(2)

PN-AIS1-827
37926

REVISED DRAFT - [redacted]

RECOGNIZING, FOSTERING AND MODELING THE EFFECTIVENESS OF SCHOOLS AS ORGANIZATIONS IN THIRD WORLD COUNTRIES

John Schwille, Alemu Beeftu, Richard Navarro,
Robert Prouty, Stephen Raudenbush, William Schmidt,
Mun Tsang, Christopher Wheeler

August 4, 1986

Institute for Research on Teaching

and

Office for International Networks in Education and Development

Michigan State University

Prepared for submission to Harvard University in fulfillment of literature review and synthesis requirements for Project BRIDGES. Except for the research team leader, authors are listed in alphabetical order; each has taken responsibility for and made substantial contributions to major parts of this report.

TABLE OF CONTENTS

INTRODUCTION

What are the attributes of effective schools? What strategies might improve schools? What issues do policymakers need to confront as they consider different strategies? This review explores these important and complicated questions.

For Third World countries, as for industrialized countries, such questions are not trivial. In developing countries, school quality and school management are now recognized as high priority issues, especially at the primary level. There are several reasons for this increased interest. First, individual and social returns to investments in primary schooling have been found to be greater than returns to investments at higher levels of education (Psacharopoulos and Woodhall, 1985; Colclough, 1980). Second, Third World countries have disproportionately invested in primary schooling in herculean efforts to expand their education sector (Williams, 1984; Porter, 1984). Third, the social demand for formal education remains very high (Heyneman, 1984; Colclough, 1980), creating an unmet need for primary education which is hard to finance, given current financial circumstances in Third World countries. Fourth, recent research indicates that investments to improve the quality of schooling can have a higher rate of return than investments to expand access or level of achieved education (Fuller, 1985; Behrman and Birdsall, 1983).

Simultaneous demands to increase access to primary schooling and to improve the quality of existing schools--in the context of scarce resources for

education--create acute policymaking dilemmas for Third World governments. In most countries it is unrealistic to expect an influx of fresh resources for education, so crucial policy choices now require evaluation of alternative means of increasing the efficiency of allocating existing resources. But successful evaluation requires empirical evidence about costs and likely consequences of manipulating key variables over which policymakers have some, but not complete control. Providing such evidence constitutes one of BRIDGES' primary goals.

Within this context, the central role of Michigan State University's research team is to make accessible to policymakers research findings which can illuminate the costs and consequences of alternative strategies for improving the functioning of schools as organizations. These research findings may be derived from the synthesis of research findings already available, the analysis or reanalysis of data already collected, and the collection and analysis of new data. From the standpoint of cost-effectiveness, new data should be collected only when it can be demonstrated that available data does not contain answers to crucial questions. Our present literature review seeks to provide a broad synthesis of available literature on school improvement, to identify already available findings which can be helpful to policymakers and to create a conceptual basis for further needed research.

To identify attributes of effective schools, to consider possible strategies for improving schools and to reduce the unintended effects of such strategies, we have carefully examined literature from developing and industrialized countries. While we have not been exhaustive in selecting literature for review, we have gone well beyond what is commonly known as the effective schools literature in the United States. To provide an analysis

appropriate to schools in developing countries, we have examined literature on the school as an organization, the anthropological literature that puts school characteristics into a context of cultural and social variation, and the growing literature on the importance of certain school resources in raising student achievement in developing countries. Such a quantity of literature could not be covered in depth given the time frame for our review. What we did was to adopt a selective, though still systematic approach to the review by: (1) assembling and prioritizing lists of relevant references (one for industrialized countries and one for Third World countries); (2) systematically screening key journals; (3) obtaining unpublished materials, such as papers presented at the 1986 meetings of the American Educational Research Association; (4) using bibliographies from key reviews already identified; (5) preparing abstracts for works found to be particularly important. In selecting studies to review, we also paid particular attention to existing reviews, usually prepared for other purposes.

Conceptualization of the synthesis

Our interpretation and synthesis of the literature suggests that the success of attempts to improve effectiveness and efficiency of schools depends primarily on the improvement of teacher effectiveness in the classroom. In our view, school organization is a context which either fosters or inhibits teacher engagement in instructional activities, teacher expectancies for their own and their students' success, teacher knowledge and convictions about what to teach, teacher effective use of instructional materials, and teacher adaptability to cultural and other background differences in their students. These central measures of organizational health are portrayed as Box C in the figure which illustrates our conceptualization (Figure 1). We further postulate that

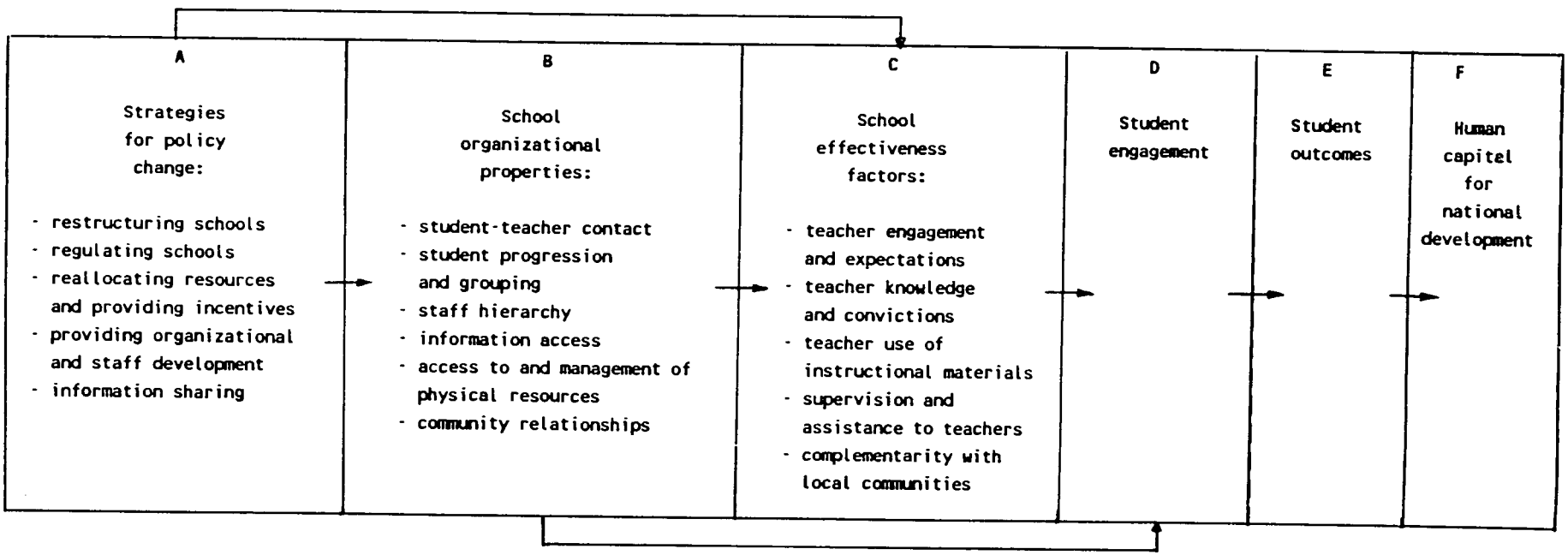

Figure 1: A conceptual model for school improvement

policy-directed changes in schools are primarily intended to increase the effective engagement of children (Box D). This increased engagement, in turn, serves to raise student outcomes (Box E) and to increase a nation's human capital: the skills, knowledge, habits, and predispositions needed by the population to contribute to national economic and political development (Box F).

The organizational context within which school effectiveness factors operate is illustrated by Box B in Figure 1. These structural and environmental constraints and facilitators are conceived in terms of parameters which determine the extent of student-teacher contact, the criteria for student progression and grouping, the nature of hierarchical relations within the school staff and between the staff and the larger organizational entities within which the school is embedded, the facilities which exist for giving students and teachers access to educationally relevant information, the stock of physical resources available to the organization and provisions for their use, and the nature of the links between the school and its various community constituencies. To the extent that these organizational factors affect student outcomes, they are expected to do so through the mediation of the school effectiveness factors (Box C) and/or independently through effects on student engagement (represented by the arrow from Box B to Box D).

Since policymakers are not actors directly involved in the life of individual schools, a set of strategies is included in the diagram to represent the options available to them for intervening to improve the effectiveness of schools as organizations. These strategies are represented by Box A in the figure. They include restructuring schools, regulating schools, reallocating resources or changing the distribution of incentives, providing organizational and staff development, and engaging teachers and headmasters in reflection to

share what they already know. Again it is possible for these strategies to operate directly on the school effectiveness factors (as represented by the arrow from A to C) and/or through the mediation of the organizational factors (A to B to C). The former offers the most direct route to increased effectiveness, but the latter has appeal inasmuch as the organizational factors are more easily changed through top-down initiatives.

Relationship to other BRIDGES teams

A conceptual scheme for school improvement inevitably intersects with the goals and concerns of other BRIDGES teams, especially the classroom management, learning technologies, and external efficiency teams. In a sense all teams share the same ultimate dependent construct: that is, to maximize the human capital needed for national development (Box F). What distinguishes the teams is the set of independent variables whose effects constitute their primary focus of study. For the MSU team, the primary interest is in the organizational factors of Box B and how they interact with Box C in influencing student engagement and student outcomes (Boxes D and E). For the classroom management and learning technologies teams, the key independent variables have to do with various instructional variables (partly represented by Box C) and their effects on student engagement and outcomes. The independent variables of interest to these teams include some of the same ones that, for the MSU team, mediate the effect of school level variables on student outcomes. For the external efficiency team, the key independent variables have to do more with the student outcomes (Box E) and their effect on what students do later in contributing to national human capital (Box F). Since the objects of study of the various teams are inextricably intertwined, the BRIDGES project

will function best if these teams collaborate, each keeping in mind its special focus while maintaining a vision of the whole picture.

Schools, school effectiveness and school improvement

Schools. Although it is common to envisage schools as buildings containing classrooms or places where children go for instruction, it is more useful to think of the school as a social organization with leadership, shared values and norms, and social conflicts; where administrators, teachers, staff and students, each group characterized by different interests and perspectives, negotiate expectations for one another's behavior. Of course, the power in this negotiation is asymmetric: administrators have major control over hiring, promotion and termination of staff; and administrators and staff have means of sanctioning unwanted student behavior. However, the students, while learning to comply with adult expectations, also develop subtle and overt means of influencing adult behavior. The school is thus a social microcosm where children learn to function in society, both complying with and shaping social institutions. This function of schooling has often been termed its socialization function.

Schools also have explicit academic instructional goals, and for primary schools in Third World countries, teaching literacy and numeracy are among the most important. The socialization and academic purposes of schooling are closely interrelated. Societies intend for schools as microcosmic cultures to value academic achievement. Yet while in some schools academic achievement is indeed accorded high status, in others the conflict between academic achievement and other sources of high status defines a fundamental tension in the social relations of the school.

School effectiveness. Our discussion implies that effective schools teach children both the academic and social skills they need to navigate the byways of social institutions. This view of competency goes beyond the measures of literacy and numeracy currently in vogue among those who study school effectiveness. It suggests that measuring the ability to read and compute through standardized tests and then using the results to distinguish between "effective" and "ineffective" schools is inadequate. Test scores are proxies for learning, not direct measures of how well children can read and solve problems under conditions other than testing. Other ways of assessing a child's contributions to individual and group problem solving are needed.

Schooling has multiple outcomes, and hence defining school effectiveness operationally is challenging and likely to arouse intense disagreement. It requires specification of which outcomes are most valued. Moreover, even to the extent that consensus can be achieved on the relative worth of various outcomes, to be clear about effectiveness requires judgment about how these key outcomes should be distributed socially within a school. The question of which is the optimal social distribution of school outcomes is primarily a question of social values.

While reasonable people will disagree on criteria for school effectiveness, it seems self-evident that most academic outcomes likely to be considered important can be obtained only by engaging children actively in school pursuits. Thus, the extent of active student engagement (Box D of Figure 1) becomes a key proximal outcome for school improvement efforts, whatever the distal outcomes (Boxes E and F).

Organization of the essay: Three parts

Part one. In Part one we discuss the importance of school effectiveness factors and school organizational properties for understanding student engagement and outcomes. Active student engagement is the result of what might be called healthy social relations of instruction (Box C). These social relations summarize the extent to which the school, as a social organization, enhances or inhibits effective teacher functioning. These healthy social relations remove obstacles to teaching, reduce diversions from instruction, and eliminate annoyances and deprivations which debilitate commitment to teaching. Often these social relations are healthy primarily because they involve teachers self-consciously and collectively in creating conditions favorable to effective instruction.

These factors can be best understood as school level variables. It is true that teachers' expectations for their students' success result in part from the beliefs, values, and experiences they bring to a school when they are hired. Nevertheless studies of schools reveal that institutional norms and support mechanisms shape teachers' views of their own and their students' possibilities for growth. Similarly, the level of a teacher's knowledge in part represents prior educational experiences, but our interest focuses on the extent to which policy changes and organizational factors foster either the growth of new knowledge or, alternatively, devalue and attenuate the knowledge that teachers bring with them to the job.

This point of view may not appear to give enough weight to more tangible and easily measurable variables such as school facilities, resources and structural characteristics. We do not ignore these variables. However, they should be viewed as manipulable aspects of schooling which can be made to improve schools primarily, though not solely, by improving the social

organization of instruction. These tangible variables are discussed at the end of the first part of our essay.

Part two. Knowing about the attributes of effective schooling is important, but it does not tell us how to change schools to ensure that such attributes become important parts of the way a school works. Namuddu's research in Kenya is a case in point. Her efforts to create videotapes as the basis for improving the teaching of other teachers encountered considerable difficulty. Teachers who might have learned something by watching and reflecting on the episodes of effective teaching dismissed them as irrelevant because the schools portrayed were organizationally different and therefore seen as irrelevant to their own situations. Due to the perceived organizational differences between two types of schools, the analysis of effective teaching by itself did not suffice to bring about change (personal communication).

Since recognizing the characteristics of effective school organization is not sufficient to bring about change, the second part of our essay is devoted to a discussion of the strategies that policymakers often use in attempting to bring about school improvement (Box A of Figure 1).

Part three. A principal finding of policy implementation research is that every policy has unanticipated consequences that undermine the accomplishment of intended goals---an ironic social equivalent to the physical law that for every action there is an equal and opposite reaction. This means that any analysis which stops with descriptions of the attributes of effective schooling and of strategies policymakers often use to bring about school improvement is incomplete. An ability to structure a process of implementation that enables policymakers to better anticipate unintended consequences and therefore to manipulate conditions to reduce their effect is critical to successful change.

Part three provides a way of analyzing this important issue with an eye towards helping policymakers create such a process.

It is also in this last part of our paper that the reader will see how we ultimately view effectiveness. Effectiveness, we argue, should not be elaborately defined in a priori fashion since any definition requires judgments from policymakers as well as other stakeholders about the kind of student outcomes desired and the desired distribution of such outcomes. It is also necessary to know what the school contributes to these outcomes over and above the contributions of out-of-school and prior factors. Finally, it requires knowledge of the cost of the contributions the school makes.

The results of this review show that much is known about the nature of school effectiveness. It also shows, however, that modeling and making good policy about school effectiveness for a particular nation or region requires specific knowledge about these settings. Our review has much to say about what knowledge is required and how it can be analyzed and used.

PART I:

Factors Important to School Success

In recent years, the nature of school effectiveness has been the subject of numerous research studies and various reviews (e.g., Brookover et al., 1979; Cohen, 1983; Edmonds, 1978; Lezotte and Bancroft, 1985; Odden and Webb, 1983; Purkey and Smith, 1983, 1985; and Rowan et al., 1983). Our analysis of the attributes of effective schools represents a departure from these and other studies in several respects. First, we do not focus exclusively on research from the United States or other Western industrialized countries (cf. Cohn and

Rossmiller, 1985). Instead, we draw from their findings and synthesize them with research findings from Third World countries. Second, we have consciously chosen to highlight variables that offer the most promising avenues to school improvement in Third World countries. And finally, we have considered the interaction between school effectiveness factors and their organizational context.

The five factors we analyze below are meant to have face validity for policymakers, practitioners and researchers alike. They represent judgments based on our reading of the most relevant literature we could obtain. They do not represent a definitive list, given the need for substantial additional research and the imperative for new policy initiatives in this area.

I-A. Teacher engagement and expectations

An academically effective school, according to Purkey and Smith (1982), is distinguished by its culture: a structure, process and climate of values and norms that channel staff and students in the direction of successful teaching and learning. Students have maximum opportunities to learn, teachers and administrators have developed clear goals related to student achievement, and teachers, parents and administrators expect students to achieve. Such a culture, studies show, represents more than the sum of individual teacher expectations; individual efforts have become transformed into a common endeavor (Brookover et al., 1979; Grant, 1982; Rutter et al., 1979; Rutter, 1983; Cohen, 1983; Spady, 1982, Clark et al., 1984; Hallinger and Murphy, 1985a, 1985b; Purkey and Smith, 1983, 1985). Such findings are not limited to the United States. Studies in the Philippines and Malaysia (Avalos and Haddad, 1981), for example, also describe the positive effects on student achievement of this kind of school culture.

Such a culture, ethos or moral order (Grant, 1982) also affects the learning environment in another way. Students have difficulty learning in a distracting or unsafe environment. Studies show that where teachers and administrators can agree upon and be consistent and explicit about standards, student learning improves (Rutter et al., 1979; Lufler, 1978). Moreover, when such agreement exists, studies show that discipline policies become more than simply a punitive process--they become opportunities for organizational self-improvement. For example, a 1979 Phi Delta Kappa commission on schools with few or no discipline problems found disciplinary codes were the result of input from students, teachers and administrators; problems with student behavior, such as fighting, were seen as symptomatic of other problems and emphasis was placed on positive behaviors and preventive measures rather than formal rule enforcement or punishment programs (Lasley and Wayson, 1982).

Such a climate of teacher expectations and appropriate student behavior provides teachers with the maximum opportunity to create what they value most: a productive relationship with their students. In the United States a substantial body of literature shows that seeing the "light bulb" go on inside a student's head as he or she understands a concept constitutes one of the most powerful motivating forces for teachers (e.g., Lortie, 1975).

This is entirely understandable once one grasps the central difference between teaching and many other professions: teachers depend on students for their own success. As Cohen (1984) puts it:

> Now teaching is a trade in which one can only succeed if one's students succeed. It is not like baking bread, where most of the satisfaction lies in turning out a good loaf--even if another part is someone else's use of the bread. If one's students do not do well, or do not even want to do well how can a teacher do well? Or gain much satisfaction, or self-respect? (p. 12)

Intrinsic rewards for individual teachers that result from collectively pursuing a common endeavor and creating productive relationships with students are school level variables found in effective schools. But schools exist in a larger society and rewards established at that level for individual teachers often make it more difficult to create the kind of school culture just described. Consider, for example, the negative effects extrinsic rewards such as poor advancement potential and low salaries can have on individual teacher engagement and morale.

In the United States, the careerless nature of teaching discourages engagement or commitment. A uniform reward structure schedule based on seniority is largely insensitive to variations in talent and effort. Sykes (1983) argues that the lack of advancement opportunities coupled with the scarcity of rewards for excellence and effort create the following:

> The twenty-year veteran is indistinguishable from the neophyte and the typical salary schedule features a generally rising income slope, with the top-scale salary no more than double that at entry level. To advance, men move into administration or leave the profession altogether. Career mobility for women has been conspicuously absent . . . with the result that women favor an in-and-out pattern, an accommodation to the birth of children and the demands of motherhood (p. 110).

Salaries also affect individual teacher engagement. Although income potential in the United States appears not to be teaching's most potent recruiting device, those who attempt to account for the high defection rates place most of the blame on low salaries (cf. Goodlad, 1984). Wangberg, Metzger, and Levitov (1982), for example, found that most of the job dissatisfaction they had located among female elementary teachers was rooted in low, noncompetitive salaries.

In developing countries material rewards also affect teacher engagement and morale in a major way, as Hurst (1981) reminds us:

> Those who have worked in schools and colleges in developing countries will be well aware of the generally poor state of teacher morale. This has a number of causes. The rapid expansion of educational systems in recent years has led to the creation of whole armies of teachers, many of whom are not adequately trained for the jobs they are supposedly undertaking. In addition their pay is usually very low. . . . Their physical working conditions are usually very poor . . . and the profession has become a dumping ground for under-achievers at school, college and university level (p. 190).

Intrinsic and extrinsic rewards, moreover, are reciprocally related. For example, the amount of status teachers enjoy in adult society is an extrinsic reward. Poor esteem because of low salaries, however, is felt internally by individual teachers, often with devastating effects on morale. Industrialized as well as developing countries have sought to increase social esteem and, thereby, individual teacher morale by providing symbolic rewards to teachers (e.g., special teacher recognition ceremonies in the United States in the wake of critical reports such as A Nation at Risk and similar activities in the People's Republic of China after the Cultural Revolution and in Tanzania after independence).

It should be stressed, however, that describing certain school level characteristics or societal level reward structures is different from prescribing specific policies to affect either. Take school culture, for example. The literature on school innovation in the United States shows that administrative and bureaucratic decisions can put into place certain organizational and structural arrangements (e.g., clear educational goals, meetings for planning or model evaluation systems) which can facilitate changes in faculty attitudes and values. But this literature also cautions against prescriptive policy changes. It appears that the attitudes which create a productive school culture are in part derived from teachers' cumulative experiences from classroom interactions with students and in part derived from

collaboration and participatory decision making on a collegial basis over time (cf. Purkey and Smith, 1985). Hurst (1981) forcefully argues that the same conclusions apply to developing countries; his critical review examines efforts to improve the quality of education through top-down mandated change instead of collaborative efforts involving teachers as meaningful partners. The promise and limits of top-down change are discussed at greater length in Parts II and III of this essay.

