupil
	Tertiary French in years	Teacher salaries per pupil
	Method training	Foreign language percentage budget
	Years teaching foreign language	Number full-time staff
	Foreign language association member	Pupil/teacher ratio
		Teacher salaries per staff
		Number of teachers male

STUDENT

TEACHER

Use of mother tongue, beginning

Use of mother tongue, intermediate

Use of mother tongue, advanced

Emphasis choices : listening comprehension

Emphasis choices : speaking fluency

Emphasis choices : correct pronunciation

Emphasis choices : reading comprehension

Emphasis choices : ability to write

Order of spoken and written French

Grammar teaching

Speaking emphasis

Pronunciation methods

Teaching aids, blackboard

Teaching aids, pictures

Teaching aids, "props"

Teaching aids, film strips

Teaching aids, sound movies

Teaching aids, phonograph records

Teaching aids, tape recorder

Teaching aids, language lab.

Teaching aids, variety and extent

SCHOOL

Number of teachers in French

Number of teachers in English

Number of teachers in Social Studies

School librarian

French foreign language assistant

English foreign language assistant

Foreign language laboratory technician

Total number of language auxiliaries

Admission criteria, residence

Admission criteria, performance

Admission criteria, interview

Admission criteria, examination

Admission criteria, graduation

Admission criteria, membership

Streaming practices

School programme, variety

Type of course

Language of instruction

Student decision making

Hours per week schooling, Pop.I

Hours per week schooling, Pop.II

Hours per week schooling, Pop.IV

Weeks per year schooling

Hours per year schooling, Pop.I

Hours per year schooling, Pop.II

Hours per year schooling, Pop.IV

When spoken French introduced

When Reading and Writing French introduced

Class conducted in French

Selection of French students

Percentage study French, Pop.I

Percentage study French, Pop.II

Percentage study French, Pop.IV

STUDENT	TEACHER	SCORE
		Number of periods French, beginning
		Length of periods French, beginning
		Level present, beginning
		Number of periods French, intermediate
		Length of periods French, intermediate
		Number of periods of French, advanced
		Length of periods French, advanced
		Total time French, beginning
		Total time French, intermediate
		Total time French, advanced
		Level present, advanced
		Native teachers French

Annex Table 3

Contribution by blocks to Total Variance from Between-student Regression Analyses

SCIENCE

Increment / Country	Population I				Population II				Population IV			
	Block 1	Block 2	Block 3	TOTAL	Block 1	Block 2	Block 3	TOTAL	Block 1	Block 2	Block 3	TOTAL
Australia	-	-	-	-	16	7	11	39	13	3	20	44
Belgium(Fl.)	4	2	8	21	8	3	12	26	-	-	-	-
Belgium(Fr.)	12	0	21	36	-	-	-	-	-	-	-	-
Chile	4	0	9	26	13	4	6	25	19	3	8	32
England	21	0	3	32	23	17	7	52	13	2	41	61
Fed.Rep.of Germany	8	2	10	24	18	2	14	34	13	12	8	39
Finland	14	0	4	26	22	6	10	44	25	19	7	56
France	-	-	-	-	-	-	-	-	16	26	5	49
Hungary	8	0	7	20	14	3	5	31	11	13	10	41
India	1	0	20	29	3	10	8	24	4	4	17	26
Iran	6	14	6	32	5	1	9	17	3	0	4	11
Italy	4	0	4	14	10	4	6	24	10	2	16	30
Japan	17	0	1	22	23	0	4	40	-	-	-	-
Netherlands	16	1	7	29	19	15	10	49	21	5	31	63
New Zealand	-	-	-	-	17	12	8	45	13	1	31	54
Scotland	22	1	5	36	29	11	9	55	19	1	34	63
Sweden	16	0	5	24	18	0	7	36	18	8	20	52
Thailand	-	-	-	-	10	3	23	37	-	-	-	-
United States	18	1	9	34	22	2	7	36	18	9	8	39
MEAN	11.4	1.4	7.9	27.0	15.8	5.9	9.2	36.1	12.7	6.4	15.3	38.8
Highest	22	14	21	36	29	17	23	55	25	26	41	63
Lowest	1	0	1	14	3	0	4	17	3	0	4	11

Annex Table 4

Contribution by blocks to total variance from between-student Regression Analyses