I-B. Teacher knowledge and convictions about what to teach

Schools can be organized in ways which lead naturally to teachers learning more about subject matter content, students and effective instructional practice and to their sharing this knowledge with other teachers. Little (1981), for example, describes two kinds of work norms found in the unusually successful schools she studied. The first is collegiality, the notion that the work of teachers is shared, not to be done exclusively in the isolation of a classroom. Continuous improvement, the second work norm, reflects an expectation that improvements in teaching are continuous and life-long (rather than limited only to beginning teachers). In effective schools, teachers engage in continuous analysis, evolution and experimentation with instructional practices.

Schools where such norms are present and salient are characterized by frequent talk among teachers about the practice of teaching (instead of about the family background of students); frequent observations of teaching by teachers; and teachers working together to plan, design, research and prepare materials for teaching.

Since the teacher is the active agent through which most student learning takes place and serves as the basic resource for knowledge (in spite of

attempts in certain countries to develop teacher-proof curricula), such aspects of a school's culture are as critical to an effective school as are teacher expectations. How to go beyond describing such norms to actually creating them in less effective schools remains as problematic and difficult to achieve as efforts to promote higher teacher expectations.

Moreover, to make matters more complicated, societal factors affect individual teacher competencies which can make the task of creating productive work norms in a particular school even more difficult. Take the need for an adequate knowledge base as an example. Where teacher knowledge is deficient, concepts may be taught incorrectly, or more commonly, not at all. Under such circumstances, teachers are likely to avoid discussions that deal with content, since doing so would make their ignorance visible. This in turn makes it more difficult to build norms of collegiality and continuous improvement and makes even more salient the quality of teacher preparation programs.

In the United States, teacher education has been repeatedly and severely criticized for its miserable quality (Conant, 1963; Powell et al., 1985). In many developing countries a number of factors work together to prevent individual teachers from also developing an adequate knowledge base. Frequently, primary school teachers have had fewer than twelve years of schooling, the quality of which has necessarily been limited by lack of materials and adequate training on the part of their own teachers. Even where their background training has been adequate, it is frequently difficult for these teachers to obtain the sorts of materials which would allow them to keep up-to-date on current trends and developments in education. In many rural areas there is also a lack of access to current newspapers and journals, leading to broad geographical differences in levels of teacher knowledge. A further problem has resulted from the rapid expansion of access to primary

education. More and more, countries have been forced to relax credentialling requirements due to a simple lack of availability of the necessary personnel.

Attempts to upgrade less qualified teachers have often focused on methodology and classroom management. Beeby (1979) called for a strengthened role for school inspectors as one means of addressing classroom management deficiencies. But when the Ministry of Education in Kenya, for example, decided to implement a combination radio/correspondence in-service teacher training program, the emphasis was entirely on improving the general knowledge level of teachers.

> The main emphasis . . . was to be in-service training in the light of the urgent national need for teacher upgrading. The plan was not for training teachers in classroom methodology. It was aimed principally at upgrading their basic knowledge and general educational level, although there was always the possibility that teachers' methods would improve as a result of the examples placed before them in the unit's courses (Hawkridge, et al., 1982, p. 181).

A difficulty here, moreover, is how to determine the actual levels of teacher knowledge, as well as the overall distribution of teachers in need of upgrading. Most countries have generally relied on credentialling procedures for this purpose, rather than testing of teachers. There is a strong recent trend toward teacher testing in the United States, but it is unclear at present what the ultimate results of this emphasis will be. Clearly, any teacher testing which emphasizes a weeding out rather than an upgrading procedure would be counterproductive in countries where there is not a ready pool of qualified candidates.

Teachers' convictions about what is important to teach may also depend largely on their own level of general knowledge and personal experience. This is an area where more research is needed before any firm conclusions can be drawn, but we believe that teachers are more likely to emphasize those areas in

which they themselves have an interest. Other factors influencing teachers' convictions about what is important to teach are school level decisions on how to use available textbooks, school level cooperation between senior and junior teachers, influence of principals and inspectors, the way in which a school organizes teacher evaluation procedures, the way a school interprets national curriculum guidelines, and the way a school responds to external examinations (for related research on teachers' content decisions, cf. Porter et al., 1986; on teacher evaluation, cf. Darling-Hammond et al., 1983). A fruitful area of research may well be the interaction between each of these factors and teachers' general knowledge in influencing what is ultimately taught.

In order for upgrading programs to be successful, policymakers may well need to create stronger incentives for teacher participation. However, even where financial incentives are difficult to produce, other inducements such as increased job security, recognition, and improvement in working conditions may be offered at minimal cost. These could well be necessary if not sufficient conditions for teacher involvement in upgrading programs. A strategy to enable policymakers to become aware of current teacher constraints and needs could help them target policies to help teachers themselves feel a stake in personal and professional upgrading. These issues are discussed at greater length in Parts II and III of this essay.

I-C. Teacher use of instructional materials

Schools where teachers are engaged, hold high expectations for their students, and work productively with one another are also places where instructional materials can be put to effective use. This is important because instructional materials such as student textbooks, teacher guides, maps, chalk and audio-visual equipment are but one component of an overall package of

resource allocation priorities. Since the other components, such as teacher salaries and physical plant facilities tend to be more highly visible, instructional materials tend to rank low on the list of priorities. This situation, of itself, would perhaps be acceptable if it could be shown that instructional materials were of relatively less importance than other inputs. Recent evidence for both developed and developing countries suggests, however, that such is not the case.

For audio-visual equipment like television and projectors, the evidence of a strong positive effect on learning is limited. For textbooks, however, the evidence is quite clear. Heyneman et al. (1978) conclude that investments in textbooks are more likely to produce learning gains than investments in any other educational interventions. This effect appears to hold true for all countries but there may be a considerable amount of variation in the actual impact of textbooks depending on country-specific variables such as availability of other books and so on. Heyneman suggests that the impact of textbooks on student achievement appears to be much greater in developing countries than in the developed countries.

Jamison et al. (1981) have documented the impact of textbooks on mathematics achievement in Nicaragua. Heyneman, Jamison, and Montenegro (1984), examining an extremely well-organized textbook intervention in the Philippines, found that by improving the student-textbook ratio from 10:1 to 2:1 per subject in grades 1 and 2, student learning improved in a single year by an amount equal to between .18 and .51 standard deviations. Further improvement (to a 1:1 ratio) seemed to bring no further gains in achievement. These gains were achieved at the cost of approximately a 1% increase in per student annual costs. It also appeared that the textbook intervention resulted

in equity gains, with greater achievement for students of low socioeconomic backgrounds.

But textbook availability per se is not the key; it is the way they are used which makes the difference. Luna and his associates (1986) in their study of mathematics achievement in the Dominican Republic found schools in which modern textbooks were widely available but which showed no significant difference in achievement. Similarly, the ZIMSCI project in Zimbabwe found that availability of textbooks produced no change in achievement if teachers were not trained in their use (Cowden Chikombah--personal communication.) Data from Chile tell a similar story:

> Far from showing . . . that young, inexperienced teachers are more likely to use textbooks than those with more experience . . ., the reverse was true: less experienced teachers are less likely to use textbooks than those with more experience, and 78 percent of all the teachers in the survey expressed negative or ambivalent attitudes toward the use of textbooks (Schiefelbein, Farrell and Sepulveda-Stuardo, 1983, as summarized in Psacharopoulos and Woodhall, 1985).

The studies conducted by the International Association for the Evaluation of Educational Achievement (cf. Chang and Ruzicka, 1985) highlighted the existence of three different dimensions to any curriculum which help to account for the above findings.

a. The intended curriculum consists of what the educational planners of a country actually specify. It can be found in textbooks and in various teacher curriculum guides.

b. The implemented curriculum is actually taught in the schools. It includes teacher decisions about what is important, instructional methodology and student activities.

c. The attained curriculum consists of the actual achievement gains of students, both in purely cognitive ways such as those measured by tests and in affective ways, such as in attitudes toward learning.

These IEA studies found only a minimal correlation among these three dimensions. This finding points up the importance of close attention not only

to the allocation of instructional materials, but also to their subsequent use by school officials.

Textbook use, moreover, is related to textbook quality. Indeed, certain textbooks may actually encourage counterproductive teaching techniques. Luna et al. (1986) found that the mathematics textbooks used by most teachers in the Dominican Republic led to the teaching of mathematics "as a set of rules, a rigid discipline in which there is little place for creativity" (p. 9). A further problem which they identify is that the need for remediation is such that teachers are often unable to follow the curriculum guides they have been given. Such findings point out the importance of encouraging, at the school level, a collegiality among teachers, in this case with more knowledgeable ones helping others to gain a better understanding of subject matter content.

Textbooks, finally, are not the only form of instructional materials which teachers and students have at their disposal. The presence of a school library, for example, can compensate for the lack of classroom textbooks (Fuller, 1985), provided school level norms and values emphasize the importance of using such a room. Windham (1985) suggests that instructional support kits available at the school level for individual classrooms could be designed and made widely available at relatively low cost, not as an alternative to textbook distribution but simply to offer some help where textbooks are not available. He would include "a chalkboard and chalk, a set of maps, a dictionary, selected instructional charts (with an emphasis on issues of health, population, agriculture, and language), and any other supplies that materials specialists feel can be justified" (p. 68). Lassa (1983) calls for use of locally occurring instructional materials--weights and measures from the local marketplace, local songs and stories, traditional games and so on. Research is needed to determine the value of maps, health

posters, alphabet cards and many other instructional materials which may be important both for their contribution to cognitive learning and perhaps just as importantly, for brightening up the school environment and contributing to the joy of schooling.

Research on textbooks and other instructional materials shows how important their existence can be for learning to occur. But maximizing their use requires school level commitment, decision and organization, which do not happen in most schools. Policies might best be directed toward creating such norms than towards simply increasing the availability of text material.

I-D. Giving direction and assistance to teachers

Two key issues inevitably shape the question of providing assistance and direction to teachers and have been dealt with differentially in the literature: 1) what degree of autonomy should teachers enjoy? and 2) what is the principal's role? The second question will be dealt with first.

The U.S. literature has identified the principal as a key component of effective schooling. Two aspects of the principal's leadership role have been emphasized. First, the effective schools literature on staff and organizational development suggests that principals play a key role in the overall evolution of the school's climate (e.g., Brookover et al., 1979; Purkey and Smith, 1983; Rowan et al., 1983; Manasse, 1985). This leadership is exerted largely through the development and maintenance of school policies and norms. That is, the principal acts as a catalyst to help teachers arrive at a shared set of goals, to establish priorities and to reach a consensus regarding the importance of student academic achievement. He or she further helps to establish an orderly climate in which teachers are able to devote greater attention to the accomplishment of instructional goals (c.f. Firestone and

Wilson, 1985, for a discussion of how principals influence instruction through the cultural and bureaucratic linkages governing teacher behavior). Second, the literature on restructuring schools indicates that principals are important to the implementation and perpetuation of organizational change and new technologies. Both of these aspects of principal leadership behavior point to the importance of the principal's role as an instructional leader (Bossert et al., 1982; Rowan et al., 1983; Hallinger and Murphy, 1985b). A Third World case in point is provided by Somerset (1984), who points out that in a number of Kenyan schools, a sudden downturn in school effectiveness could be related directly to the loss of key head teachers.

Although there is widespread agreement on the importance of the principal, researchers have taken sharply differing approaches to the question of teacher autonomy. Saunders and Vulliamy (1983), discussing curriculum reform in Tanzania and Papua New Guinea, suggest that since what goes on in the classroom ultimately determines student learning, "teachers should be given clearly defined and detailed procedures for implementing the general principles of an educational innovation" (p. 361). They argue that unless teachers receive help developing new lesson content and new pedagogical skills, they will simply revert to whatever practices they were previously following.

Windham (1985) continues this line of reasoning, arguing in fact for an even more directive approach, with national curriculum committees preparing programmed teaching materials which the teacher would then follow. He sees this as requiring less supervision and management and believes that it should be generalizable throughout sub-Saharan Africa. It would seek to "strengthen the central government's policy and planning skills related to the educational system and to provide immediate and direct intervention into the operations of individual classrooms" (p. 64).

Hurst (1981), on the other hand, calls for a much more collaborative approach. He believes that teachers can and will respond to innovation if they perceive it to be appropriate. He calls for administrators to work with teachers, adjusting timetables, information flows and decision-making processes as necessary to strengthen the institutionalization of innovations.

A similar approach is taken by Beeby (1979), describing the potential benefits to schooling of collaboration between school inspectors and teachers in Indonesia. Beeby concludes that "no radical improvements in the methods of teaching in the primary schools will take place without the active cooperation of the supervisors" (p. 96). This, he explains, is because for the teacher and the headmaster alike, the inspector (or supervisor) is seen as an instrument of the central authority. He concludes, however, that without active encouragement from the educational hierarchy, the inspectors themselves will not feel the need for introducing (or at least giving encouragement to) more innovative teaching methods in the classrooms which they visit.

He suggests that an important initial step in assisting teachers is to identify where in the system new ideas and practices are most likely to be generated and then to use the inspectors as one link in the chain of implementation. He feels that the role of the inspectors must be less that of the formal authority figure and more that of one who is willing to enter into a professional partnership with the teachers.

A factor which deserves consideration here is the nature of teachers as a group. Teachers do not constitute a homogeneous body with the same (or perhaps even, similar) needs. Teacher aptitudes vary widely and initiatives for providing assistance to teachers must first provide a means of identifying the nature (and distribution patterns) of assistance needed and then be sensitive to individual differences. Exemplary teachers will need support of an entirely

different nature than will the more ineffective teachers. Rural teachers will most likely have needs which differ from those of urban teachers. Teachers in well-equipped schools will have needs which differ from those in more poorly equipped schools, and so on. This calls for more highly-targeted policy initiatives and points up the need for a realization of the importance of instructional management (whether by headmasters, head teachers, supervisors or others) while at the same time avoiding a mechanistic view of teachers as little more than tools to be manipulated.

In conclusion, differences in the ways in which teachers and headmasters are supervised and evaluated are important to the success of Third World schools. Inspectors and headmasters are generally most effective when they find a way of working with teachers that is attentive but not overly controlling. However, although the importance of leadership may be universal, the ways in which it is manifested can be presumed to vary widely in different cultures and settings.

I-E. Complementarity between the school and the communities in which it is located

Since the turn of the century, the "relevance" of education has been debated in the U.S. by such prominent educators as John Dewey, Robert M. Hutchins, W.E.B. DuBois and Booker T. Washington. This debate over how schools prepare or fail to prepare students to assume roles in the workforce has been important to less developed countries as well (see Heyneman, 1985a, for a brief history). Vocational programs in particular have experienced highs and lows in support. Heyneman (1985a) reports that three basic problems were identified with the "diversified" curricula suggested by the World Bank in the 1970's: they are complex and expensive; matching the curricula with the employment sector is difficult; and unsatisfactory outcomes are usually associated with

unrealistic employment objectives. Psacharopoulos and Loxley (1985) reached similar conclusions in a World Bank sponsored project evaluating curriculum diversification in Colombia and Tanzania. They caution that the diversified or vocation-oriented curricula may be desirable in some circumstances, but as previously practiced, they are expensive and difficult to implement. However, what is not addressed by Heyneman or Psacharopoulos and Loxley is the degree to which the social, political and economic interests of the nonelite groups are acknowledged and dealt with in the countries studied. In other words, when education reflects only the interests of the elite, the expansion of education and the attempt to make schooling more "relevant" to the masses, whether through the traditional academic or the diversified curricula, may experience serious problems. Therefore, in order to be responsive to the interests of the national population, we argue in this section that there must be complementarity between the way in which schools are organized and their local communities (cf., Dove, 1980).

What do we mean by complementarity and what do we mean by community? Community for this purpose can be defined on the basis of politics, social organization, economics, religion, ethnicity and language. It refers to the material interests and cognitive perspectives that groups of people hold in common. If people were members of a self-sufficient, homogeneous village, they might all be viewed as members of one and only one community. But fewer and fewer of the world's peoples live under such conditions. Many people living in close proximity belong to many different communities, when all the relevant economic, sociological, political, ethnic, religious and linguistic variables are taken into account. And any one person is likely to belong to not just one, but multiple communities, each differentiated from other communities in terms of economic interests, political views, religious beliefs, etc.

By complementarity we mean a balance of support and tension between the school and the various communities to which the students and their parents belong. In other words, an ineffective school might be so divorced from the community (in terms of failing to take into account local beliefs, aspirations and employment opportunities) as to be resisted or rejected by the students and parents. In fact, the research indicates that, to be effective, it is very important that the school be strongly supported by parents. At the same time the school should not be so supportive of traditional ways that possibilities for improvement in health, nutrition, productivity or social justice are precluded. In other words, complementarity does not imply narrow cultural congruence. Pollitt (1984) and others have pointed out, for instance, that school-feeding programs constitute not merely a legitimate function of schools, but one which may in fact be vital to the attainment of equity objectives. As will become apparent in the discussion below, the appropriate state of complementarity will differ from setting to setting and cannot be predicted or determined without knowledge of that setting.

This view of the nature of effective school organization is based on a growing body of anthropological research in education which, over the past twenty years, has been used to argue that the value of the knowledge and skills learned in school cannot be assessed without regard for their usage in everyday life. Furthermore, this research asserts that the acquisition of knowledge and skills in the school can only be fully understood when one takes into account the social organization of communities outside school (Erickson, 1984; Spindler [ed.], 1982; Ogbu, 1983).

Cognitive psychology also offers evidence of the linkage between the knowledge and skills learned in school and the contexts of their use in and out of school. This is a conception of thinking as a situation-specific operation

rather than a manifestation of general ability. Drawing from their research with the Vai of Liberia, Scribner and Cole (1981) apply this perspective to literacy:

> ... Literacy is not simply learning how to read and write a particular script but applying this knowledge for specific purposes in specific contexts of use (p. 236).

This notion of learning is not intended to imply that all learning is context dependent, but it does highlight the importance of social relations as an aspect of the learning process that must always be taken into account if one is to improve learning. Just as changes in the objects of learning change the learning task, so also a change in the relationship of the learner to the objects of learning profoundly alters the learning task and what in common parlance is referred to as the ability of the learner. That is, ability is conceived as socially constructed, as a matter of judgments arising from a specific set of social interactions and expectations and not as context independent and located inside the individual alone. Erickson (1984) gives a concrete illustration of this point in comparing in-school and out of-school competence in mathematics:

> From this point of view, it is not surprising that a child can display arithmetic competence while dealing with change at the grocery store and yet seem to lack that performance when doing what seems to be the "same" arithmetic problem on a worksheet or at the blackboard in the classroom, even if the problem were displayed using pictures of coins with which the child is familiar rather than using numerals with which the child might be less familiar. Still, a picture of a coin is not a coin, and relations with the teacher and fellow students are not the same as relations with a store clerk (when one has money) and one's little brother or friend (before whom one's display of appropriate performance carries no negative social and emotional consequences). The nature of the task in the store and in the classroom is very different and so is the nature of the abilities required to accomplish it (pp. 10-11).

Studies of face-to-face encounters as well as observations of cross-cultural classroom interactions have demonstrated that cultural differences can lead to interactional difficulties (teacher not understanding pupil, pupil not understanding teacher, etc.), which may become grounds for conflict which can then escalate and exacerbate cultural differences (Erickson, 1984). An illustration of successful avoidance of conflict is reported by Barnhardt (1982). The case involves an Alaskan Athabaskan native school in which the teachers, all native and lifelong residents of the village, implemented the State-mandated, standard curriculum in English, but used a slightly different pattern of social relations in their instruction that was compatible with the cultural patterns of interaction in the village.

Barnhardt attributes the dramatic increase in achievement, based upon standardized tests as compared to the performance of similar students in other village schools, to this effort to bring about an appropriate level of complementarity between the state-mandated, standard curriculum taught in a second language on the one hand and the delivery of instruction in a more culturally "appropriate" manner on the other. The teachers simply used means of maintaining social control that "made sense" from the perspective of both the teachers and the students. Barnhardt's research as well as that of others (e.g., Au and Mason, 1981; Erickson and Mohatt, 1982; Philips, 1982) demonstrates that even a slight change in the nature of social relations between the teacher and learner as illustrated by their co-membership in the community and their use of their cultural knowledge to teach in an appropriate manner may have a profound effect on the outcome.

In many countries the attraction of transnational cultural and economic forces competes with and often distorts the complementarity of local values and knowledge with modern values and knowledge. According to Sunkel and Fuenzalida

(1979), the "transnational community" is made up of people from different nations, but who have similar values, beliefs, ideas and a lingua franca--English. This transnational community shares a "transnational culture" which has two main components: specialized and common culture. The specialized component of the transnational culture takes the form of scientific and technological activities. This specialized knowledge is rooted in the belief that there are "scientific methods" which can be systematically applied to every aspect of reality. But in addition to this belief in the power of specialized, technological knowledge, members of the transnational community have remarkably similar patterns of beliefs, values and behavior in matters of family life, housing, consumption patterns, and other aspects of everyday life. The popularization of these consumption patterns beyond the transnational community, in many settings, can lead to the distortion of more beneficial patterns, not only in the case of consumer durables, but also in basic foods as when bread is substituted for maize or manioc (Barnet and Muller, 1974; Brown, 1974; Sauvant, 1976a, 1976b).

Sri Lanka provides an example of a situation in which educational planners tried to be more responsive to the mass of the population who did not share in the benefits of participation in this transnational community (Nairn, 1985; Little and Lewin, 1984). Policies were developed to replace the traditional academic curriculum, with its accompanying GCE O and A level examinations, which had given priority to attaining international acceptance of educational credentials for a small elite. In its place was a common comprehensive curriculum featuring prevocational studies and emphasizing education about the world of work, plus a new external examination--the National Certificate of General Education (NCGE). In order to give more students access to employment,

the new program was intended to reduce the privileges of the well-established, elitist schools.

The new common comprehensive curriculum remained in effect for five years before a new government reversed policies in 1977. Its demise is attributable to obstacles of both institutional capacity and attitudes. Capacity issues involved lack of resources and insufficient expertise in research and management for curriculum development, assessment, implementation and follow up. But even if there had been adequate resources, there was still attitudinal resistance on the part of different communities which were crucial to the successful implementation of the reform. These difficulties are discussed further in Part II, as examples of reform through regulation.

Each of the failures in Sri Lanka can be addressed in future reforms which attempt to ensure an appropriate degree of complementarity between community interests and perspectives on the one hand and the nature of school reform on the other. But success will depend on the extent to which policies in other sectors of society provide a setting which is responsive to the political, social and economic interests of nonelite parts of the population. At the grassroots level of teaching and learning, it is possible to envision a school organization in which a diversity of communities is taken into account. Educational systems attempting to separate themselves too much from the context of the community in which they operate not only work to the disadvantage of rural and marginalized peoples, but hinder national development as well.

I-F. Organizational properties that facilitate or inhibit school effectiveness

There are various ways in which the organizational properties of schools can facilitate or inhibit the five factors of school effectiveness discussed above. In discussing these possibilities, we need to start by considering three

propositions upon which this statement about the relation between school properties and effectiveness is based. First, primary schools in their prevailing form are organized as sets of classrooms. Second, intentional learning in these schools takes place largely through the intervention of teachers. Third, it follows that the effects of organizational factors are also largely mediated through classrooms and teachers. Organizational factors are not proximal or direct determinants of learning in the sense that teachers are.

Primary schools may range from schools with one room and one teacher to schools of many classrooms, but they all can be viewed as sets of classrooms. In the future, the extensive use of computer-based and other instructional technology is likely to lead to major changes in the organization of schools, but at present the primary schools of the world are organizationally more similar to each other than different. The widespread norm is one or more groups of children in a specially designated space, each group taught mostly or entirely by one adult.

By use of the word "classroom" we do not mean to imply a particular spatial configuration although it is true that something that might be called a room is found in at least some of the schools of every country. The significance of the term classroom is to be found elsewhere, however. It lies in the fact that a certain number of children are considered manageable and teachable by an adult or older adolescent and are therefore put together for that purpose under the close spatial and visual supervision of the teacher. It is this definition that conveys the sociological reality of the classroom.

Finally, teachers teach, but schools as such do not teach. The attempt to bring about intentional learning is the province of teachers. Dreeben and Barr (1983) have characterized school-level organization as follows:

> Contrary to conventional belief, schools are not organizational units of instruction. They are structures akin to switching yards where children within a given age range and from a designated geographical area are assigned to teachers who bring them into contact with approved learning materials specified as being roughly appropriate to age, during certain allotted periods of time. Schools deal in potentialities; they assemble a supply of teachers, of students, and of resources over a given period of time. Their central activities are the assignment of children to specific teachers, the allocation of learning materials to classrooms, and the arrangement of a schedule so that all children in the school can be allotted an appropriate amount of time to subjects in the curriculum (p. 83; cf. also Barr and Dreeben, 1983).

The organizational portrait drawn here is overly mechanistic, but it does serve to point out that organizational properties of schools per se are not likely to have direct effects on learning. The effects of such properties are more likely to be mediated through classrooms and teachers. That at least is our working hypothesis and as such is subject to empirical test. Moreover, even if one suggests that school level organizational factors can have an influence that is independent of classrooms, to estimate the size of this independent effect will require knowing how much of the total effect of any given organizational factor is mediated through classrooms and teachers. Otherwise, if the total effect is not disaggregated in this way, the independent effect is likely to be overestimated.

To take a simple example, Fuller's (1985) review suggests that school libraries have a positive effect on student outcomes. He reports that 15 studies show a positive effect and three no effect (none of the studies had negative effects). If we are to have adequate support for a causal interpretation of these findings, we need more information, for libraries may well have both an effect that is dependent on teachers and an effect that is independent. One hypothesis would be that library effectiveness is contingent on teachers providing direction and assistance in the use of the library. For

example, teachers might explain how the library is organized, describe its contents, give assignments in the library and evaluate students on their use of the library. A second hypothesis could claim that the effect of libraries depends on student inclinations and is independent of teachers' directions. Thus, students who have an interest in reading, who perceive reading to be useful, who are self-starters and who have time for reading might make much use of the library on their own. Hence, to understand the effect of libraries we would optimally like to know the size of the teacher-mediated effect as opposed to the teacher-independent effect.

If policy is made without this knowledge, policymakers will not know how much to invest in preparing teachers in the use of libraries. If the effectiveness of libraries is largely dependent on how well teachers are prepared to use libraries, the investments in libraries will not have a large payoff without investments in the preparation of teachers. On the other hand, if the library effect is largely independent of teachers, money invested in teacher preparation might not lead to much of a payoff and would perhaps be better invested in acquisition of libraries themselves. Still, one would have to ask if further investment in teacher preparation would raise library use and therefore increase the effects of the library among students whose inclinations to use libraries are currently low. In short, to make effective policy concerning school libraries, it is desirable to find or create conditions where not only availability of libraries varies, but also where teachers vary in the extent to which they guide students in the use of libraries and students vary in the extent to which they are naturally inclined to use the library.

There is still another pitfall to be addressed in interpreting the effects of libraries on student outcomes. Research is needed not only to help understand the conditions under which libraries have an effect, but also to

avoid attributing an effect to libraries when there is none. At least some of the effects reported in Fuller, for example, might be due to confounded variables. Let us say, for example, that library activity is correlated with the extent to which teachers guide students in the use of the library and that this teacher guidance is in turn correlated with higher teacher expectations for students in general. It is, therefore, possible that the apparent effects of libraries are due to high teacher expectations and that, if one controls for teacher expectations, libraries would have little or no independent effect (contrary to what Fuller suggests). This finding, too, would have implications for policymakers; in such a case, instead of spending money for libraries, they would be better advised to work at raising teacher expectations. In short, to make well-informed policies about libraries will require either studies of natural variation which account for potentially confounding variables or experimental designs which reduce such confounding through random assignment.

We therefore advocate examining the following sets of organizational properties in terms of their direct and indirect effects on student outcomes. The direct effect is represented by the arrow from B to D in Figure 1. The indirect effects are hypothesized to be mediated through the five school effectiveness factors: teacher engagement and expectations; teacher knowledge and convictions; teacher use of instructional materials; supervision and assistance to teachers; and complementarity with local communities. This perspective is not adequately represented in the present literature. Many of the studies reviewed (e.g., in Fuller) are misleading since, in the absence of adequate causal models and appropriate multilevel analyses, they are interpreted to mean that variables like school size directly affect student achievement (for a more detailed critique of Fuller, see Appendix One).

The organizational properties of schools are conceived as follows. Given the scarcity of relevant literature and its conceptual shortcomings, references to the literature are used for illustration only.

a. Teacher-student contact factors. Here we include size and time variables such as size of school, size of classes within school, number of classes per school, number of shifts, length of school day and hours of teaching assigned each teacher (i.e., direct student contact). These variables have to do with the teaching load of individual teachers, the amount of individual attention that students can potentially receive and the extent to which teachers and principals can be personally acquainted with students and their families. Hence they are likely to have effects on teacher expectations and engagement, teachers' opportunities to learn, teachers' use of instructional materials, the possibility of providing appropriate supervision and assistance to teachers as well as the nature of the teachers' relationships with community adults. For example, in the U.S., Rosenholtz (1986) found that teacher commitment and learning were significantly related to such school level characteristics as pupil-teacher ratios and school size.

These variables are likely to be highly interrelated and interactive in their effects. Take multiple shifts, for example. A Malaysian study (Beebout, 1972, cited by Fuller, 1985) reports double shifts at minimal loss of student learning. Two other studies found no negative effects on student achievement associated with multiple shifts. An intriguing study reported by Husen (1972) but largely overlooked since then, reported that children attending school every other day achieved just as well as those who attended every day. But the aim for unqualified generalization about the effects of different shift schedules is likely to prove fruitless. The question is not whether multiple shifts can be used without negative effects, but rather under what conditions

can multiple shifts be used to improve access to and quality of primary schools. Answering this question requires data on a number of variables. Multiple shifts might be seen as a way of increasing the number of students using a given school facility, but at the same time it could be used to reduce teacher-student ratios, to change the number of hours individual students spend in school, to allow two sets of teachers to use the same school or to reduce or extend the number of hours worked by individual teachers. Thus, the term multiple shifts stands for a complex of variables. To project the effects of a change to multiple shifts, it is necessary to specify these variables and then through careful research to estimate, insofar as possible, the independent and interactive effects of each variable.

Hypotheses concerning these possibilities can be easily generated. For example, if more than one set of teachers and students share the same facility, multiple shifts are not likely to have any negative effect. Also, if schools are hopelessly ineffective to begin with, changing to shifts is not likely to be detrimental. If the same teachers teach both shifts, then the effect is likely to depend on how long the shifts are and what the teacher-student ratio is within each shift. If multiple shifts involve reducing the amount of student time in school, the effect of multiple shifts can be expected to vary with the extent of this reduction. Such effects are not all likely to be linear. It would be interesting to know if a change from one to two shifts would have a <u>positive</u> effect (a) when the total number of children remains the same (extremely large class sizes being divided between two shifts), (b) a limited number of textbooks is shared across shifts and (c) the hours of school are reduced to keep the teachers' hours from being too long. Whatever the results of such experiments, the effects are likely to be mediated in part

through such factors as teacher engagement and expectations, teachers' use of instructional materials and teachers' opportunities to learn and improve.

b. Student progression and grouping factors. This category includes such variables as the age at which students are allowed to begin school, other school admission criteria, the criteria by which students pass from grade to grade and teacher assignment to classes by grade and subject matter. Hence the category includes organizational properties which make the student composition of classrooms more or less homogeneous in such characteristics as social background, achievement and age. Also included are questions such as the number of grade levels and subjects taught by individual teachers and whether they continue teaching the same cohort more than one year.

Within-class grouping does not appear to be widely practiced in most developing countries and tracking is a phenomenon mainly associated with secondary schools. The lack of within-class grouping may be due to the fact that diversified materials which would allow for grouping are not available. It is more likely, however, that grouping assumes a different and more extreme form in many countries, with lower-performing students being left to fend for themselves. As a result they are eliminated early from schooling. In many schools in rural Zaire, for example, well over 50% of the students in the lower secondary grades are failed each year (Prouty, unpublished, 1986). Attrition is one school level variable which will have a direct effect on student outcomes by restricting access to knowledge and indeed to schooling. It may also have indirect effects. Since teaching is a profession in which satisfaction derives largely from student performance, it can be assumed that teacher commitment could be seriously affected by these failure rates. This indirect effect has yet to be shown, however.

c. Staff hierarchy. This category has to do with assigned relations of subordination and supervisory control within the school staff and is particularly concerned with assignment of responsibilities to supervise and evaluate teachers. It involves such questions as these: Is there a principal, head teacher or inspector with delegated authority over teachers in the school? What are the duties of these persons? How much responsibility are they given for observation and evaluation of teachers? How are they selected and trained?

Variation in institutional definitions of the principal's role makes it difficult to generalize from the role of principals in the United States to the role of headmasters in developing countries. This is particularly the case when the principal's role can be defined in formal, official terms within the system but may be seen quite differently if the unofficial, informal aspects of the organization are taken into account. It should be noted that in some systems principals have nothing to say about such matters as curricula, testing of students and selection of teachers. In some systems, principals may have no teaching assignments whereas in others this may be seen as a critical component of their jobs.

In addition to these differences in definitions of the principal's role, there may also be differences which derive from variation in organizational conditions existing in the schools themselves. Hallinger and Murphy (1985b) have documented a number of differences in the way effective U.S. principals work, depending on whether the school has a majority of students of high or low socioeconomic status.

d. Access to information. How access to information is structured represents an important organizational property of schools. Important knowledge that is channeled to teachers and other staff members includes knowledge about the content to be taught, knowledge about methods of teaching,

knowledge about students, and knowledge about the community context and future opportunities for students. The channels include officially legitimated sources such as textbooks, teachers' guides, ministry publications, school libraries, radio and other electronic media as well as persons offically designated as authorities or trainers. Thus, the channels vary along several significant dimensions: some are based on text, some are not; some rely on face-to-face interaction, some do not; some bring about collective discussion (e.g., teachers who get together to listen to radio in-service), some are meant for use by individual teachers in the isolation of their classrooms. Some sources are heavily sanctioned and authoritative. In some francophone countries of Africa, for example, juries composed of teachers from other schools, plus an inspector, orally examine students and, in so doing, provide teachers with a chance to learn what content they ought to stress. The school's location will also have an important influence on these channels insofar as it will be more or less easy to reach the school in person, through electronic media or through distribution of written materials. Our framework leads us to expect that channels of information will have the greatest influence on student outcomes when they work to strengthen the collective culture of teachers in a school and thereby to increase teacher engagement, expectations and agreement on what to teach. We further expect that where there are one-teacher schools, the effectiveness of these schools could be increased by devising ways to increase the availability of relevant information and at the same time create a sense of shared understanding with other educators in the area, through visits or meetings.

e. Access to other physical resources. Although studies of industrialized countries have generally failed to document the importance of building quality and other facilities (e.g. Hanushek, 1981), Heyneman and others have challenged

the applicability of these findings to the Third World (Windham, 1985, is more skeptical). In our view, although there may be a threshold beyond which increased inputs no longer result in higher student achievement, building quality in developing countries generally fails to come up to this bare minimum required for good functioning. Fuller's (1985) review indicates that availability of desks and building quality in general have been shown to be correlated with achievement, but these studies do not explain why building quality enhances achievement. Here again we expect some of the effects to be mediated through teachers. For example, it can be presumed that teacher motivation will be higher in more appropriate surroundings and that they can more adequately care for instructional materials under good building conditions.

f. Community relationships. By community relationships we mean such variables as the location of the school, its status as a boarding or day school, the role assumed by parents in the operation of the school, and the rights and obligations of local communities concerning support and control of the school. It is within this web of relationships that the degree of complementarity discussed earlier is established. These relationships are also important in influencing whether the community supports or is hostile to the development of teacher engagement and expectations, changes in teacher knowledge and convictions about what to teach, and ways in which teachers use instructional materials. Boarding schools, for example, derive much of their influence over students from the fact that the school staff is able to enforce its culture on students with less interference from parents than might be the case in a day school. This potential power is particularly important when the the objectives of the school in some sense oppose the values of parents (for example, in advocating changes in the role and treatment of women).

PART II:

Strategies to Change Schools

Understanding the characteristics just discussed in no way guarantees that we can use this information to advantage in improving schools. In the first place, knowledge of factors associated with school effectiveness and their organizational concomitants does not necessarily identify the causes of effectiveness. For example, knowing that effective schools have committed teachers with high expectations does not guarantee that improving teacher commitment and expectations will transform an ineffective school into an effective one. Second, a successful strategy for change requires a sound understanding of organizational change as well as a theory for explaining school effectiveness. For example, to foster teacher commitment and positive expectations for student achievement requires a sophisticated understanding of how schools function and how policy can influence that functioning. To illustrate these points and the issues they pose, we discuss five strategies that policymakers in the U.S. or elsewhere have used to promote school change; these include policies to restructure schools, to regulate schools, to reallocate the resources available to schools, to deliver staff and organizational development to schools, and to enable teachers and headmasters to share more effectively what they already know. Our aim is briefly to assess the potential of each of these strategies for improving schools and to identify the issues that must be addressed if these strategies are to have the hoped-for effects.

II-A. Restructuring schools

Underlying conventional policy analysis on the relationship between educational policy, school organization, instructional practice and student outcomes is a longstanding debate about the rationality of schools as organizations. The debated attributes of rationality, posited in an ideal type for organizations in general, were well characterized by Corwin (1974) as including the following:

a. Clearcut goals are understood by organizational members and they are committed to them.

b. Organizational activities are planned.

c. The activities are closely coordinated.

d. Information is available as needed to make informed decisions for achieving the organization's goals.

e. Managers have sufficient control over the organization to ensure compliance with long-range plans.

Criticism of the applicability of this model to schools can be found in the literature over many decades. Currently, although many administrators and activists continue to look to rational models for guidance, this critical perspective constitutes the dominant scholarly point of view on schools as formal organizations. Critics have proposed alternative formulations in the form of loose coupling (Weick, 1976), garbage can theory (March and Olsen, 1976), hyper-rationality (Wise, 1979), institutionalized ritual (Meyer and Rowan, 1978; Rowan, 1981), interacting spheres (Hanson, 1981) and institutionalized bargaining (Sedlak et al., in press). These critiques have attempted to explain why top-down attempts to bring about school improvement have so often been ineffective and even counterproductive in the United States. This ineffectiveness and ways to avoid it have also been explored in the innovations literature (e.g., Rand change agent study as reported in Berman and McLaughlin, 1975-78).

At the same time policymakers have continued to try to reorganize schools to see if ineffective organizational characteristics can be eliminated. For example, in the United States there was a series of major school change experiments in the late 1960s and early 1970s. One of the most ambitious attempts to restructure schools to improve student outcomes was the Individually Guided Education (IGE) system developed and implemented by the University of Wisconsin Education R & D Center with funding from the U.S. federal government (Popkewitz, Tabachnick, and Wehlage, 1982). Under this program an organizational framework was developed to capitalize on the individualized instructional technologies then in vogue. The IGE model consisted of seven components. The first component was a school unit, the organizational building block of a school that consisted of a nongraded group of about 100 students, four teachers, a unit leader, aides and clerical staff. Each unit was given responsibility to carry out the second IGE component, which was its Instructional Programming Model.

This Instructional Programming Model consisted of setting objectives for the building, unit and individual students, pretesting students and giving them appropriate instruction, and finally assessing how well they had done and beginning a new cycle. Other components supported this instructional component with evaluation, instructional materials, development of home and community relations, development of networks of external agencies to assist in implementation and, finally, provision for continued research and development. The Wisconsin Center supported this model with much research on learning processes, organizational factors, and cost-effectiveness issues. In spite of this comprehensive effort, when Popkewitz et al. (1982) studied six IGE schools considered exemplary they still found, from school to school, much variation in basic conceptions of learning, teaching and social relations. In these

respects organizational change remained superficial, without major impact on the most fundamental aspects of schooling.

William Spady's theoretical and political efforts on behalf of outcome-based education provide a second example of reform through organizational change (1982). He asserts that if mastery learning is to be successfully implemented on a widespread scale, major organizational changes are entailed (e.g., change from age-grade structure to a time-flexible, continuous-progress system). Similar, though somewhat less far-reaching attempts at change have been the subject of research by Elizabeth Cohen et al. (1979) and others. These include individualized instruction, nongraded classrooms, and team teaching. All of this literature indicates that radical change with profound consequences for school effectiveness is possible, but that it requires extraordinary conditions and support. Through empirical research, Cohen and others (e.g., Berman and McLaughlin, 1975-78) have found that such reforms are fragile and vulnerable to loss of funding, staff turnover, opportunistic administrators, etc.

The structural changes that seem most readily able to achieve their intended aim are those which do not strongly challenge the existing practice of teachers and headmasters. For example, a current project in Zambia involves a change for small rural primary schools serving a low density population from a school with grades 1-4 and a separate class for each grade to a school with grades 1-7 and more than one grade per class. The intention is to increase class size and thereby reduce the number of teachers required without loss of instructional effectiveness. This experiment is partly funded by the Swedish International Development Agency and is in part inspired by Swedish experience with small rural schools with more than one grade per classroom (Ingemar Fagerlind, personal communication).

In another example of where marginal organizational change might be possible, Beeby (1979) suggests that too many primary schools in Indonesia were kept to a pattern of six classes (one per grade). With student attrition from grade to grade, this pattern was inefficient since it resulted in either lower primary teachers having more or upper primary teachers having fewer than the desired number of students in their classes. This could be avoided if parallel classes in the earlier grades fed into single classes in the higher grades. Beeby suggests that one reason individual schools have not had larger numbers of classes and teachers is that this would have reduced teacher chances for promotion to principal, which was the one nonseniority-based financial incentive for teachers.

In summary, the literature on organizational change suggests that top-down attempts to bring about major school restructuring are unlikely to achieve their intended objectives if they assume schools are organizations with a high degree of rationality. Substantial impact requires extraordinary conditions and support, whereas marginal impact seems much more feasible. If marginal impact is considered insufficient, other strategies for policy change may be needed.

II-B. Regulating schools

By school improvement through regulation we mean the practice of promulgating rules and guidelines for how schools should be organized and run, how headmasters, teachers and students should act and what they should accomplish. Some such regulation is part of most educational reform efforts, particularly in centralized systems, but in this section we are interested in the issue of whether regulation can be effective when largely or wholly unaccompanied by increased resources and technical assistance.