READING COMPREHENSION

Increment / Country	Population I				Population II				Population IV			
	Block 1	Block 2	Block 3	TOTAL	Block 1	Block 2	Block 3	TOTAL	Block 1	Block 2	Block 3	TOTAL
Belgium(Fl.)	1.7	0.0	11.6	22.5	7.5	10.7	9.3	33.7	12.7	18.5	3.4	42.3
Belgium(Fr.)	16.7	1.1	18.4	38.8	11.2	14.5	10.3	43.2	7.2	8.4	9.0	29.7
Chile	1.4	1.2	8.3	22.3	20.1	5.9	6.6	38.6	14.9	4.9	5.1	28.9
England	22.1	0.4	1.5	31.4	27.3	13.7	2.6	50.8	2.4	1.8	4.6	16.4
Finland	17.7	0.0	2.8	27.8	20.3	13.4	4.3	46.8	11.5	2.3	2.3	22.0
Hungary	18.7	0.3	3.8	28.4	18.6	3.9	4.0	36.2	6.9	8.5	4.8	27.1
India	1.6	0.3	14.9	30.8	1.4	3.1	9.7	20.0	4.0	2.2	4.7	12.9
Iran	8.7	9.1	7.2	30.3	6.1	3.2	6.6	20.0	5.5	0.0	7.6	14.5
Israel	25.4	1.9	3.6	36.6	25.2	13.9	4.3	48.3	8.6	19.4	15.2	46.9
Italy	9.6	0.2	4.4	18.0	10.4	12.4	3.2	33.6	9.6	7.8	5.0	25.3
Netherlands	11.1	1.6	4.0	26.4	12.5	20.7	3.9	45.8	1.0	7.4	4.8	16.3
New Zealand	-	-	-	-	13.5	17.6	8.2	46.8	9.0	1.6	1.9	19.2
Scotland	23.7	0.3	2.3	33.8	26.1	12.3	3.9	50.8	4.0	1.4	3.5	20.2
Sweden	11.4	0.3	2.9	18.3	16.1	0.0	2.9	33.7	4.2	14.0	1.8	24.9
United States	19.8	2.0	3.8	31.8	22.1	3.4	6.5	42.6	17.5	8.0	2.9	33.6
MEAN	13.5	1.3	6.4	28.4	15.9	9.9	5.8	39.4	8.0	7.1	5.1	25.3
Highest	25.4	9.1	18.4	38.8	27.3	20.7	10.3	50.8	17.5	19.4	15.2	46.9
Lowest	1.4	0.0	1.5	18.0	1.4	0.0	2.6	20.0	1.0	0.0	1.8	12.9

Annex Table 5

Contribution by blocks to Total Variance from Between-student Regression Analyses

CIVICS COGNITIVE

Increment / Country	Population II				Population IV			
	Block 1	Block 2	Block 3	TOTAL	Block 1	Block 2	Block 3	TOTAL
Fed. Rep. of Germany	17.6	6.5	17.7	57.2	4.7	0.6	6.8	27.8
Finland	19.0	11.2	9.1	60.6	13.8	2.9	0.9	37.2
Ireland	18.6	4.7	17.0	62.2	9.0	0.8	9.0	43.5
Italy	8.5	10.5	12.2	56.3	-	-	-	-
Netherlands	15.3	20.2	7.5	55.2	8.2	7.9	5.2	31.2
New Zealand	14.4	14.7	12.1	60.0	8.0	0.2	3.3	33.7
Sweden	-	-	-	-	7.5	15.0	6.0	43.2
United States	20.4	7.0	2.8	61.8	18.3	9.0	4.1	56.7
MEAN	16.3	10.7	11.2	59.0	9.9	5.2	5.0	39.0
Highest	20.4	20.2	17.7	62.2	18.3	15.0	9.0	56.7
Lowest	8.5	4.7	2.8	55.2	4.7	0.2	0.9	27.8

Annex Table 6

Contribution by blocks to Total Variance from Between-student Regression Analyses

ENGLISH - READING

Increment	Population II					Population IV				
Country	Block 1	Block 2	Block 3	Block 4	TOTAL	Block 1	Block 2	Block 3	Block 4	TOTAL
Belgium(Fr.)	14.2	21.0	6.5	9.8	57.0	6.0	3.4	10.7	15.6	41.9
Chile	-	-	-	-	-	8.2	8.5	4.7	15.2	42.0
Fed.Rep.of Germany	15.1	30.7	14.0	13.0	78.5	1.0	0.5	7.5	8.9	34.0
Finland	25.7	40.4	5.8	2.2	78.2	4.3	0.2	19.3	2.3	43.1
Hungary	-	-	-	-	-	2.9	12.7	2.1	17.5	41.6
Israel	22.4	4.1	5.5	4.4	53.5	9.7	12.1	7.7	5.2	42.3
Italy	1.2	7.5	5.4	13.2	36.1	16.4	30.9	5.8	15.2	75.0
Netherlands	10.6	51.7	8.4	2.1	77.5	0.6	1.2	2.8	5.8	22.4
Sweden	12.5	0.0	8.6	4.2	54.4	3.1	19.5	1.5	2.3	47.0
Thailand	8.2	1.4	14.5	9.9	47.1	12.8	5.3	5.0	17.1	48.3
MEAN	13.7	19.6	8.6	7.4	60.3	6.5	9.4	6.7	10.5	43.8
Highest	25.7	51.7	14.5	13.2	78.5	16.4	30.9	19.3	17.5	75.0
Lowest	1.2	0.0	5.4	2.1	36.1	0.6	0.2	1.5	2.3	22.4