U.S. experience with enforcement of civil rights requirements shows that opportunities for students historically discriminated against can be significantly expanded through vigorous enforcement of federal regulations. For example, between 1965 and 1968 the dual system of segregated facilities in the South was dismantled as federal officials used the power created by the Civil Rights Act of 1964 (Title VI) to withhold federal funds from school districts that discriminated on the basis of race (Orfield, 1969). Court decisions between 1971 and 1973 (Swann and Keyes v. School District No. 1, Denver) approved busing as an acceptable means of desegregation and laid the basis for findings of de jure segregation in the North and for district-wide remedies (Kaestle and Smith, 1982). As a result, many cities and towns throughout the United States experienced fundamental changes in the organizational structure of their school systems as minority children began attending previously all-white schools.

The success in gaining access to previously all-white schools for minority children had its counterpart in gender for males and females through federal enforcement efforts of Title IX of the Education Amendments of 1972. Courses previously open only to one sex were made available to both young men and young women. Athletic opportunities were made more equal. Counselors were prohibited from steering students into traditionally stereotypic classes. These were a few of the more significant changes.

Handicapped students gained access to public school programs in a qualitatively new way as the result of a new federal program (The Education of All Handicapped Children's Act of 1975) and a new civil rights law (Section 504 of the Rehabilitation Act of 1973). School districts, under these laws, were required to find all eligible handicapped children in their district, evaluate their needs by developing an Individualized Education Plan (IEP) and educate

them to the maximum extent possible in the regular classroom (the Least Restrictive Environment).

While regulatory policies had the effect of creating changes in policies and practices at the school district and building levels for children with different characteristics, discriminatory practices at the classroom level proved less tractable than anticipated, in part because they were more difficult to reach in any effective way. In short, while students could be moved around, mingled, and evaluated, what really counted in the final analysis was their treatment in classrooms. Intended and unintended teacher bias, stereotypic curricular materials, peer group pressure and lack of coordination between classroom teachers and specialty teachers were some of the factors that led to differential access to content for these very students.

If the 1960s and 1970s reflected a national commitment to improving school responsibility for equal opportunity and equity, the "movement for excellence" in U.S. secondary schools became the major national agendum for the 1980s. To date this movement has relied on regulatory reforms at the state level. These reforms have concentrated on credentialling and curriculum, calling for graduation testing, increases in course requirements, changes in the length of the school year and day, testing of teachers and mandatory limits on class size.

Research indicates that this approach has at best produced mixed results. Greater organizational rationality has improved opportunities for youth already advantaged by the present system but has negatively affected students at risk. Existing programs for academically-talented youth have been strengthened and new initiatives funded, while curriculum changes for academically-marginal students have come to focus exclusively on competency items and drop-out rates have increased (Sedlak et al., 1986).

At the elementary level a ten-year program of research on how the content decisions of U.S. teachers are influenced by external factors provides another example indicating that certain types of regulatory policy can have a considerable, though not necessarily positive effect on teachers. Part of this research consisted of a three-year longitudinal study that closely examined the teaching of two teachers per school in three schools in each of six districts. The districts were systematically chosen from medium and large districts in a four-state region for their differences in district curriculum policy. The evolution of these policies was followed over the three-year period and policy strength was analyzed in terms of its prescriptiveness, consistency, authority and power. With the exception of one district which had virtually no district policies, each of the teachers was influenced by district policy. Policies which combined guidance on topics to be taught with testing and instructional materials tied to those topics were particularly powerful. In such cases, however, the policies would not have had their effect without the availability of the instructional materials and effective management strategies for making teachers aware of the policies. In the case of the most powerful policies, there were also perceived penalties associated with not following the policies (Schwille et al., 1986).

In the Third World, one of the most obvious and often used vehicles for regulatory reform has been changes in external examinations. Heyneman (1986) argues strongly for such testing while at the same time recognizing and discussing the dilemmas it poses. He asserts that testing, when accompanied by widely disseminated information on problems students experience with particular tests, can be and has been used as a national strategy to improve teaching. This argument in favor of testing is familiar and plausible, but empirical evidence is needed to document the effects.

In Sri Lanka, we have evidence a regulatory attempt to combine examination and curriculum reform backfired. As already mentioned, this reform was adopted in 1972 and rescinded in 1977. The reform included prescribing a new national secondary school leaving examination, raising the school entry age to six years, decreasing the total length of preuniversity schooling from twelve to eleven years, abolishing early selection in favor of a common curriculum up to the ninth year and providing compulsory prevocational studies during this period of schooling. Although these changes were intended to benefit the 98% of the population that did not go on to university study, public support for the reform was undermined by the following factors: (1) large numbers of unemployed graduates with prereform credentials were available and employers tended to favor them over the more recent graduates; (2) the middle and upper classes felt opportunities for overseas study were threatened and, as much as possible, made arrangements for their children to sit for the old exams outside the country; (3) in the short run, the better equipped urban schools were able to adapt more quickly to the changed curriculum, thus actually reducing the proportion of rural students gaining admission to further education; and (4) both urban and rural parents were dissatisfied with the prevocational emphasis while rural parents perceived their children's chances of social mobility were decreased. In addition to these economic and attitudinal reasons for the failure of reform, provision for its implementation was totally inadequate: appropriate teacher training was not provided; consideration of the nature and limitations of the teaching force did not play a prominent role in the planning process; consultation with parents was not sufficient to reconcile traditional views of schooling with the new program; and planning was not sufficient to ensure the curriculum was congruent with the needs of the workforce (Little and Lewin, 1984; Nairn, 1985).

In Indonesia, Beeby (1979) reported a similar gap between intentions and likely effects when he discussed the abolition of the external secondary school leaving exams in 1971. He warned that, as a result, the schools might either lower standards or be too much influenced by the entrance examinations used by different universities. He suggested that it might have been better to retain the secondary leaving examinations for a while and gradually increase the number of questions requiring more than rote recall.

These examples indicate that regulatory reform, even when accompanied by other measures to increase the likelihood of success, is still prone to unintended effects. Before undertaking such reforms, therefore, it is incumbent upon policymakers to understand the decision-making of persons subject to the reform so that the effects of the reform can be more accurately predicted rather than simply presumed. (See Part III for more discussion of this point).

II-C. Reallocating resources and providing incentives

The main question for this section is whether policy decisions about resource allocation alone can lead to improvement at the school level. This issue can be broken down into a number of different, though closely related questions. How much would increased resources help? Are increased resources likely to be available? If not, how much difference could reallocation of resources make? Which of various reallocation strategies would be politically feasible?

In the United States we have seen much research based on skepticism of the idea that student gains can be achieved solely through increased resources (higher teacher pay, more and better instructional materials, class-size reduction, etc.) targeted toward the needs of particular subpopulations (cf.,

for example, Hanushek, 1981). In general, the U.S. research suggests that, in the ranges studied, variation in resources per se has generally not been very effective in increasing student achievement as measured. One explanation for this (Cohen, 1983) may be that many input-output analyses start by measuring whether particular resources are available in a school, not by assessing how proficiently they are used by teachers and other staff to promote learning.

Heyneman and others have rightly challenged the applicability of this research to developing countries where variation in resources has a greater range and has often fallen below the minimum needed to support the desired levels of student outcomes (Heyneman, 1976; Heyneman & Loxley, 1982). In such situations, increases in supplies and other non-salary parts of a school's budget could result in increased retention and increased learning.

In many Third World countries, though, especially in sub-Saharan Africa, economic projections indicate that, for the coming years, there will not be increased resources for education (Heyneman, 1984; 1985b). Under such conditions, improvement in the quality of education may depend on reallocation of existing resources. In this section, therefore, we are concerned with policies which rely on reallocation of resources and changes in monetary incentives and which are minimally, if at all, concerned with restructuring organizations, regulating educational practice or providing special programs of staff or organizational development. Increasing skepticism about school improvement through restructuring or regulating has led to increasing interest in such policies, which range from changes in formulas for allocating funds to full-blown attempts to convert public schools to a market driven system (e.g., voucher systems).

In briefly considering how much difference reallocation of resources would make to school improvement, we consider three major categories of expenditure:

school facilities, instructional materials, and teacher salaries. In each of these areas, the prognosis is different:

a. Facilities. There is general agreement among researchers that some minimum quality of facilities is important to improving student outcomes. Beeby suggests that while good teaching can be done by competent teachers where facilities are poor, nevertheless the transition to a higher level of meaningful teaching is impossible unless teachers have extensive, safe storage for books and equipment, a place to leave class projects over night and so on (Heyneman and White, [eds.], 1986). This is a situation however, where there appear to be rapidly diminishing returns for expenditures beyond a certain threshold level.

Several studies have found facilities such as desks to be associated with student achievement (Fuller, 1985). At the same time, a number of researchers (Simmons and Alexander, 1978, for example) have reported finding no correlation between expenditures per pupil on facilities and student achievement. We could argue that this may well be because the effect of facilities quality depends on intervening variables such as instructional materials and teacher knowledge about how to make use of them (this relationship between inconsistency of findings and interaction effects is further discussed in our critique of Fuller--see Appendix One).

b. Instructional materials. Here the prospects for cost-to-outcomes ratios are still more favorable than what was just discussed for facilities. Available evidence suggests that textbooks are important to student achievement (see, especially, Jamison et al., 1981; Heyneman, Jamison, and Montenegro, 1984), and yet the share of instructional materials in education budgets has generally been very small compared to teacher salaries. In developing countries, the absolute amounts are minuscule. Heyneman (1984)

reports that whereas the U.S. has been spending over $200 per pupil in classroom supplies and Sweden over $300, the corresponding amounts range between one and five dollars per pupil in Malawi, Ivory Coast, Indonesia, Peru, Thailand and Brazil and less than one dollar in Bolivia. The shortage of these materials leads in turn to major inequities in their distribution. In the public urban schools of the Dominican Republic, only 19% of the eighth grade students had mathematics textbooks and in public rural schools 16%, whereas in the private schools sampled 63% of the students had these books (Luna, Gonzalez, and Wolfe, 1986).

This need for textbooks makes the financing of them a pressing issue. Should they be free or purchased by students? In Mexico the government has had a successful program to provide free textbooks to all children. On the basis of this experience, it has been estimated that developing countries can do the same by allocating about 2-4% of the education budget annually to textbooks (Neumann and Cunningham, 1982). By contrast, Liberia tried to finance textbooks through private sector publication and sales to students. A recent report concluded that "the two major goals of the program (1) to make the books readily available to all students and (2) to operate the program as a strict commercial, profit-making enterprise, are in some respects incompatible." (Modu, 1986, p. 22).

c. Salaries. Nwagwu (1981) reports that government efforts to improve salaries, promotions and fringe benefits between 1972 and 1977 were directly responsible for increased interest by males and females in considering primary school teaching. In this way the teaching profession was put in a position to compete for higher levels of talent. However, since teacher salaries comprise such a large part of the education budget (96% in Africa and 91% in Asia, according to Mingat and Tan, 1985), any nonnegligible increase in teacher

salaries would ordinarily require an increase in the overall education budget. That is, at present, unlikely in most developing countries. Reallocation from nonsalary parts of the budget will likewise not be feasible and, in light of what has already been discussed about the cost-effectiveness of textbook expenditures, likely to have a negative effect on student outcomes. As a result, there are two main reallocation issues involving teachers. First, given the different impact on student achievement and the different costs involved, it should be possible to raise student achievement by reducing expenditures on teacher salaries and increasing expenditures on instructional materials (this argument is made by Mingat and Tan, 1985, but note that their results are based on simulation, not empirical data).

Second, we may want to reallocate funds within the salary budget to change salary differentials and thereby improve incentives for effective performance by teachers and headmasters. The literature provides various illustrations of the lack of incentives due to lack of differentials in salary schedules. To be sure, secondary teachers earn more than primary teachers--up to five times more in some countries (Heyneman, 1985b), and there are differentials related to seniority and educational credentials. These differentials, though, may not stand up under the scrutiny of research on their cost-effectiveness, while still other differentials may be needed. Windham (1985) suggests that paying all teachers on the same salary scale works to the disadvantage of rural schools and school subjects that are marketable outside education, such as science and mathematics. He also notes that in some situations teacher trainees have had higher real incomes before than after they graduate when training stipends, free food and housing are taken into account. Beeby (1979) emphasizes the lack of salary incentives for superior performance by teachers in Indonesia. At times teachers were even unable to get the pay raises to

which they were entitled by reason of seniority and education. In Kenya, where correspondence and radio courses were being used effectively to provide in-service to teachers, declining enrollments in these courses were attributed to the fact that there was little financial incentive to participate; in this case salary increases were not automatic once the in-service certificate had been obtained (Hawkridge et al, 1982; Bude and Greenland, 1983, give further insight into the nature of financial incentives for participation in in-service in Africa).

In general, while the overall importance or lack of importance of facilities, instructional materials and salaries is becoming increasingly clear, we need to know more about their effects in particular settings. Before we can deal adequately with the issue of salary differentials and other incentive questions, we have to be able to estimate the effects of a particular mix of incentives in a particular setting. For example, in the specific countries, regions and types of schools for which policy is being made, it would help to know the incentives and disincentives for teacher attendance, teacher learning and effective use of instructional materials. Likewise, it is important to know the incentives and disincentives that affect the performance of headmasters and inspectors.

It is not enough, of course, to identify these incentives and disincentives. One must also ask whether policymakers have policy instruments to change them, whether the resources needed could be made available, and whether such changes would obtain sufficient political support to be implemented. These questions cannot be answered independently of other strategies for school improvement and will therefore be considered in Part III of this essay on modeling a mix of policies.

II-D. Prescribing organizational and staff development

The difficulties many developing countries experience in improving the quality of their rapidly expanding education sector can be traced in part to a failure to address organizational development, especially at the primary school level. Organizational development as it is usually defined seeks to increase effectiveness by using behavioral science knowledge to intervene in a planned way in the organization. Generally such intervention is managed from top levels in the system and has effects throughout the organization (see Beckhard, 1969, as cited in Dalin and Rust, 1983).

In their comprehensive review of school organizational development in industrialized societies (particularly European and North American societies), Dalin and Rust found that effective school improvement required that teachers and administrators be directly involved in needs assessment, goal setting and program development. These findings led to the development of a comprehensive set of procedures, now sponsored by International Movements Toward Educational Change (IMTEC), located in Norway, for organizational self-assessment and development involving steps ranging from initiation to data collection and analysis, setting objectives, implementation and evaluation. Survey instruments and procedures play a large role in this approach, but Dalin and Rust are careful to emphasize that these are only a part of a larger process. Their survey instrument (Guide to Institutional Learning) has been field tested in one hundred teacher training colleges, secondary schools and primary schools in five industrialized countries. While it could be adapted and used in developing countries, in its present form the approach is time consuming. For example, at least 20 working hours from all staff members are needed to complete the assessment phase alone (steps 1-6).

In constrast to organizational development with its more open-ended approach to goal setting, the kind of staff development described in this section is more directive. Teachers are trained in specific practices found to have improved student achievement. Such an approach, according to Sparks (1983), usually has the following components:

1. analysis of current staff teaching practices;
2. explanation and demonstration of new recommended practices;
3. discussion of possible applications, usually in small groups;
4. practice and feedback through microteaching, roleplaying and peer observation; and
5. coaching in the classroom where efforts are critiqued, suggestions for improvement are made and desirable practices are demonstrated.

This requires considerable time, effort and resources to implement successfully. To date, while staff development is commonplace throughout the United States, these requirements have not been met. Instead, as Lanier's extensive review (1986) has shown, staff development tends to be fragmented and without the depth described above (see also Griffin, 1983; Fuller, Noel and Malouf, 1985, explicitly compare the regulatory and staff development approaches).

Taylor's (1983) study of three teacher in-service programs in Africa, however, shows that in-service programs need not be so labor intensive (and, therefore, expensive), if more limited goals are sought. The programs he studied were distance in-service programs where teachers pursued studies through correspondence materials, group study, tutor visits and, in some cases, residential summer study. In the case of Kenya, radio was also used. Once opportunity costs of school time (lost earnings for the student or loss of teaching time for the school system) were included in a cost effectiveness analysis, teacher upgrading clearly emerged as a cheaper and more realistic

strategy for replacing or upgrading unqualified teachers than more intensive methods of training, such as full-time study at a teacher training college.

II-E. Engaging teachers and headmasters in reflection to share what they already know

Closely related to, but distinct from the organizational development approach that is prescriptive of procedures, and also from the staff development approach that is prescriptive of both process and outcome, is a strategy which relies still more heavily on explicating and sharing what teachers and headmasters already know about teaching and learning. Granted, this process may be more difficult to apply in some settings than others. It requires colleagues, who are not readily available to teachers in one-teacher schools. Yet it is perhaps more important for these teachers than for others that some formal support for information sharing be arranged to supplement the informal support that is lacking.

Also, the process will work better in schools where at least some of the teachers are particularly competent. Where none of the teachers are knowledgeable about a particular subject-matter, there will always be a need for inservice from outside. But we do not accept any claim that particular populations of teachers are completely without worthwhile knowledge and irrationally opposed to improvement. This view is similar to the discredited view that traditional farmers lacked the knowledge they needed to improve, in part, because they were tied to irrational traditional practices. But later it was realized that these farmers had good reasons for their practice, and that it was the adoption of new ideas that had not been made fully rational from the farmers' perspective (Hurst, 1981; Schultz, 1964). We presume that most teachers would be willing to change in directions which they would regard as improvements if the personal cost of such changes to them were not too great.

Hence teachers' reluctance to adopt new standard operating procedures should not be seen as recalcitrance or stupidity, but rather as the result of a view of classroom practice that differs in fundamental, qualitative ways from the views held by many educational planners, policymakers, curriculum developers, school administrators, and educational researchers (Campbell, 1981). School administrative procedures and policymaking have not provided a significant place for teachers' own accounts of their practice--how they think about what they do, what their students' test scores mean, what other outcomes of their teaching they can point to and why these outcomes are important (Schwille et al., 1983).

One result of this is a gap in perspective between teachers and administrators concerning the nature of daily life in classrooms and the implications of that life for the implementation of mandated policies and procedures. Bridging this gap is difficult. In order to utilize teachers' practical wisdom for school improvement, teachers must be helped to articulate and deepen their insights into practice, while administrators are made to familiarize themselves with teachers' bottom-up views of school policy and classroom practice. Understanding this situation is crucial to understanding the different and often competing organizational cultures that interact within schools and shape the nature of student outcomes.

An important reason for the neglect of teachers' views is clarified by recent research from the Institute for Research on Teaching which suggests that much of what teachers know about what they do is implicit (Erickson, 1984). The meanings of pedagogical action are locally specific to the immediate topics and people at hand (e.g., what to do next with this child, that reading group, this instructional topic, that child's conceptual difficulty, this child's socio-emotional context, etc.). Over the years, teachers' cumulative

experiences lead to implicit and explicit conceptions of relationships among children, subject matter content, instructional methods and classroom social arrangements (Yinger & Clark, 1983). In short, teachers acquire working hypotheses, as they generalize enough from their increasing knowledge to develop what anthropologists call "folk theories" and psychologists call "causal attributions."

In addition to their knowledge of practice being local and tacit, teachers also perceive patterns of relations as embedded in the context from which they draw their meaning. Teachers do not categorize information about children or classroom events in the way that is typical of much of the research on teaching (e.g., the cognitive and affective domains or the academic and social domains). Rather, a student's math achievement test score may be viewed in the context of how that child plays at recess and performs in reading and/or the child's home situation (Shultz, Florio & Erickson, 1982).

Although what teachers know about their practice is limited in important respects, this need not prevent taking the practitioner's knowledge seriously in its own right, nor preclude treating the limits of tacit knowledge as a starting point for developing its more explicit articulation by practitioners in the policy formation process (Navarro, 1985). However, the teacher's job as typically organized simply does not provide the time and resources necessary for this reflection, nor are schools organized to provide an audience for this information.

The role of the headmaster in developing an appropriate context for the emergence of the teachers' view seems crucial, yet it is also problematic given the demands placed on headmasters as middle level managers in the school hierarchy. While demands on the headmaster to be both administrator and teacher advocate is the bane of many instructional leaders, the development of

survival skills (e.g., code switching, managing different discourse strategies, familiarity with the community, etc.) can also facilitate greater receptivity to innovation and organizational change at the grass-roots level than would otherwise be possible (Navarro, Berkey and Minnick, 1986). Although most of the headmaster's day is spent mediating between individuals of different status, the role also entails opportunities to fundamentally change relations among actors in a school, but not without risk.

In short, this information sharing strategy entails (1) identifying teachers and headmasters whose practice is noteworthy (even if they cannot articulate exactly why this is so), (2) providing these teachers and headmasters with opportunities to deliberate in small groups with other teachers and headmasters on a regular basis, in particular, to observe and to provide nonjudgmental commentary on one another's teaching, (3) relief from other duties for sufficient time to be able to take on these other tasks, and (4) presenting the results of this reflective process to different audiences such as other teachers and headmasters, higher level administrators, parents, planners, and policymakers. Research has shown the importance of teachers' work-related views on policy implementation. If school improvement is to occur, then channels should be developed through which grassroots information, which only teachers can make available, can be effectively utilized in policymaking deliberations.

PART III:

Modeling an improved mix of school improvement policies

The preceding sections of this essay have addressed two fundamental questions: 1) What collective characteristics of schools, teachers and classrooms have been found predictive of school effectiveness? and 2) what strategies of intervention are available to policymakers and planners for changing schools in ways that are likely to enhance those characteristics? But even if one had certain knowledge of the causes of school effectiveness and available strategies for change, that would be insufficient for selecting among alternative policies. It is the purpose of this section of the essay to deal with the issues that remain and to provide a systematic approach for using knowledge yielded by earlier sections, however uncertain and incomplete that knowledge may be.

To do this, this section proceeds as follows: first, we introduce an example of an improvement in student outcomes that one might want to bring about in a Third World country and the issues raised by this example. Second, we use this example to discuss the importance of the choice of outcomes to be emphasized in schooling and why it is important to consider how the distribution of each outcome ought to appear if we are successful in improving schooling. This consideration requires, in effect, looking at more than one parameter in the distribution of each outcome. Third, we discuss the necessity of a multilevel approach to estimating the effects of policies designed to improve the outcomes of interest. Fourth, we show how costs can be taken into account in making comparisons of alternative policies, and how the

choice of outcomes calls for a certain type of cost analysis. Fifth, we argue it is necessary to frame these analyses in terms of an understanding of how decision-makers at each level of the educational system determine their course of action and that unless we do so we will not be able to implement policies that in the abstract appear logically straightforward and compelling. Finally, we show that models of how to increase the effectiveness of schooling must be better specified in the sense that they must include not only the variables susceptible to control by policymakers, but also other variables that must be taken into account if we are to estimate precisely the effects of policy.

Hypothetical outcomes and interventions. Much of the effective schools research is based on standardized achievement measures of reading and mathematics (but cf. Rutter et al., 1979; Rutter, 1983; and Heyneman, 1985b for consideration of other outcomes). Let us suppose that policymakers seek to raise achievement in reading and mathematics in the primary schools of a particular developing country, and have proposed three school level strategies to do this: 1) develop new instructional materials in reading and mathematics and train teachers in their use; 2) reduce class-size; 3) provide radio in-service for teachers (similar in part to the Kenyan example discussed above--cf. Bude and Greenland, 1983, for other examples of distance in-service). For the sake of this illustration, we will assume that we are dealing with a population of primary schools in a country where the national language of instruction is Spanish, French or English and where a substantial number of students enter school with limited or no exposure to this language.

Although this hypothetical example is simple, it poses a set of dilemmas for policymaking which are characteristic of the more complex problems in "real world" settings. First, consider the choice of outcomes. Although our attention is restricted to math and reading achievement, the relative

importance of the two is undetermined. Since different strategies for change may affect these outcomes differentially, part of the problem of choosing strategies for change involves evaluation of the relative importance of outcomes. Second, consider the problem of the social distribution of outcomes. Strategies for improving achievement of linguistic minorities may not be identical to strategies for improving achievement for those proficient in the language of instruction. Thus a crucial component of the policymaking process is to decide how achievement gains might optimally be distributed. Third, the hypothetical example poses three alternative strategies. To choose one strategy over the other implies an assumption about the causal relationship between those strategies and the outcomes. Fourth, the magnitude of the causal effects of each strategy must be weighed against the cost of implementing it.

Thus, even for our simple, hypothetical example, sound decision-making presupposes a framework within which goals and available means may be clarified and addressed systematically. Therefore, before we continue with the example, we introduce such a framework, which is multilevel in character. Policies are implemented at different levels of the social organization of schooling and are intended, in essence, to influence the social distribution of outcomes within those units. By implication, then, alternative policies are rooted in conceptual models for processes occurring within and between units. The framework we suggest requires that these conceptual models be made explicit.

Choosing, changing and realizing intended outcomes. In addressing the issue of how to improve schools, we see five questions that must be considered if improved policies are to be formulated:

1. What outcomes of schooling are most important as criteria for evaluating school effectiveness?

2. What is the optimal social distribution of those outcomes within a school, school district or country?

3. Given a set of manipulable variables which are assumed causally related to these outcomes, what is the magnitude of each effect on the distribution of the outcome?

4. What is the unit cost of changing the alternative causal variables?

5. What policies are most likely to bring about the intended results?

Together, questions (1) and (2) require that we specify the criteria by which education effectiveness is to be defined. In effect, these questions require that we identify the values to be maximized by policy. Together, questions (3) and (4) enable us to assess the efficiency of alternative strategies for realizing those values. That is, they tell us the expected gain for a unit increase in the causal variable and they tell us the cost of achieving that gain.

In effect, answers to the first four questions may be located within an "expectancy x value" framework. Thus, alternative policies may be weighed according to the expected gains they will yield across a number of outcomes, where the outcomes are associated with values.

Question (5) forces us to address the complex issue of policy implementation. Expected gains from policy can be realized only if unintended problems are held to a minimum and if policies are carried out in a fashion consistent with incentives that are important to grassroots decision-makers and stipulate means that are, in fact, available to local actors. We will present a way to take these considerations into account, providing information that may in turn affect the initial choice of policy.

We will start by examining the logical requirements for selecting criteria which define school effectiveness and for evaluating alternative strategies for increasing effectiveness so defined. The criteria consist of outcome variables and explicit notions about how those outcomes are optimally socially

distributed. Alternative strategies must be evaluated, both for their expected effects on the criteria and for the cost of achieving those effects.

III-A. Determining the criteria for evaluating school effectiveness

In evaluating any social program, one key decision is unmistakably value-driven: choice of the dependent variables or outcomes. In the mid-seventies in the U.S., for instance, controversy over regulation of pre-school programs focused on the "cost vs. quality" dimension. The debate reflected the varied goals of pre-school education: to enhance cognitive development in disadvantaged children, and to increase labor force participation of their mothers. The two goals appeared to conflict: those who emphasized the importance of maximizing cognitive gains advocated regulations which required small classes with highly trained teachers; those who wished to maximize labor force participation argued that such regulations would inflate costs of day care, and given fixed resources for the support of day care, such inflated costs would limit access. Thus, success according to the standard of quality might mean failure according to the standard of enabling impoverished mothers to go to work.

A similar "access vs. quality" debate is prominent in discussions of Third World educational policy. Fuller (1985) argues that the rapid expansion of education in Third World countries has undermined quality. He also cites literature which seems to indicate that improvements in school quality and academic achievement have bigger payoffs for Third World economies than do increases in access. Moreover, the payoff for increasing quality appears more pronounced for primary grades than for later grades.

Fuller's analysis suggests using academic achievement of primary schoolers as a key outcome rather than, say, the proportion of children completing ten years of schooling. Choice of outcomes is crucial because policies which appear effective with regard to one outcome may appear harmful with respect to another.

The social distribution of outcomes. The choice of outcome measures, however, does not eliminate the influence of values in shaping the criteria against which a program is to be judged. All outcome variables have distributions, distributions have parameters, and the choice of which parameters to study may have consequences which are as weighty as those of choosing the outcome itself. A parameter is a characteristic of the distribution of a variable. Examples of parameters are the mean, the standard deviation of variance, the proportion of children judged either maximally competent or unusually gifted. Many parameters of multivariate distribution are important substantively: the gap between majority and minority children or the strength of association between social class and achievement are examples. To illustrate how choice of parameters may influence the outcomes of an evaluation, consider a program designed to improve mathematics achievement. If the program minimizes the proportion of children failing to achieve minimum competency it might also limit the number of children achieving excellence. On the other hand, a "gifted and talented" mathematics program might exacerbate sex or racial differences in mathematics achievement.

Despite its consequences for shaping the evaluation, the choice of which parameters of the social distribution of an outcome to study is typically submerged by a tradition which makes study of the mean axiomatic. At times, the consequences of this tradition may be absurd. Some policymakers of the late 1950s sought to increase the number of mathematicians and scientists in

the United States. In the late sixties, other policymakers sought to reduce socioeconomic disparities in achievement. The tradition of evaluating schools by evaluating their mean achievement is equally irrelevant to both policy concerns.

The notion of specifying the distribution of outcomes as the criterion for evaluating school effects potentially enriches the class of issues accessible to policy analysis. For instance, Murnane (1975) examines the relationship between hours children spend in school and achievement in 875 inner city U.S. schools. Controlling for prior ability and socioeconomic background, he found important differences in the strength of association between hours and achievement, and concluded that schools influence the productivity of their students.

A second example illustrates how variance may be conceived as an "equity" measure. Some U.S. states have enacted legislation to achieve a more equitable distribution of resources as part of the "school finance reform" movement. To evaluate such a reform, Monk (1981) advised evaluators to utilize the variance of per pupil spending in school districts as the outcome. If successful, the policy ought to reduce variance.

The social distribution of the outcomes of schooling has been a prominent concern in the conceptualization of policy in Third World countries. Fuller (1985) identifies an enormous gap in achievement between urban vs. rural students in such countries as Egypt, Kenya, Peru, Brazil, and Haiti. According to him, the high level of absenteeism and massive grade repetition in rural areas have "massive" cost implications for Third World education. Thus, a major goal in some countries could be to reduce the gap between urban and rural school quality and achievement.

Fuller also reports evidence that in some developing nations, increases in achievement of poorer children provide proportionately greater economic payoffs. The implication is that policies which are especially beneficial to poor children will have a larger payoff than policies which simply improve average achievement.

In each case above, we have emphasized that certain parameters of achievement distribution may be viewed as the outcome measure. To make this notion concrete requires specification of the unit of analysis. Moreover, the conceptualization implies both a within-unit and a between-unit statistical model.

In the example from Murnane (1975), schools were evaluated for their productivity. Though this was not Murnane's approach, one might conceive of a within-school model relating hours spent to achievement, controlling for important student level variables. Between schools, productivity (association between hours spent and achievement) serves as the outcome.

In the school finance reform, the school district is the unit. Within districts, per pupil spending varies across schools. Between districts, variance in spending itself varies. Districts viewed as equitable are those which have small variance in per pupil spending.

In the case of urban vs. rural achievement, the unit may be a country. Within a country, achievement depends on urban vs. rural location and other factors. Between countries, the urban-rural gap serves as an outcome. One goal of educational policy is to reduce that gap.

To clarify the notion of using the social distribution of outcomes as a criterion for evaluating the effectiveness of schools, we consider examples in which parameters of the distribution of outcomes are crucial. Two examples come to mind.

First, Benjamin Bloom's (1984) goals for mastery learning are to increase mean achievement and to reduce variance. In this ideal setting, everyone benefits from the program, but those at the lower end of the outcome distribution profit most. This idea is depicted in Figure 2. Person A starts out at the bottom of the distribution and moves to A' as a result of the new curriculum. Person B starts out at the top of the distribution and moves to B'. Both subjects improve but the distance between A and A' is clearly greater than the distance between B and B'. It would be impossible to discover whether such a program achieved its intended effects without studying changes in both the mean and the variance.

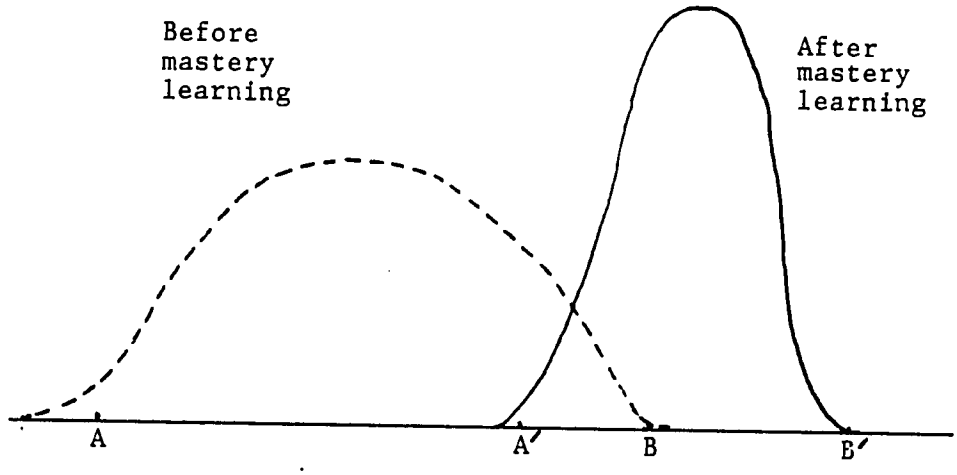

Figure 2: Intended effects of mastery learning

In the second example, James Coleman and his associates (1982) have asserted that Catholic schools have a different social distribution of achievement than do public schools, which enables them to approximate the American ideal of the "common school". In such a school, mean achievement is high, with social class background not strongly predictive of future achievement. Thus, the effect of social class on achievement is asserted to be

weaker in Catholic than in public schools. This idea is depicted in Figure 3. In statistical terminology, the figure depicts a social class-by-sector (public vs. Catholic) interaction such that the effect of social class is less pronounced in the Catholic sector than in the public sector.

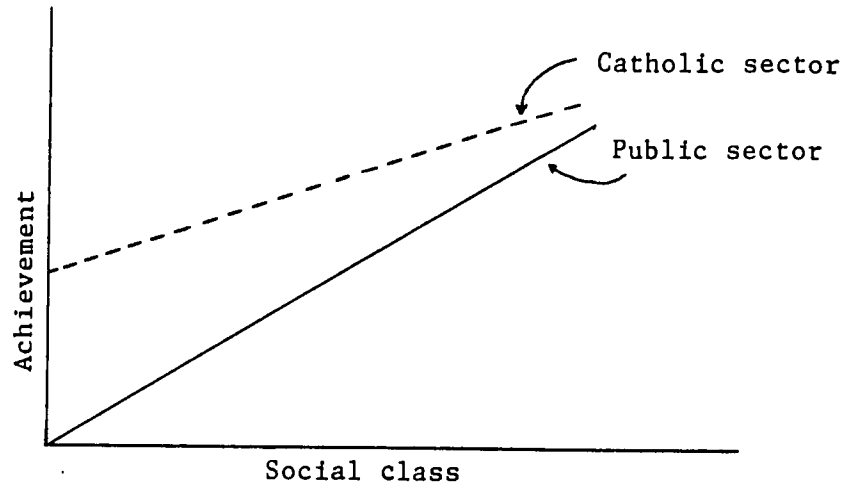

Figure 3: Relationship between social class and achievement in Catholic and public sectors.

III-B. A multilevel approach to estimating effects in the search for effective schools

The two examples above involve evaluation of two alternatives: mastery learning vs. no-mastery learning; Catholic vs. public schools. The search for effective schools and for the correlates of effectiveness is quite different in that there are as many "treatments" as there are schools. To discover by means of traditional methods which of 1000 schools, for instance, best achieve the "common school" ideal would be unwieldy. Figure 3 would become a tangle of 1000 lines describing the association between social class and achievement within each of 1000 schools. However, recently developed multilevel methods of statistical analysis now enable investigators to achieve the same goal without

covering large quantities of graph paper with black ink. The methods also have useful statistical properties which counteract the tendency for some fraction of schools in such studies to yield skimpy or untrustworthy data.*

The essential idea behind these developments is the explicit modeling of processes occurring within each school. In essence, each school may be viewed as having a production function which relates student background characteristics to outcomes. However, the parameters of each production function are presumed to vary across schools as a function of school policies, practices and organizations. One parameter of the school's production function is the mean, which is typically the focus of research. However, other parameters which characterize the social distribution of achievement may at times be equally important.

The multilevel model as developed in Raudenbush and Bryk (1986) is developed below in the context of our hypothetical example. The model postulates explicit models for processes occurring within schools and between schools.

Within-school model. Within each school, the achievement of a given student has two sources: the average achievement level in that student's school and the effect of that child's proficiency in the language of instruction. Thus, for each school two parameters are estimated: mean achievement and effect of language proficiency on achievement. The within-school model may be summarized as follows: achievement = school mean + language effect and effects of other student background variables. The two

*Appendix Two reviews for the interested reader recent methodological literature which has paved the way for improved statistical modeling of educational processes having a multilevel character. The appendix also provides a general statistical representation of the model.

parameters (school mean, language effect) summarize the social distribution of achievement in each school.

Between-school model. Between schools, each school's mean and language effect constitute outcome variables to be predicted by knowing about school policies and contexts. That is, both the school mean and the language effect are assumed to vary as a function of manipulable policy variables, such as availability of instructional materials, class size, use of radio in-service; and as a function of contextual variables. An important contextual variable in this instance might be the average level of the students' proficiency in the intructional language and the urban vs. rural character of the school.

Other levels. Multilevel statistical models provide a methodology for simultaneously examining educational effects at several levels of school organization, for instance, individuals, classrooms, schools, districts, countries, etc. Although it has been convenient for purposes of explication to restrict attention to two levels of organization (individuals, schools), an adequate representation of school effects would also incorporate classrooms. This is because research indicates that a substantial proportion of variation in school outcomes is between classrooms and because our theoretical model suggests that the efficacy of school improvement strategies is typically contingent on the response of teachers.

A multilevel model for evaluation and planning. The ideas in the previous section may be extended to a multilevel model for:

1. specifying outcomes;
2. specifying the parameters of social distribution of outcomes which are to be studied;
3. studying the effects of alternative policies on parameters, that is, determining the expectancy of success associated with policy options and;
4. specifying the cost of achieving the return.

The model may be used with schools, districts, provinces or countries as the grouping unit (see also discussion in Cronbach and Webb, 1975).

To estimate all the effects specified by such a multilevel model requires considerable information. However, when sufficient information is unavailable, the model may be useful as a heuristic for specifying key decision points in educational program evaluation or policy analysis.

Whether used as a method of statistical analysis or as a tool of conceptualization, it is critical to recognize several potential sources and consequences of its misspecification. By misspecification we refer to the failure of the model to incorporate key variables. Misspecification on the outcome side would occur, for instance, if policies which positively influence outcomes in the model had unanticipated negative consequences for important outcomes ignored by the model. Misspecification of causes would occur if the model failed to specify contingent conditions required for the success of a policy or if effects attributed to policy were actually consequences of causal variables not included in the model. Later sections of this essay explore methods of examining the adequacy of specification of causal models in the context of policymaking.

Application of the multilevel modeling approach to the hypothetical example. Let us continue the analysis of our hypothetical example by discussing the implications of this multilevel approach for the social distribution of outcomes. We have already said that we would concentrate on improving achievement in mathematics and reading as opposed to other outcomes that might be discussed, for instance, increasing average years of schooling or reducing rates of retention in grade. Even so we are faced with a situation in which more than one outcome is of interest and where no single test score or

parameter of a distribution will adequately represent the outcomes of interest. Moreover, the "distribution-of-outcomes" approach calls upon us to be more precise about what we mean by improvement of achievement. Earlier research and evaluation has generally operationalized improvement as an increase in mean achievement scores. However an increase in mean achievement could result from change processes which have different meanings for sub-groups of pupils (i.e. which involve statistical interactions). In the context of the example, mean achievement could go up under any of the following conditions:

1. Those fluent in the instructional language show improved achievement while achievement levels for other students remain the same;
2. Non-fluent children improve, others remain constant;
3. Everyone improves;
4. Gains of one group offset losses of the other.

Moreover, the failure of the program to improve mean achievement might reflect differential effectiveness rather than across-the-board failure.

Whether the sought-after improvement in mean achievement is a "good thing" depends on values. How important is the incorporation of the non-fluent groups in the educational mainstream of the country--as opposed to the rapid advancement of the fluent group? The observed mean achievement gains (or the lack thereof) cannot be pronounced a success (or failure) until values concerning the distribution of achievement are clarified and until analysis provides a more fine-grained sense of how policy shifts the entire <u>distribution</u> of outcomes.

For the moment, assume that we want both to bring about an increase in mean achievement of each school and, in the interests of equity within each school, to decrease the effect of prior knowledge of the language of instruction on the subsequent outcomes. Therefore, for each school we must consider four outcomes of interest, the mean achievement in reading and mathematics and the

regression-estimated effect of prior knowledge of the language of instruction on each of these two subjects.

Next we are ready to estimate the effect of the three strategies for improving mathematics and reading: (1) provision of new materials and training in their use; (2) reduction in class size; (3) provision of radio in-service for teachers. Assume we have no direct evidence of how these strategies might improve student achievement in mathematics and reading or affect the relationship between prior knowledge of the language of instruction and mean achievement. We can still, however, make estimates based on logical deductions or on extrapolations from studies done elsewhere. We might then conclude that provision of new materials and training in their use is more effective than the other strategies in raising mean achievement in mathematics and reading, but is not likely to be effective in reducing the relationship between prior knowledge of the language of instruction and achievement. That is, in both reading and mathematics a child's prior knowledge of the language of instruction will probably continue to be a very good predictor of achievement and children from homes where there is no familiarity with the national language of instruction continue to be at risk in the school setting. On the other hand, a major decrease in class size is not likely to be so effective in raising mean achievement, but we might hypothesize that it does reduce the effect of prior knowledge of the national language of instruction since the teacher could have more time to devote to developing the skills of children who have little prior knowledge of this school language.

Deciding whether to adopt one of these strategies requires still more information, however. Which strategy would provide the largest outcome per unit cost? Will it be sufficiently congruent with the way decision-makers at various levels of the school system make decisions? Do the data on effects

leave out important variables that, if included, would change our predictions about these effects? We take up each of these issues in turn, starting with cost.

III-C. Examining cost from effectiveness, utility and feasibility perspectives

Methods of cost analysis. Given the dearth of resources facing schools in many Third World countries, it is important to optimize the use of the available resources to achieve desired improvements of schools in these countries. In this context, cost analysis is a useful methodology that helps decision-makers in the efficient allocation of resources to achieve given goals. The following presents a brief overview of this methodology and continues the analysis of our hypothetical example to illustrate the issues to be considered in conducting a cost analysis.

Cost analysis is a methodology for evaluation that takes into account both the costs and outcomes of selecting alternatives, making it possible to choose those alternatives that provide the best results for given resources outlay or that minimize the resource utilization for a given outcome (Levin 1983).

The basic procedure for cost analysis consists of (1) identifying alternative strategies for achieving a given objective, (2) determining the costs and outcomes of each alternative, and (3) comparing outcome per unit cost (or cost per unit outcome) for the alternatives.

A cost analysis begins with an identification of interventions or strategies that can achieve a given objective. If a relevant intervention is omitted, the result of the cost analysis may be invalid. Also, if no alternative interventions are identified, there is no need to conduct a cost analysis.

Next, the cost of an intervention can be determined using a simple and logical method called the ingredients method. According to this method, the ingredients used in the intervention are first identified. The ingredients usually fall into one or more of the following categories: personnel, facilities, equipment and materials, other program inputs, and client inputs. The cost of each ingredient is then determined. In cost analysis, the cost of an ingredient is its opportunity cost, that is, the cost incurred as a result of the ingredient being used in the given intervention and thus not available for use in alternative activities. It is defined as the worth of the ingredient in its best alternative use. Finally, the source of financial support for each ingredient is also specified. Using this method, the total cost of the intervention as well as the distribution of the cost burden can be determined.

The outcomes of an intervention can be measured as benefits, effects, or utility, depending on the objectives to be achieved. Benefits refer to those outcomes of an intervention that can be assessed in monetary terms. For example, a benefit of a training-for-employment program can be measured to be the additional lifetime earnings of participants of the program. Effects refer to those outcomes on an intervention that can be assessed in their own attributes. For example, an effect of a remedial mathematics program can be measured to be the additional gain in test score for mathematics for participants in the program. Utility refers to those outcomes of an intervention that are measured in terms of their subjective value to a key decision-maker. For example, in deciding which programs to eliminate to achieve a given budget cut, the key administrator can show the utility of the programs by assigning a utility score to each program based on his or her own liking. The measurement of outcomes of alternative interventions often

constitutes the major work of an evaluation. And depending on the nature of the outcomes of alternatives, there are different approaches to cost analysis. They include cost-benefit (CB) analysis, cost-effectiveness (CE) analysis, and cost-utility (CU) analysis.

CB analysis refers to the evaluation of alternatives by comparing their costs and benefits when each is measured in monetary terms. Since both outcomes and costs are in monetary terms, this approach has the advantages of (1) identifying worthwhile interventions for which benefits exceed costs, (2) ranking interventions in terms of cost-benefit ratios, and (3) comparing interventions in different areas (e.g., education, health, transportation) or interventions for different goals to assess their relative returns to investment. The major disadvantage is that it is often difficult to measure the outcome(s) of an intervention in monetary terms. In education, CB analysis can be used on matters concerning the "external efficiency" of schooling such as maximizing the economic returns to investment in alternative education and training programs (Thias and Carnoy, 1972; Hu, Lee, and Stromsdorfer, 1971).

CE analysis refers to the evaluation of alternatives according to both costs and their effects with regard to producing some defined outcome. It can only be applied to evaluating alternatives with a common goal and a common measure of effect. However, CE analysis is a widely practised approach because (1) the measures of educational effectiveness can be those which a decision-maker will normally consider (e.g., improvement in student test scores), (2) CE analysis builds on a standard approach to evaluation by adding a cost dimension to the overall evaluation design, and (3) CE analysis is less costly than CB analysis in that the outcome of an intervention does not have to be expressed in monetary terms. In education, CE analysis can be used on matters concerning the "internal efficiency" of schooling such as improving the

quality of learning (Levin, Glass, and Meister, 1984; Mayo, McAnany, and Klees, 1975).

CU analysis refers to the evaluation of alternatives according to a comparison of their costs and the estimated utility or value of their outcomes. CU analysis is appropriate when subjective assessments must be made about the nature of outcomes as well as their relative values. CB and CE analyses are usually time consuming (and costly too). The advantage of CU analysis lies in its flexibility to deal with a range of alternative interventions in a relatively short time, without stringent requirements on the measurement of outcomes. But since the approach is subjective, its results are usually not replicable.

In addition to the three approaches discussed above, there is a fourth approach to cost analysis, cost feasibility (CF) analysis. CF analysis evaluates alternatives for their costs only to determine which alternatives are feasible under a given budget constraint. Since CF analysis ignores outcomes, it cannot be used to choose among alternatives that are feasible.

Finally, the computation of cost/outcome ratios (C/B, C/E, or C/U) for alternative interventions will produce a ranking for the interventions. The ranking can serve as a source of information to inform decisions, and it should be combined with other relevant considerations (ease of implementation, distribution of costs and outcomes among different target groups, etc.).

Cost considerations applied to our hypothetical example. To illustrate the issues to be considered in conducting a cost analysis, let us return to our hypothetical example. Since we are not prepared to express student achievement in reading and mathematics in monetary terms, we cannot do a cost-benefit analysis. Moreover, if we wish to do a cost effectiveness analysis we must choose to focus on one common outcome. For cost effectiveness analysis,

shifting from one objective to another usually requires more than one cost analysis. Thus, we must decide whether to focus on (1) increase in mean mathematics achievement, (2) increase in mean reading achievement, (3) decrease in the effect of prior knowledge of language of instruction on mathematics achievement, or (4) decrease in the effect of prior knowledge of language of instruction on reading achievment. This choice is particularly problematic in the case of school level interventions such as reduction in class size (within primary schools) which are meant to impact on more than one subject matter area. Nevertheless, for the sake of conceptual clarity in explaining cost-effectiveness analysis, we will at this point focus solely on increases in mean mathematics achievement. We will discuss later how cost-utility analysis can be used to handle the multiple-outcome situation.

Given the objective to be achieved and having identified the three strategies discussed above, we can proceed to determine the costs and effects of the strategies.

The ingredients method is used for cost estimation. For each of the three interventions, this involves a careful specification of all the ingredients involved in the intervention, an estimation of the cost of each ingredient, as well as an identification of the source of financial support for each ingredient. To provide a basis for comparison, one can estimate the total cost of each strategy for a class of primary-school students (say, 45 students per class) in one year of instruction. The total cost of the instruction-material strategy consists of the annualized costs for a teacher (salary and fringe benefits), a classroom (including utilities and maintenance), classroom furnishings, instructional materials, as well as training in the use of instructional materials. The strategy that reduces class size requires additional teachers, classrooms, and furnishings. The total cost of this

strategy is the total annualized cost for the additional resources needed in reducing class size (say, from 45 students per class to 30 students per class) per subject. The total cost for the radio in-service training for teachers involves annual expenses for obtaining broadcasting time on national or local radio, an instructor, a classroom and its furnishings, as well as broadcasting equipment such as radios (see Taylor, 1983, for a similar example). It is obvious that the omission of any relevant ingredient, whether it is common to all strategies (for example, a classroom) or specific to a strategy (for example, radios for in-service teacher training) may bias the results of a cost analysis. It is also important to identify the source of financial support for each ingredient. For example, if broadcasting equipment is donated to a school, the cost of the in-service training strategy will be greatly reduced for the school while the cost will remain the same for other schools not having free broadcasting equipment.

Consider next the effects of the strategies. Here we have to determine the effect of a one-year treatment of each strategy on mean mathematics achievement. As our common measure of effects, we will use a mathematics test score from a test designed to cover the primary school curriculum in the country of interest. The effect of a strategy can be expressed in terms of the standardized gain in test score. It is estimated from data from an experiment or a study of natural variation discussed previously.

Having determined the costs and effects of the three strategies, their effect-to-cost ratios (effect per dollar) can be computed to produce a ranking. This ranking is a useful piece of information for decision-makers. In general, a difference of ten percent in magnitude between any two ratios is considered small and insignificant; a difference of 100 percent or more is very significant and should not be ignored. The ranking should also be combined with

other relevant considerations to inform decisions such as cost feasibility, and ease of implementation. For example, the strategy that reduces class size may be the most cost-effective among all alternatives, but its implementation may be hindered by an acute shortage of well-trained mathematics teachers. Also the total cost for the drastic reduction in class size may be so high for Third World schools that the strategy is not feasible. The in-service teacher training strategy may be the least cost-effective of the three strategies; it may, however, become more attractive for schools that receive free broadcasting equipment. The strategy that provides instructional materials has to take into account the time and publishing requirements that are needed to develop and produce culturally relevant instructional materials. Finally, if the strategies affect students of different social backgrounds differently, the political ramifications of the strategies have to be taken into account too.

The above discussion on cost-effectiveness analysis applies only to a common measure of outcome. If it is desired to provide a cost analysis of the four outcomes discussed above (increases in the two achievement means and decreases in the effects of prior knowledge of language of instruction on achievement in mathematics and reading), it will be necessary to employ a cost-utility approach. The policymakers and other significant stakeholders in education in the country concerned would need to be called upon to state the relative importance of the four outcomes. These judgments of importance are used in conjunction with estimates of the effects for the four outcomes and the costs of the three strategies to compute a combined valued-based outcome (utility) to cost ratio for each of the three strategies. The cost-utility ranking can then be used to inform decisions, with consideration given to other relevant factors. Since different countries may have a different assessment of

the relative importance of the outcomes, the results of the cost-utility analysis are usually not replicable for other countries.

In sum, cost analysis provides a framework to optimize desired outcomes of schooling for given resource outlay. It makes explicit the importance of identifying all relevant alternatives for achieving given school objectives, and it takes both the outcomes and costs into consideration in evaluating the alternatives. When combined with careful consideration of other relevant factors, it is a useful methodology to promote efficiency in school operation. Its application is particularly useful for many Third World schools, in light of the scarcity of financial resources facing these schools. It is possible that certain student outcomes can be increased through reallocation of school resources so little or no increase in the school budget is needed.

III-D. Backward mapping to guard against unwarranted assumptions about how various levels of decision-makers will respond to policy

Up to this point different approaches have been discussed from a conventional policy perspective that assumes policymakers at central levels of decision-making can identify and specify key variables. Policy options are then developed, expected outcomes weighed and the most promising option chosen. Richard Elmore, in a series of papers (1979; 1983; 1984), has demonstrated, however, that this procedure is logically incomplete; the policymaker has only established a hypothetical cause-and-effect relationship between a policy strategy and an expected effect. What has not been done is to reverse the logic and assess the cause-and-effect relationship from the perspectives of those who have to implement the policies (teachers and administrators) and their clients (students). Elmore argues that the policy analyst would end up choosing a better mix of policy options if he/she also reasons backward, starting with how decisions at the grassroots level get made, what factors

influence those decisions, and what means policymakers have to affect such decisions. This "backward mapping" strategy not only helps the policy analyst identify unanticipated negative effects of policy, but can also lead, according to Elmore, to different conclusions about the set of policies most likely to bring about intended results. The "backward mapping" approach is an especially appropriate strategy to analyze the complexities of education. It takes into account the following attributes which produce this complexity.

a. Schooling has multiple outcomes, which do not lend themselves to any simple maximizing strategy.

b. Each category of personnel involved in the production of these outcomes (students, teachers, principals) is heterogeneous enough to respond in diverse ways to external conditions.

c. Various conditions need to be taken into account (e.g. student entry capabilities, student home background, community social structure, peer groups, school norms, labor markets, migration patterns etc.) if one is to understand the processes by which schooling leads to desirable (and undesirable) outcomes.

d. These conditions (even in socialist countries) are but partly under the control of government policy.

When accurate knowledge is available on how students, teachers and principals view their work and make decisions, it is possible for educational policies to influence these decisions in ways that lead to school improvement. Conventional forward policy analysis is particularly handicapped because it neglects this kind of knowledge which means the likelihood of missing important variation in outcomes, target populations, and associated conditions is great. Backward mapping allows for more systematic consideration of such variation and its consequences. Again following Elmore, this can be done systematically by addressing the following questions in the order listed.

a. What grassroots decisions are critical to the attainment of certain desired outcomes (e.g., certain specific improvements in education, health, agriculture, etc.)? Who are the makers of these decisions?

b. In particular, what do central government policymakers want to result from these decisions?

c. What other external conditions influence these decisions?

d. What are the agencies that could be charged with implementation of government policies in this area and what could they do to promote the desired outcomes (with the least adverse effect on attainment of their other objectives)?

e. What tools are available to central government policymakers to promote the desired outcomes through designated implementing agencies?

To illustrate how a conventional (or "forward mapping") policy perspective might produce a policy that could fail to achieve its intended goal, let us hypothetically examine what might happen to the policy of providing radio in-service for teachers. Then we will show how a "backward mapping" approach might have helped policymakers anticipate problems that occurred and thereby have encouraged them to consider more seriously other options.

In discussing this approach, we will assume that policymakers used evidence from the Kenyan example discussed earlier that showed a government sponsored radio/correspondence in-service program increased the general knowledge level of teachers with resulting increases in mean student achievement scores. policymakers in the country under study, we will assume, drew the inference that a radio program targeted specifically on mathematics and reading might have a similar effect in their country. Since no evidence was available on how the program affected the relationship between prior knowledge of language of instruction and achievement, this issue was not analyzed further. Because this country received free broadcasting equipment from another country, cost effectiveness analysis produced a rank ordering that prioritized the radio in-service over a policy to provide new materials and instruction in their use or a policy to reduce class size. Policymakers in the Ministry of Education then succeeded in gaining Cabinet approval and subsequent legislative support for the radio in-service initiative.

It took only a couple of months to see that the initiative was in serious trouble. Monitoring teams from the Ministry of Education found the programs had not been widely listened to and that there was considerable teacher hostility among those who had listened. Further discussions at the central level led to a set of explanations for failure that focused on the teachers; their antipathy to new ideas and their fear of being shown up as incompetent. But what really happened? In our hypothetical example, the causes were in part organizational. As it turned out schedules of when the programs were to be aired usually got lost; radio sets were often not in running order; and reception had been poor.

Had policymakers focused on how teachers, students and principals viewed their work as a part of their initial decision-making, they would have discovered that such problems were likely to occur. To illustrate this let us show what could have happened had policymakers proceeded step-by-step through the backward mapping process. As will become clear below, such a procedure requires central policymakers to have an intimate, usually first-hand, knowledge of local conditions.

a. Identification of target populations and grassroots decisions.

Teachers were the target population for this initiative. Organizational decisions, not teacher attitudes, were the key to problems that emerged and these organizational decisions depended on the location of the school and its size. Urban schools had many fewer problems receiving schedules than did rural schools, for example. Radio sets were generally in better repair in the urban areas because parts were more accessible, as were trained service technicians. Finally, reception proved better in urban areas because they were not surrounded by mountains like many of the rural areas.

Had policymakers at the central level involved a sample of schools in their initial discussion about such a policy, they likely would have heard of some of the organizational pitfalls that lay in wait from those who have to implement decisions. Moreover, they probably would have seen that it was generally not best to approach the target population (teachers) as if they were a homogeneous category uniformly influenced by standardized policies. Clearly the organizational context of the school where teachers worked (urban-rural) affected how teachers responded and ultimately the success or failure of the policy itself.

b. Outcomes desired by policymakers.

In discussing our hypothetical example, we have already given much attention to the primary interest policymakers had in raising mean achievement in reading and math and their desire to reduce the effect of prior knowledge of the school language on achievement in those content areas. The backward mapping approach helps us to see not only the policy preferences of those at the center but also those at the local or school level. In this case, the organizational problems were such that local policymakers (i.e. teachers and administrators) felt it was more important to get on with teaching, regardless of its quality, than to waste time and energy carrying out the government's policy.

c. Other external conditions affecting decisions.

A backward mapping approach allows policymakers the opportunity to see what other conditions might affect efforts to improve mathematics and reading achievement. We will return to this point below by showing how a condition such as teacher absenteeism might be discovered.

d. Implementing agencies and means of promoting outcomes.

Since the Elmore approach would have surfaced the organizational issues described above, the opportunity would have presented itself to policymakers to determine which conditions could be modified by government policy and to identify the actions and organizations that could have been given responsibility at the operating level for implementing policy.

e. Tools of implementation.

The backward mapping approach, finally, provides policymakers the opportunity to reconsider whether the tools of implementation that appear beneficial from a central level perspective are in fact the most promising and most feasible in terms of local perspectives. Given what we learned using this strategy, would the radio in-service intervention have provided the most beneficial per unit cost compared to the other two alternatives? Could changes have been made to create incentives for grassroots level implementation? This approach forces one to pose such questions.

In summary, this hypothetical example illustrates the kind of reasoning that might produce policies more congruent with existing local level incentives and disincentives. But before we leave this approach, another point should be made. As we noted above, knowledge of how teachers, students and administrators actually make decisions (not how they ought to make decisions) helps policymakers estimate more precisely the full range of consequences that policy is likely to have. As alluded to above, moreover, backward mapping has the additional advantage of helping to identify other external conditions that can affect the choice of policy in critical ways.

For example, policymakers using this backward mapping strategy might well discover other problems that must also be addressed because of their ability to negatively affect efforts to improve mathematics and reading achievement. In

the first section of this essay we discussed five factors which influence the effectiveness of schools in producing student outcomes. If teacher absenteeism were a substantial problem in this country, teacher engagement and expectations would also be negatively affected. After all, if a teacher is frequently absent from school, any hope of improving mathematics and reading achievement would be disappointed regardless of any other intervention taken. Discovering this problem would lend itself, in turn, to a backward mapping strategy to discover factors, such as the following, which might account for the problem. Some of these factors might be known a priori; others might emerge in the process of backward mapping.

i. Regulations of the Ministry of Education (including not only prescriptions, rewards, penalties relating specifically to attendance, but also other requirements that make it more or less likely for teachers to be in school);

ii. Alternative employment opportunities (including any possibilities for increasing personal income through self-employment or dual employment) and the resulting opportunity costs of being in school;

iii. Family responsibilities (child care, non-income generating food production, housekeeping, duties for entertaining relatives);

iv. Expectations of the religious community to which teacher belongs;

v. Conditions affecting health of teacher;

vi. Expectations of school principal and other teachers concerning attendance;

vii. Opportunities for engaging in leisure activities during school day;

viii. Unfavorable working conditions leading to job dissatisfaction (e.g. student absenteeism, disorderly students, low student engagement).

Of course, in pursuing this strategy, it would be important to take into account the possible variations of teacher attendance by type of schools. Reasons accounting for the absence of rural teachers, in other words, might be different from reasons for the absences of urban teachers.

What we have seen in this example is the utility of backward mapping as a strategy to surface problems that might need to be confronted if the original goal is to be addressed effectively. In the process policies that appear to hold promise for accomplishing a goal may give way to other alternatives as such new information comes to light.

III-E. Need for better specified models

Each of the three preceding sections requires causal analyses of the outcomes of schooling in developing countries. Each approach is based on a somewhat different framework for explicating the determinants of these outcomes. The approaches required for the multilevel linear modeling and the cost analyses are more heavily quantitative and based on statistical modeling while the other--the backward mapping approach--can incorporate more qualitative analyses. But no matter which of these approaches is employed, one of the difficulties inherent in any causal analysis is the determination of a model that is sufficiently well specified to allow a prediction of policy effects that is accurate enough to be useful.

The need for a better specified model derives from the importance of properly identifying the causal structure inherent in a given context. Once this is done, one can be more certain of (a) correctly estimating the effects of the factors of interest, (b) including all relevant factors that might be the targets of educational interventions, and (c) being able to design workable strategies of intervention through understanding of the causal relationships involved.

Whenever there are multiple factors influencing a particular outcome of schooling, then in order to estimate the magnitude of the effect attributable to each of these factors, all the factors need to be included in the analytical

model. Otherwise, effects that ought to be attributed to a factor not included in the model will, by default, be misappropriated to a different factor that has been included in the model. This happens when the variable left out is correlated with the variable included, and both are determinants of the student outcomes. The result can be either overestimation or underestimation of the effect of the variable that is included and, clearly, underestimation of the effect of the variable left out. If the variable that is included represents an alternative educational treatment, the result can be the estimation of invalid cost-effect ratios with correspondingly misleading conclusions.

The example we have discussed earlier can be used to illustrate this point. Recall that we are investigating the effect of instructional materials on student outcomes in mathematics and reading and comparing the magnitude of this effect to the magnitude of effects for changes in class size and provision of radio in-service for teachers. Suppose now, as we have earlier, that another important determinant of student outcomes in reading and mathematics is teacher absenteeism. Suppose also that teacher absenteeism tends to be high when provision of instructional materials is low. Then if teacher absenteeism is left out of the model when instructional materials is left in, the effect of instructional materials will be incorrectly estimated by the omission of teacher absenteeism. This would lead to an invalid cost-effectiveness ratio for instructional materials and subsequently misleading comparisons with the ratios for alternative interventions. The exclusion of teacher absenteeism from the model also means that we might be neglecting a possible alternative strategy for improvement. A strategy to reduce teacher absenteeism might even turn out to be the most cost-effective strategy.

This need for adequately specified models means that policy analysis must pay attention to many variables which by themselves are not susceptible to

change through policy. For those who believe that the ability of children is fixed, ability is a variable of this type. Social class, for example, is a variable that may indeed be susceptible to change through social conditions or policy, but which in terms of education policy cannot be modified in the short run. Likewise, ethnicity, religion and gender are important variables that are not readily susceptible to change through policy.

As another example, let us analyze the determinants of teacher absenteeism. Both alternative employment opportunities and teacher working conditions are important determinants of teacher absenteeism. Assume that we have a case in which good alternative employment opportunties are correlated with bad working conditions. Further assume that working conditions is in the model and alternative employment opportunities is not. Then the effect of working conditions will be incorrectly estimated and policymakers might be led to think that changes in working conditions would have a larger effect on teacher absenteeism than would actually be the case if policies to improve working conditions were put into effect.

In short, our models of school effectiveness will have to take into account various variables with which educational policymakers might not otherwise be concerned. The backward mapping approach is particularly useful in such cases in that omissions from the models can be identified even after the fact and adjustments made in the analyses. It is only through this search for all the variables that ought to be included in the model that the true importance of the policy-relevant variables can be assessed.

School improvement policy: What we know and what we have to find out in school improvement policy

Much has been learned from research on schooling in industrialized and developing countries--much about the attributes of effective schools, much about strategies to improve schools, and much about the issues that policymakers need to confront as they consider these strategies. This research suggests that schools have much in common across countries and across settings within countries. Nevertheless, Kagia (1986) of Kenya has good reason for criticizing some of the publications on the Third World by researchers from industrialized countries for being overgeneralized. We agree that, before good policy about effective schools can be made, more must be known about specific settings.

There are, in fact, several conditions that must be met before one could reasonably expect a recommended set of policies in one country to be replicable elsewhere. In particular, the following would have to obtain:

a. The relationships among causal determinants, contingent conditions and outcomes would have to be shown to be similar in the replication site and in the original site. Our discussion of backward mapping suggests that such replicability is unlikely (in a discussion of the ASEAN countries, Postlethwaite and Thomas, [1980], provide a good example when they infer from earlier research that different class sizes are optimal under different conditions). In general, if we believe that the perspectives, interests and capacities of local decision-makers influence and constrain the effects of centralized policies, then the centralized policies can be expected to have predictable results across countries only if those local perspectives, interests and capacities are similar across settings. Otherwise there will be various statistical interactions between these local conditions and the policy determinants that have to be taken into account.

b. Replicability of policy recommendations also demands that decision-makers hold the same constellation of values across countries. That is, the required assumption would be that the same outcomes are valued and the same social distribution of outcomes sought so that the same set of utilities drive decision-making. Again, this similarity of utilities across countries must be demonstrated on a case-by-case basis.

c. The same relative costs of implementing alternative policies must hold across sites. Yet it is likely that resources which are extremely scarce in one country will not be so scarce in another country.

Still, even if specific policy choices are not replicable across settings, our framework for modeling and selecting policies will apply. The logical sequence of forward and backward analysis is replicable.

To illustrate what we know and what we have to find out about particular settings, let us return to the three parts of our essay, beginning with the five factors judged to be critical to school effectiveness. We know a good deal about the nature of the conditions that influence teacher engagement in a positive or negative direction. We do not know the relative importance of these factors in all the specific settings for which policy might be made. We also know that teacher expectations and teacher knowledge are crucial to student success. We do not know what the organization and character of particular sites contribute to increased knowledge or higher expectations. We know that the availability and effective use of instructional materials can contribute in important ways to student outcomes, but we do not know how to organize schools to get the most effective use across a variety of particular settings. We know that the provision of some direction and assistance to teachers is essential to raising the average level of teaching. But we do not know how to make this direction and assistance consistent with cultural norms and supportive of good practice in any given setting. Nor can we be sure that this direction and assistance will be helpful both to the truly exceptional teachers and the ineffective teachers that are found in every school system. We know that it is important to build upon and not obstruct community support and student demand for education, but we do not know in advance what will constitute complementarity in any given setting.

Likewise, we know something about when it is appropriate to use the five strategies for school change. But here again there are limitations to our knowledge. They include not only a lack of systematic evidence concerning specific settings, but also the fact that each of the strategies taken by itself has but limited chances of success, can be used only under certain conditions, and if used indiscriminately will lead to many unintended consequences. For example, we know that radical change in the organization of schooling can produce profound changes in effectiveness, but to make these changes requires extraordinary conditions and support. These conditions may well be present in Third World countries. Indeed the changes that have already taken place in these countries have often been profound, but that does not mean that we can reorganize these systems in whatever fashion and at whatever time seems desirable to us or to policymakers.

School improvement through regulation appears to be a more feasible option since schools in Third World countries are, in general, already highly centralized and therefore the objects of centralized regulation. However, the literature makes clear that only certain areas are susceptible to change through regulation and that central areas of teaching and learning are very difficult to change in intended fashion through this strategy. In fact, as the example of examination change in Sri Lanka indicates, the strategy of reform by regulation even when it brings about change often has unintended negative consequences that outweigh the positive benefits.

Reallocating resources and changing the distribution of incentives is also potentially a very powerful strategy. But under the conditions of a zero-sum game--no increase in overall resources--it is likely to arouse powerful political opposition even when to a detached observer the benefits appear to greatly outweigh the costs.

Providing time, resources and support for organizational and staff development is, all things considered, perhaps the most powerful strategy for bringing about intended improvements in schooling. It is based on the premise that the incentive of assistance and support must be commensurate with the effort required in any change.

Information sharing and participation in decision-making is also a promising strategy for school change. Whereas other approaches may overlook or even undermine already existing examples of excellent practice, this approach capitalizes on them. By its nature it is less predictable and less susceptible to centralized control--a feature which may be viewed as virtue or vice depending on one's point of view. Since attempts to maintain a free flow of information and to assert the right of individuals to make their own decisions figure so prominently in American political ideology, it could be argued that this strategy is particularly appropriate for U.S. development assistance projects.

In the last section of this essay, we have discussed how to model the process of improving schools in a way that will be helpful to policymakers. This discussion has attended to the following issues:

a. how to estimate the effects of various school change strategies in order to bring about improvement in the factors we have chosen to represent effective schooling;

b. why analyses have to take into account the different levels of schooling (individual students, classrooms, schools, etc.);

c. how to deal with the multiple outcomes of schooling and the different values that are placed on these outcomes;

d. how understanding of grass-roots decision-makers contributes to effectiveness of policymaking;

e. how to specify what information needs to be collected in order to make the internationally valid findings on school improvement nationally valid in terms of the particular needs of a given Third World country;

f. why policymakers have to pay attention to variables which are not directly susceptible to policy manipulation;

g. how to factor cost into decisions about school improvement policy.

The more we attend to these issues, the more precisely we will be able to predict the effects of any proposed changes in school improvement policy.

In short, this literature review indicates that there is a good base for further research and policy development on school effectiveness in Third World countries. We know a good deal about why some schools are more effective than others, and a fair amount about how to change schools and how to model the processes of schooling and decision-making upon which effectiveness depends. However, this knowledge does not lead to a recipe for school improvement, but rather to specification of what is needed for policy analysis in a given country.

APPENDIX ONE

CRITIQUE OF FULLER

Bruce Fuller's article ("Raising school quality in developing countries: What investments boost learning?") provides an important and useful orientation to policy issues surrounding the evaluation of alternative educational improvement strategies in Third World countries. The discussion of post World War II historical developments, including the rapid expansion of access to schooling, the leveling off of resources for education, and widespread inattention to school quality nicely sets the context for the BRIDGES project.

BRIDGES' key purpose is to facilitate sound resource allocation decisions to optimize school quality in contexts where many basic educational needs are unmet and resources for meeting such needs are in extremely short supply. In this context, Fuller further focuses the discourse by asserting that the quality of the schooling experience is more important for development than years of schooling, that improving quality of existing schools yields a bigger pay-off than further school expansion, and that investments to primary education pay off more than investments to higher levels of education for most Third World countries' development goals.

By setting the context in this way, Fuller helps us see that the ways primary schools are funded, structured, and managed influence their effectiveness and efficiency. Moreover, he takes an important beginning step by gathering empirical evidence on these issues and subjecting them to

systematic synthesis. Hence, the Fuller paper is crucial for us to study. However, it is also important to clarify its shortcomings.

In its conceptualization, Chapter 3 ("What school factors boost achievement?") illustrates the problems which inevitably arise when analysts seek to estimate how inputs measured at a macro level affect student outcomes without attempting to specify the causal mechanisms through which such inputs are presumed to work. This point is developed with examples in Section I-F of this essay. Briefly, our argument is that in most cases, macro level inputs take on educational significance only when one examines the social purposes for which they are used; and only then do their often inconsistent effects become comprehensible. Fuller's failure to consider causal mechanisms leads to findings which are at times incomprehensible and must therefore be judged of questionable validity.

In its method, this chapter summarizes whole streams of literature by tallying up significant positive, null, and significant negative effects and then reporting the "vote count" as a summary of the effect of a particular variable. Aside from its statistical inadequacies, which are discussed later, the method tends to reinforce the "black box" character of the conceptualization. Inconsistencies in the finding of studies of a particular input's effect, which reflect the fact that people in different contexts use inputs for different effects, tend to be suppressed by this method. The result is an analysis which encourages stronger generalizations for policy than are warranted by the evidence.

Conceptualization

To guide investment decisions aimed at improving school quality, Fuller seeks to distinguish between those "elements" of school functioning which

increase achievement from those which do not. This analysis fragments conceptualization by encouraging policymakers to think of schools as lists of isolated elements which either do or do not "matter". His qualifying statements, which acknowledge the interrelatedness of these factors, which recognize their confounding with other ecological variables, and which admit that the effect of any intervention is contingent upon contextual conditions, provide only minor distractions from his primary task of persuading the reader that some investments (e.g. instructional materials, teacher training) pay off while others (e.g. class size, laboratories) do not.

The problem is that the paper lacks a theoretical framework which might render his findings interpretable. As a result, the data seem to do the talking. However, what the data say often does not make sense. Consider for example, the finding that per pupil expenditures have a positive effect in six countries, but no effect in five others (p. 34). Now it seems obvious that if no money is spent on schools, school quality will be low. Spending money on such schools will improve quality unless the money is completely wasted. If a study asserts that per pupil spending in Third World countries doesn't matter, our <u>a priori</u> temptation would be to discount the study. Either the schools studied display no meaningful variation on the independent variable (in which case the hypothesis can't be tested), the outcome variable is invalid, or the study is otherwise badly flawed, for instance by failing to account for important confounding variables.

The paper also encourages the reader to seek broad generalizations about the "elements" that affect learning. For instance, Fuller strongly asserts that investments in instructional materials and libraries pay off but that investments in reducing class size or building laboratories do not. This approach tempts the reader to think that correlations between inputs and social

outcomes are like physical scientific laws which are not context dependent. It is an approach, we have argued, which supplies a weak basis for policymaking. Moreover, by failing to consider how an input might come to have an effect, Fuller fails to take advantage of one way to evaluate his data.

Consider the tabulated results on pp. 35-38 (Table 9) of his study. It appears for Thailand, that instructional materials do not matter, that laboratories do help, and that larger classes are more effective than smaller classes! Not only do the findings for Thailand contradict his vigorously asserted global generalizations, but also they fail to make sense. By what mechanism does increasing class size improve achievement in Thailand? Since we have no theory that can possibly explain such a finding, the most prudent course is simply to dismiss it as an artifact of flawed methodology, a failure to properly specify a statistical model, so that class size is hopelessly confounded with some unspecified variable that does matter.

In short, one major problem is not that some of the studies reviewed are untrustworthy, since that problem is likely to haunt any analysis, but that the paper provides no coherent theoretical framework for critically evaluating the findings of the studies under review.

Methodological issues

The paper is also flawed methodologically and these flaws combine with the lack of theory to produce unfortunate results.

First, the modal design of the studies synthesized is apparently the cross-sectional survey. Yet the modal inference is strong causal inference about the expected effects of policy changes. Such inferences are unwarranted.

Knowing that availability of instructional materials is related to achievement does not imply that providing more of them will boost achievement.

Nor does the lack of correlation between class size and achievement imply that boosting class size will do no harm. The point is not just that "correlation is not cause." The point is that the consequences of intervening in social systems cannot be known without intervening in them. Thus experimentation, not observation, is the study design that justifies strong inferences about policy effects. The experimental studies that Fuller reviews provide some of the most convincing evidence in the paper, but most of the paper's strong causal inferences are based on correlational evidence.

The cross-sectional designs of most studies he reviewed also make it impossible to distinguish between schools which have high achievement (but may be declining) and schools which are improving. Gifford and Stoddard (in press) show that the correlates of improvement can be quite different from the correlates of status.

Second, the paper uses a method of quantitatively synthesizing study results which is described in the methodological literature on "meta-analysis" as the "vote count" method. In employing this method, the reviewer totes up the number of significant positive, significant negative, and null findings. The modal category is then proclaimed the winner. The method routinely produces misleading results for three reasons.

a. Weak power. For instance, Fuller found no effect of class size in 11 of 21 studies, and concluded that class size has little effect on outcomes. However, since the expected number of significant effects from 21 studies is about one at the five percent level of significance, the correct inference is that the effect of class size on achievement is highly significant statistically. That is, the probability of obtaining 10 non-null findings from 21 studies under the null hypothesis of no effect of class size is so minuscule that it can be dismissed as a practical impossibility. Of course, five studies

reported a significant negative effect. But, as mentioned, this finding says more about the credibility of the studies than it does about the effect of class size.

b. The equation of statistical and substantive significance. The vote-count method provides no basis for assessing the magnitude of an effect, and therefore, no basis for assessing its importance for policy. In Fuller's defense, he makes it clear that many studies fail to report effect sizes, and this failure of reporting plagues social science research generally. However, from a sample size and a p value one can estimate effect size. In the absence of such a measure, the policy importance of a finding is confounded with the sample size of the study reporting it. A small effect will be highly significant if the sample is large enough and a large effect will, with high probability, be dismissed as null in a small-sample study.

c. Failure to assess consistency. The vote-count method provides no basis for assessing consistency of findings across studies. Even when all studies report positive effects, the findings can be highly inconsistent. This would occur if the estimated magnitudes of effect varied by more than could be expected by chance given a common true effect. For instance, a particular input might demonstrate a trivially small positive effect in one study, but one which is nonetheless statistically significant because of a large sample. In another study, the same variable might have a positive effect of major importance. Fuller's approach would summarize these two findings as consistent despite their contradictory implications for policy.

When about half of a set of studies report no significant effect and the other half report an effect (which occurs for several variables in Fuller's synthesis), unless sample sizes are small, the results are probably inconsistent. If the studies are well-designed, inconsistency means that the

factor in question interacts with unspecified contextual conditions, so that its effect depends on characteristics of the setting which are unknown.

In this case, the best remedy is to explore both empirical and theoretical evidence for mechanisms which moderate the effect of the variable in question since findings regarding the average effect of the variable provide a misleading basis for policy.

Lessons for future synthesis

The Fuller paper provides an important start for quantitative synthesis of research on school quality in Third World countries. To improve upon it, the following steps are needed:

1. A conceptual framework is needed to clarify how different kinds of inputs are interrelated and to specify the mechanisms by which these inputs might plausibly affect student learning. Such a framework provides both the beginnings of the modeling process proposed by BRIDGES and a vantage point from which to interpret and to critically evaluate the study findings under review.

2. In quantitatively synthesizing study results, measures of effect size should be constructed. Studies which fail to provide sufficient information to do this (n and p or a test statistic) should probably be discarded.

3. Consistency of effect across studies should be evaluated. When a set of studies report inconsistent results, the studies should be examined closely to discern whether such inconsistency reflects methodological variations or interactions with context. If possible, statistical models for explaining variation in study effect sizes should be formulated and tested.

4. The strength of policy-relevant inferences should reflect the adequacy of the design of the studies on which they are based.

5. Learning from research is a process of both questioning the validity of new evidence on the basis of a priori knowledge and revising beliefs on the basis of new evidence. A synthesis should explicate how the research process has achieved both of these functions. It can do so by making a priori and a posteriori conceptual models explicit and by describing both how study findings were evaluated and how they influenced changes in thinking.

APPENDIX TWO

MULTILEVEL ANALYSES: TECHNICAL REVIEW

It has become increasingly clear that traditional statistical methodology provides an inadequate basis for studying most complex educational processes, especially when the goal of the research is to assess the effects of policies implemented in classrooms or schools on student outcomes (cf., Cronbach, 1976; Cronbach and Webb, 1975). For example, if students are the units of analysis, traditional methods require the assumption that each student responds independently of other students to educational programs or treatments. Yet, because students are "nested" within ability groups, classrooms, schools, and districts, such independence seldom obtains. Some analysts respond by employing classrooms or schools as the units of analysis, and a whole stream of methodological controversy addresses the choice of the appropriate unit. However, the most thoughtful commentators (Cronbach, 1976; Haney, 1980; Rogosa, 1978) have come to believe that the "unit of analysis" controversy misses the mark. The reality is that educational processes occurring within each level of the social organization of schooling ultimately influence children's growth and that these multi-level processes are interactive. In most cases the methodological challenge is not to choose the "correct" unit, but to formulate and test explicit statistical models for processes occurring within each level and between them.

On one level, the challenge is purely technical: to avoid the misestimated precision (Walsh, 1947), aggregation bias (Robinson, 1950), and model misspecification (Cooley, Bond, and Mao, 1981) which seem almost inevitably to result from ignoring the hierarchical, multilevel character of the educational process. Recent breakthroughs in statistical theory now afford, at least in principle, the opportunity to solve these technical difficulties. More important, however, these statistical methods now provide a basis for enriching the kind of research questions accessible to empirical investigation.

After years of debate over the appropriate statistical methods for assessing educational effects of multilevel contexts, a rare but refreshing convergence of opinion has emerged. The ideas underlying this convergence are not new; they were presented elegantly in Lindley and Smith's (1972) classic article. That article presented a hierarchical linear model in which parameters estimated at a lower level of aggregation were presumed to vary, in part as a function of variables measured at a higher level of aggregation. Similar models had been proposed and discussed in the context of the study of growth curves (cf. Elston and Grizzle, 1962) but the constraints on applicability of these models were too restrictive for most field research applications in educational research. It was the generality of Lindley and Smith's formulation which laid the basis for later applications.

Despite the stir it caused in the world of statistical theory, Lindley and Smith's work had little effect on the practice of educational research until recently (for an exception to this rule, cf. Novick, Jackson, Thayer, and Cole, 1972). The problem was that implementing Lindley and Smith's conceptualization was too computationally complex for the large data sets yielded by educational survey data.

The key computational breakthrough was the development of the EM algorithm (Dempster, Laird, and Rubin, 1977) which provided an accessible method for estimating the variances and covariances of the parameters viewed as varying randomly across aggregates. Dempster, Rubin, and Tsutakawa (1981) illustrated application of the methods and provided computational formulae.

Some statisticians might cringe at the assertion that a consensus has been achieved, for heated debate continues over the optimal estimation theory for these models (cf. Deeley and Lindley, 1981; and discussion of Morris, 1983). However, this debate focuses on how to estimate the parameter of the model, not on the model itself, which is quite widely thought to well represent multilevel processes. Moreover, the various estimation procedures which have been proposed yield quite similar results when data sets are reasonably large. The EM algorithm has achieved widespread use because of its relative simplicity. Mason, Wong, and Entwisle (1984), and Raudenbush and Bryk (1986) review how earlier developed statistical models for contextual effects can be incorporated under this more general model.

The statistical approach we propose for modeling multilevel effects has been labelled empirical Bayes estimation (Morris, 1983), covariance components modeling (Dempster, Rubin, and Tsutakawa, 1981), and regression with randomly dispersed parameters (Dielman, 1983). We prefer the term hierarchical linear modeling (HLM), because it highlights the class of substantive problems to which these methods apply.

To illustrate HLM, suppose that researchers wish to investigate how school policies and practices influence the social distribution of achievement. For each of many schools we might formulate a statistical model which relates student background characteristics (socioeconomic status, ethnicity, gender, etc.) to outcomes. The structural parameters of that model, which begins to

describe the social distribution of achievement, are viewed as varying randomly across the whole population of schools. A second, between-school statistical model can then be formulated to explain variation among those structural parameters as a function of differences between schools.

Below we illustrate how the model might be applied in school effects research and how it might be extended to study cost-effectiveness.

Within-school model

Consider for each school an educational production function which related student input variables (background characteristics) X_{ijk} to outcomes Y_{ij} for student i in school j:

$$Y_{ij} = \mu_j + \Sigma \beta_{jk} X_{ijk} + R_{ij}.$$

Here, μ_j represents the average level of the outcome in school i (adjusted for the background characteristics X_{ijk} of its students); and β_{jk} represents the strength of association between background characteristics and achievements. Together μ_j and the β_{jk} characterize the social distribution of achievement in school i. For compactness, we consider $\beta = (\mu_j, \beta_{j1}, \beta_{j2}, \ldots, \beta_{jK})'$ to be a K by 1 vector describing this distribution.

Between-school model

Between schools, the social distribution of achievement, described by β results from school characteristics W_j, which are contextual variables, and Z_j, which are policy/practice variables. Thus, we have:

$$\beta_j = W_j\gamma + Z_j\theta + U_j.$$

Here γ tells us the strength of association between contextual variables and β; θ denotes the strength of effect of the policy/practice variables.

Expectancy x value model

The above model fits readily into an expectancy x value framework. Associated with any change in Z_1 are two things: (1) a cost of implementing that policy; and (2) a change in β_1 equal to

$$(Z_{new} - Z_{old})\theta.$$

For the above expression, θ represents the rate of return for a unit increase in Z_1. Thus, associated with $(Z_{new} - Z_{old})$ are a cost, an expected gain, and a value associated with that gain.

Example

U.S. high school data suggest that, after controlling for student level characteristics and school level contextual variables, the number of math courses required related positively to mean achievement and negatively to the strength of effect of SES on achievement (Lee, 1986). The benefit of increasing required math courses can be approximated by (a) determining the social value associated with increasing means and "flattening slopes;" and (b) estimating the expected change in means and slopes for a unit increase in course-taking requirements. The cost of attaining these gains can also be estimated, and would undoubtedly require one to determine the expense of hiring or training additional math teachers.

REFERENCE LIST

Au, K. H. & Mason, J. (1981). Social organizational factors in learning to read: The balance of rights hypothesis. Reading Research Quarterly, 17(1), 115-152.

Avalos, B. & Haddad, W. (1981). A review of teacher effectiveness research in Africa, India, Latin America, Middle East, Malaysia, Philippines, and Thailand: Synthesis of results. Ottawa: International Development Research Center.

Barnet, R.J. & Muller, R.E. (1974). Global reach: The power of the multinational corporations. New York: Simon and Schuster.

Barnhardt, C. (1982). Tuning-in: Athabaskan teachers and Athabaskan students. In R. Barnhardt (Ed.), Cross-cultural issues in Alaskan education, Vol. 2. Fairbanks, AK: Center for Cross-Cultural Studies, University of Alaska.

Barr, R. & Dreeben, R. (1983). How schools work. Chicago: University of Chicago Press.

Beeby, C. E. (1979). Assessment of Indonesian education. Wellington: New Zealand Council for Educational Research in association with Oxford University Press.

Behrman, J. & Birdsall, N. (1983). The quality of schooling: Quantity alone is misleading. American Economic Review, 73(5), 928-946.

Berman, P. & McLaughlin, M. (1975-1978). Federal programs supporting educational change, 1-8. Santa Monica, CA: Rand Corporation.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13(6), 4-16.

Bossert, S., et al. (1982). The instructional management role of the principal. Educational Administration Quarterly, 18, 34-64.

Brookover, W. B., et al. (1979). School social systems and student achievement. New York: Praeger.

Brown, L. (1974). By bread alone. New York: Praeger.

Bude, U. & Greenland, J. (Eds.). (1983). The in-service training of primary school teachers in English-speaking Africa. London: Macmillan.

Campbell, D. R. (1981). Going for the answers with questions in a Philippine elementary mathematics classroom. Unpublished doctoral dissertation, School of Education, Stanford University, Stanford.

Chang, L. & Ruzicka, J. (1985). Second International Mathematics Study: United States Technical Report 1. Champaign, IL: Stipes.

Clark, D., Lotto, L., & Astuto, T. (1984). Effective schools and school improvement: A comparative analysis of two lines of inquiry. Educational Administration Quarterly, 20, 41-68.

Cohen, D. K. (1984). "...The condition of teachers' work..." (In Symposium on the year of the reports: Responses from the educational community). Harvard Educational Review, 54, 11-15.

Cohen, E. G., Meyer, J. W., Scott, W. R., & Deal, T. E. (1979). Technology and teaming in the elementary school. Sociology of Education, 52, 20-33.

Cohen, M. (1983). Instructional, management, and social conditions in effective schools. In A. Odden & L. Webb (Eds.), School finance and school improvement: Linkages for the 1980s. Cambridge, MA: Ballinger.

Cohn, E. & Rossmiller, R. (1985). Research on effective schools: Implications for less-developed countries. Washington, DC: World Bank.

Colclough, C. (1980). Primary schooling and economic development: A review of the evidence (World Bank Staff Working Paper No. 399). Washington, DC: World Bank.

Coleman, J. J., Hoffer, T., & Kilgore, S. (1982). High school achievement. New York: Basic Books.

Conant, J. B. (1963). The education of American teachers. New York: McGraw-Hill.

Cooley, W. W., Bond, L., & Mao, B. (1981). "Analyzing multi-level data. In A. A. Berk (Ed.). Educational Evaluation Methodology. Baltimore, MD: Johns Hopkins University.

Corwin, R. (1974). Models of educational organizations. In F. Kerlinger (Ed.), Review of research in education (Vol. 2). Itasca, IL: Peacock.

Cronbach, L. J. (1976). Research on classrooms and schools: Formulating of questions, design, and analysis. Occasional paper, Stanford University Consortium.

Cronbach, L. J. & Webb, N. (1975). Between- and within-class effects in a reported aptitude-by-treatment interaction: Reanalysis of a study by G. L. Anderson. Journal of Educational Psychology, 67, 717-724.

Dalin, P. & Rust, V. (Eds.). (1983). Can schools learn? Berkshire, Windsor: The NFER-NELSON Publishing Company Ltd.

Darling-Hammond, L., et al. (1983). Teacher evaluation in the organizational context: A review of the literature. Review of Educational Research, 53(3), 285-328.

Deeley, J. J. & Lindley, D. V. (1981). Estimation in covariance components models. Journal of the American Statistical Association, 76, 833-841.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society (Series B), 39, 1-8.

Dempster, A. P., Rubin, D. B., & Tsutakawa, R. D. (1981). Estimation in covariance components models. Journal of the American Statistical Association, 76, 341-353.

Dielman, T. E. (1983). Pooled cross-sectional and time-series data: A survey of current statistical methodology. The American Statistician, 37, 111-122.

Dove, L. (1980). The role of the community school in rural transformation in developing countries. Comparative Education, 16(1), 67-79.

Dreeben, R. & Barr, R. (1983). Educational policy and the working of schools. In L. Shulman & C. Sykes (Eds.), Handbook of teaching and policy. New York: Longman.

Edmonds, R. (1978). A discussion of the literature and issues related to effective schooling. Paper presented at the National Conference on Urban Education, St. Louis, MO.

Elmore, R. (1979). Backward mapping: Using implementation analysis to structure program decisions. Political Science Quarterly, 94, 601-616.

Elmore, R. (1983). Social policymaking as strategic intervention. In E. Seidman (Ed.), Handbook of social intervention. Beverly Hills, CA: Sage.

Elmore, R. (1984). Forward and backward mapping: Reversible logic in the analysis of public policy. In K. Hans, & T. Toonen (Eds.), Policy implementation in federal and unitary systems. Boston: Nijhoff.

Elston, R. C. & Grizzle, J. E. (1962). Estimation of time-response curves and their confidence bands. Biometrics, 18, 148-159.

Erickson, F. (1984). School literacy, reasoning, and civility: An anthropologist's perspective. Review of Educational Research, 54(4), 525-546.

Erickson, F. & Mohatt, G. (1982). The cultural organization of participation structures in two classrooms of Indian students. In G. Spindler (Ed.), Doing the ethnography of schooling: Educational anthropology in action. New York: Holt, Rinehart, & Winston.

Firestone, W. & Wilson, B. (1985). Using bureaucratic and cultural linkages to improve instruction: The principal's contribution. Educational Administration Quarterly, 21(2), 7-30.

Fuller, B. (1985). Raising school quality in developing countries: What investments boost learning? (Discussion Paper, Report No. EDT7). Washington, DC: World Bank.

Fuller, B., Noel, M. M., & Malouf, D. B. (1985). Polity and competence: Can the state change teachers' skills? Educational Evaluation and Policy Analysis, 7(4), 343-353.

Gifford, B. R. & Stoddard, T. (in press). Looking for effective schools: A value-added approach. The School Administrator.

Goodlad, J. I. (1984). A place called school: Prospects for the future. New York: McGraw-Hill.

Grant, G. (1982). Education, character and American schools: Are effective schools good enough? Syracuse, NY: Syracuse University Press.

Griffin, G. (1983). Implications of research for staff development programs. Elementary School Journal, 83(4), 414-425.

Hallinger, P. & Murphy, J. (1985a). Instructional effectiveness and school socio-economic status: Is good for the goose, good for the gander? Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Hallinger, P. & Murphy, J. (1985b). Assessing the instructional management behavior of principals. Elementary School Journal, 86(2), 217-247.

Haney, W. (1980). Units and levels of analysis in large-scale evaluation. New Directions for Methodology of Social and Behavioral Science, 6, 1-15.

Hanson, E. (1981). Organizational control in educational systems: A case study of governance in schools. In S. Bacharach (Ed.), Organizational behavior in schools and school districts. New York: Praeger.

Hanushek, E. (1981). Throwing money at schools. Journal of Policy Analysis and Management, 1(1), 19-41.

Hawkridge, D., et al. (1982). In-service teacher education in Kenya. In H. Perraton (Ed.), Alternative routes to formal education. Baltimore: Johns Hopkins University Press.

Heyneman, S. (1976). Influences on academic achievement: A comparison on results from Uganda and more industrialized societies. Sociology of Education, 49, 200-211.

Heyneman, S. (1984). Research on education in the developing countries. International Journal of Educational Development, 4(4), 293-304.

Heyneman, S. (1985a). Diversifying secondary school curricula in developing countries: An implementation history and some policy options. International Journal of Educational Development, 5(4), 283-288.

Heyneman, S. (1985b). Investing in education: A quarter century of World Bank experience (Seminar Paper No. 30). Washington, DC: World Bank.

Heyneman, S. (1986). Educational testing to maximize national economic performance. Paper presented at the annual meeting of the Comparative and International Education Society, Toronto.

Heyneman, S., et al. (1978). Textbooks and achievement: What we know. (Staff Working Paper No. 298). Washington, DC: World Bank.

Heyneman, S., Jamison, D. T., & Montenegro, X. (1984). Textbooks in the Philippines: Evaluation of the pedagogical impact of a nationwide investment. Educational Evaluation and Policy Analysis, 6(?), 139-150.

Heyneman, S. & Loxley, W. (1982). Influences on academic achievement across high and low income countries. A re-analysis of IEA data. Sociology of Education, 55, 13-21.

Heyneman, S. & White, D. (Eds.). (1986). The quality of education and economic development. Washington, DC: World Bank.

Hu, T., Lee, M. L. & Stromsdorfer, E. W. (1971). Economic returns to vocational and comprehensive high school graduates. Journal of Human Resources, 6(1), 25-50.

Hurst, P. (1981). Some issues in improving the quality of education. Comparative Education, 17, 185-193.

Husen, T. (1972). Does more time in school make a difference? Saturday Review (April 29), 32-35.

Jamison, D. T., Searle, B., Galda, K. & Heyneman, S. (1981). Improving elementary mathematics education in Nicaragua: An experimental study of the impact of textbooks and radio on achievement. Journal of Educational Psychology, 73, 556-567.

Kaestle, C. F. & Smith, M. S. (1982). The federal role in elementary and secondary education, 1940-1980. Harvard Educational Review, 52, 384-403.

Kagia, R. (1986). The impact of educational research on policy: Trends and possibilities in Africa. Paper presented at the 30th anniversary meeting of the Comparative and International Educational Society, Toronto.

Lanier, J. E. (1986). Research on teacher education. In M. Wittrock (Ed.), Handbook of research on teaching. 3rd edition. New York: Macmillan.

Lasley, T. J. & Wayson, W. W. (1982). Characteristics of schools with good discipline. Educational Leadership, 40, 28-31.

Lassa, P. (1983). The quality of learning in Nigerian primary education: A note on curriculum and community needs - an example from mathematics. International Review of Education, 29, 245-246.

Lee, V. E. (1986). Multilevel causal models for social class and achievement. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Levin, H. (1983). Cost-effectiveness: A primer. Beverly Hills, CA: Sage Publications.

Levin, H., Glass, G., & Meister, G. (1984). Cost-effectiveness of four educational interventions (Project Report No. 84-All). Washington, DC: National Institute of Education.

Lezotte, L. & Bancroft, B. (1985). Growing use of the effective schools model for school improvement. Educational Leadership, 42(6), 23-27.

Lindley, D. V. & Smith, A. F. M. (1972). Bayes estimates for the linear model. Journal of the Royal Statistical Society (Series B), 34, 1-41.

Little, A. W. & Lewin, K. (1984). Examination reform and educational change in Sri Lanka, 1972-82: Modernisation or dependent underdevelopment? In K. Watson (Ed.), Interdependence in education: International perspectives. London: Croom Helm.

Little, J. W. (1981). School success and staff development: The role of staff development in urban desegregated schools. Final report to the National Institute of Education,. 400-79-0049. Boulder, CO: Center for Action Research, Inc.

Lortie, D. (1975). Schoolteacher: A sociological study. Chicago: University of Chicago Press.

Lufler, H. S., Jr. (1978). Discipline: A new look at an old problem. Phi Delta Kappan, 59, 424-26.

Luna, E., Gonzalez, S. & Wolfe, R. (1986). The underdevelopment of educational achievement: Mathematics achievement in the Dominican Republic eighth grade. Manuscript submitted for publication.

Manasse, A. (1985). Improving conditions for principal effectiveness: Policy implications of research. Elementary School Journal, 85(3), 439-463.

March, J. & Olsen, J. (1976). Ambiguity and choice in organizations. Bergen: University of Norway.

Mason, W. M., Wong, G. Y. & Entwisle, B. (1984). Contextual analysis through the multilevel linear model. In S. Leinhardt (Ed.), Sociological methodology, 1983-84. San Francisco: Jossey-Bass.

Mayo, J., McAnany, E. & Klees, S. (1975). The Mexican telesecundaria: A cost-effectiveness analysis. Instructional Science, 4, 197-236.

Meyer, J. W. & Rowan, B. (1978). The structure of educational organizations. In M. Meyer (Ed.), Environments and organizations. San Francisco: Jossey-Bass.

Mingat, A. & Tan, J. P. (1985). Improving the quantity-quality mix in education: A simulation of policy trade offs (Discussion Paper, Education and Training Series). Washington, DC: World Bank.

Modu, C. C. (1986). Specialist services and fellowships for staff development: Annual technical report (1985). McLean, VA: Institute for International Research, Inc.

Monk, D. H. (1981). Toward a multilevel perspective on the allocation of educational resources. Review of Educational Research, 51, 215-236.

Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. Journal of the American Statistical Association, 78, 47-65.

Murnane, R. (1975). The impact of school resources on the learning of inner city children. Cambridge, MA: Ballinger.

Nairn, A. (1985). Change of curriculum = Change of attitude? A discussion of recent attempts in Sri Lanka and Britain to inject pre-vocational studies into the school curriculum. In K. M. Lillis (Ed.), School and community in less developed areas. London: Croom Helm.

Navarro, R. A. (1985). The "problems" of language, education and society: Who decides? In E. Garcia, & R. Padilla (Eds.), Advances in bilingual education research. Tucson, AZ: University of Arizona Press.

Navarro, R. A., Berkey, R., & Minnick, F. (1986). The art of becoming an institutional teacher. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Neumann, P. & Cunningham, M. A. (1982). Mexico's free textbooks: Nationalism and the urgency to educate (Staff Working Paper No. 541). Washington, DC: World Bank.

Novick, M. R., Jackson, P. H., Thayer, D. T., & Cole, N. S. (1972). Estimating multiple regression in m groups: A cross-validation study. British Journal of Mathematical and Statistical Psychology, 25, 33-50.

Nwagwu, N. (1981). The impact of changing conditions of service on the recruitment of teachers in Nigeria. Comparative Education, 17(1), 81-86.

Odden, A. & Webb, L. (Eds.). (1983). School finance and school improvement: Linkages for the 1980s. Cambridge, MA: Ballinger.

Orfield, G. (1969). The reconstruction of southern education: The schools and the 1964 Civil Rights Act. New York: Wiley-Interscience.

Ogbu, J. (1983). Minority status and schooling in plural societies. Comparative Education Review, 27, 169-190.

Philips, S. U. (1982). The invisible culture: Communication in classroom and community on the Warm Springs Indian Reservation. New York: Longman.

Pollitt, E. (1984). Nutrition and educational performance. Prospects, 14 (4), 443-460.

Popkewitz, T. S., Tabachnick, B. R., & Wehlage, G. (1982). The myth of educational reform: A study of school responses to a program of change. Madison, WI: University of Wisconsin Press.

Porter, A. (1984). Education priorities in sub-Saharan Africa. Paper presented at the Conference on Education Priorities in Sub-Saharan Africa, Windsor, England.

Porter, A., Schwille, J., Alford, L., Floden, R., Freeman, D., Irwin, S. & Schmidt, W. (1986). State policy and the control of curriculum decisions: Zones of tolerance for teachers in elementary school mathematics. (RS 173). East Lansing, MI: Institute for Research on Teaching, Michigan State University.

Postlethwaite, T. N. & Thomas, R. M. (Eds.). (1980). Schooling in the ASEAN region. New York: Pergamon Press.

Powell, A. G., Farrar, E. & Cohen, D. K. (1985). The shopping mall high school: Winners and losers in the educational marketplace. Boston: Houghton Mifflin.

Psacharopoulos, G. & Loxley W. (1985). Diversified secondary education and development: Evidence from Colombia and Tanzania. Baltimore: Johns Hopkins University Press.

Psacharopoulos, G. & Woodhall, M. (1985). Education for development: An analysis of investment choices. New York: Oxford University Press.

Purkey, S. & Smith, M. (1982). Too soon to cheer? Synthesis of research on effective schools. Educational Leadership, 40, 64-69.

Purkey, S. & Smith, M. (1983). Effective schools: A review. Elementary School Journal, 83, 427-452.

Purkey, S. & Smith, M. (1985). School reform: The district policy implications of the effective schools literature. Elementary School Journal, 85, 353-389.

Raudenbush, S. & Bryk, A. S. (1986). A hierarchical model for studying school effects. Sociology of Education, 59, 1-17.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. American Sociological Review, 15, 351-357.

Rogosa, D. (1978). Politics, process, and pyramids. Journal of Educational Statistics, 3(1), 79-80.

Rosenholtz, S. (1986). Organizational determinants of teacher commitment. A paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Rowan, B. (1981). The effects of institutionalized rules on administrators. In S. Bacharach (Ed.), Organizational behavior in schools and school districts. New York: Praeger.

Rowan, B., et al. (1983). Research on effective schools: A cautionary note. Educational Researcher, 12,(4), 24-31.

Rutter, M. (1983). School effects on pupil progress: Research findings and policy implications. In L. S. Shulman, & G. Sykes (Eds.), Handbook of teaching and policy. New York: Longman.

Rutter, M., et al. (1979). Fifteen thousand hours: Secondary schools and their effects on children. Cambridge, MA: Harvard University Press.

Saunders, M. & Vulliamy, G. (1983). The implementation of curricular reform: Tanzania and Papua New Guinea. Comparative Education Review, 27(3), 351-373.

Sauvant, K. (1976a). The potential of multinational enterprises as vehicles for the transmission of business culture. In K. Sauvant & F. Lavipour (Eds.), Controlling multinational enterprises: Problems, strategies, counterstrategies. Boulder, CO: Westview Press.

Sauvant, K. (1976b). His master's voice. CERES, 9(5), 27-32.

Schiefelbein, E., Farrell, J. P., & Sepulveda-Stuardo, M. (1983). Influence of school resources in Chile: Their effect on educational achievement and occupational attainment (Staff Working Paper No. 530). Washington, DC: World Bank.

Schultz, T. W. (1964). Transforming traditional agriculture. New Haven, CT: Yale University Press.

Schwille, J., Porter, A., Alford, L., Floden, R., Freeman, D., Irwin, S., & Schmidt, W. (1986). A longitudinal study of the effects of district curriculum policies on the content decisions of teachers. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Schwille, J., Porter, A., Belli, G., Floden, R., Freeman, D., Knappen, L., Kuhs, T., & Schmidt, W. (1983). Teachers as policy brokers in the content of elementary school mathematics. In L. S. Shulman, & G. Sykes (Eds.), Handbook of teaching and policy. New York: Longman.

Scribner, S. & Cole, M. (1981). The psychology of literacy. Cambridge, MA: Harvard University Press.

Sedlak, M., Wheeler, C., Pullin, D., & Cusick, P. (1986). Selling students short: Classroom bargains and academic reform in the American high school. New York: Teachers' College Press.

Shultz, J. J., Florio, S., & Erickson, F. (1982). Where's the floor? Aspects of the cultural organization of social relationships in communication at home and in school. In P. Gilmore, & A. A. Glatthorn (Eds.), Children in and out of school. Washington, DC: Center for Applied Linguistics.

Simmons, J. & Alexander, L. (1978). The determinants of school achievement in developing countries: A review of research. Economic Development and Cultural Change, 26, 341-358.

Somerset, H. (1984). A note on the effects of head teachers and of supervisors on school quality. Paper presented at the Conference on Education Priorities in Sub-Saharan Africa, Windsor, England.

Spady, W. (1982). Outcome-based instructional mangement: A sociological perspective. Australian Journal of Education, 26, 10-29.

Sparks, G. (1983). Synthesis of research on staff development for effective teaching. Educational Leadership, 41(3), 65-72.

Spindler, G. D. (Ed.). (1982). Doing the ethnography of schooling: Educational anthropology in action. New York: Holt, Rinehart, & Winston.

Sunkel, O. & Fuenzalida, E.F. (1979). Transnationalization and its national consequences. In J.J. Villamil (Ed.), Transnational capitalism and national development: New perspectives on dependence. Sussex: The Harvester Press.

Sykes, G. (1983). Public policy and the problem of teacher quality: The need for screens and magnets. In L. S. Shulman, & G. Sykes (Eds.), Handbook of teaching and policy. New York: Longman.

Taylor, D. C. (1983). The cost-effectiveness of teacher upgrading by distance teaching in Southern Africa. International Journal of Educational Development, 3(1), 19-32.

Thias, H. & Carnoy, M. (1972). A cost-benefit analysis in education: A case study for Kenya. Baltimore: Johns Hopkins Press.

Walsh, J. E. (1947). Concerning the effect of the intraclass correlation on certain significance tests. Annals of Mathematical Statistics, 18, 88-96.

Wangberg, E. G., Metzger, D. J., & Levitov, J. E. (1982). Working conditions and career options lead to female elementary teacher job dissatisfaction. Journal of Teacher Education, 33, 37-40.

Weick, K. (1976). Educational organizations as loosely coupled systems. Administrative Science Quarterly, 21, 1-19.

Williams, P. (1984). African education under siege. Paper presented at the Conference on Education Priorities in Sub-Saharan Africa, Windsor, England.

Windham, D. M. (1985). Internal efficiency and the African school. Unpublished manuscript, Universite de Dijon, Dijon, France.

Wise, A. (1979). Legislated learning: The bureaucratization of the American classroom. Berkeley, CA: University of California Press.

Yinger, R. J. & Clark, C. M. (1983). Self-reports of teacher judgment (RS 134). East Lansing, MI: Institute for Research on Teaching, Michigan State University.