DOCUMENT RESUME

ED 337 849 EA 023 345

TITLE Educational Evaluation and Reform Strategies: The United States of America. Country Reports.
INSTITUTION Office of Educational Research and Improvement (ED), Washington, DC.; Organisation for Economic Cooperation and Development, Paris (France).
REPORT NO SME/ET/90.12
PUB DATE 14 May 90
NOTE 39p.; "Prepared...as a contribution by the U.S. Education authorities to the Education committee's activity on Educational Evaluation and Reform Strategies."
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Assessment; *Educational Change; Elementary Secondary Education; *Evaluation Criteria; *Evaluation Methods; Evaluation Problems; Evaluation Research; School Based Management; School District Autonomy; School Effectiveness; *School Restructuring; Student Evaluation; Teacher Evaluation
IDENTIFIERS *Reform Strategies

ABSTRACT

In the United States today, policymakers and citizens believe that the quality of their schools must be improved. Reform policies and strategies are the focus of the first section of this report. The four areas identified are the perception and interpretation of the problems of quality, the methods of preparing reform or ad hoc measures, the main features of reform or ad hoc measures, and the strategies of implementing adopted policies. Evaluation and monitoring of policies and reform comprise the second section. In contrast to other countries, the limited role of U.S. federal government in monitoring and evaluation is examined. Seven areas are identified: quality evaluation policies and their methods, evaluation of the quality of schools, evaluation of educational quality, evaluation of teacher quality, evaluation of curriculum quality, evaluation of the quality of other educational services, and changes in the monitoring and supervisory mechanisms. The final section reports on the first results of evaluation and monitoring. (RR)

ORGANISATION FOR ECONOMIC
CO-OPERATION AND DEVELOPMENT

Directorate for Social Affairs
Manpower and Education

SME/ET/90.12

RESTRICTED

Paris, drafted: 27th April 1990

dist.: 14th May 1990

English text only

EDUCATIONAL EVALUATION AND REFORM STRATEGIES

COUNTRY REPORTS

THE UNITED STATES OF AMERICA

(Note by the Secretariat)

1. The attached report has been prepared by the US Department of Education, Office of Educational Research and Improvement, Washington D.C. as a contribution by the U.S. education authorities to the Education Committee's activity on Educational Evaluation and Reform Strategies.

2. The report is circulated for **INFORMATION**.

TABLE OF CONTENTS

The United States (US) Country Report to OECD on Educational Evaluation and Reform Strategies

I. REFORM POLICIES AND STRATEGIES

1) Perception and interpretation of the problems of quality

In the United States today, policymakers and citizens believe that the quality of their schools must be improved. Many think that schools must raise educational standards if the country is to compete successfully in the world economy and preserve its democratic traditions.

The current reform movement is colored by economic anxiety in an era of shifting demographics. Business leaders and politicians are concerned about the competitiveness of the US economy. The US imbalance of trade, loss of jobs to overseas locations, the low level of basic reading and mathematics skills in young employees, and heightened awareness of the global economy fuel concern about the ability of the national workforce to compete internationally. Economic success in the "information age" is thought to depend upon the success of elementary and secondary education in promoting higher levels of literacy, problem-solving skills, and the application of knowledge to new situations. At the same time demographic changes increase the proportion of disadvantaged

students who have not previously performed well in school. People value the multiple purposes schools serve, from developing a range of student abilities, to teaching citizenship and basic human values, and promoting social justice. But reform leaders tend to emphasize the need to prepare students as competent future workers in an era when that is becoming harder to do.

In the popular mind, the origins of these reforms are associated with the 1983 federal report A Nation at Risk and other public and private national advisory reports that were issued about the same time. Speaking with diverse voices from varying perspectives, all called for an improvement in educational quality. Declining student test scores, the poor showing of the US in international comparisons of student performance, and projections of future needs for highly skilled workers were all marshalled as evidence of a need for higher educational standards.

These reports were taken seriously because they rang true. They were not issued by authorities with power to mandate the recommendations into policy. Their power was only the power to persuade, but this power proved to be considerable. They expressed concerns already in the public mind. There exist in the US little systematic data on student achievement over time, but concern had mounted for a decade over the 20-year decline in college admission

(Scholastic Aptitude Test, or SAT) test scores. Businesses were struck by the literacy-training needs of new employees. Studies were conducted on what fraction of American youth did not know things the public thought they should - from locating the US on a world map, to identifying the century in which their country became independent, to fundamental provisions of their Constitution.

These concerns inspired many state legislatures in the 1970s to require tests of minimal competency for high school graduation. These tests, and a "back to basics" movement in the 1970s, focused schools on basic skills instruction - at the expense of both the less important aspects of curriculum, which had catered to student perceptions of topical "relevance," and to the more important aspects of higher order skills and content knowledge.

2) Methods of preparing reform or ad hoc measures

The federal government and others addressing a national audience helped to articulate concern about declining standards and support for educational reform, but they left almost entirely to states the jobs of formulating and financing the reforms. The high federal budget deficit and early '80s trend to cut social and domestic spending made clear to the states that the federal government would not itself legislate or finance the needed reforms.

During the 1970s the relative power of states and localities in education had changed. By tradition, control of educational policy and finances in the US have been local, resting in a locally elected school board allocating a budget derived from local property taxes. During the 1970s a few taxpayers' "revolts" (most notably Proposition 13 in California) capped local rises in property taxes. Pervasive reforms of state funding mechanisms shifted financial support of schools from about 90% local sources to about half coming from the state level. With increased state funding came increased calls for state oversight of educational performance. States increasingly saw education as important to their own economic development. Constitutionally, local control of education had derived from the state. For all these reasons, states became the level of authority that led the response to the perceived need for reform.

Because reform has been decentralized among states, there is no central authority which can speak for it. Much more information is available about state actions than about reforms undertaken at the local level or by professional organizations, where they also exist. Reform leadership varied from state to state. Governors, legislators, business and citizen organizations, chief state school officers, and others each played roles in varying combinations. In some states, a governor or a business roundtable appointed an ad hoc commission and promoted its recommendations. In others, the

state legislature held hearings, initiated legislation, and sometimes continued formal oversight of its implementation. In yet others, the chief state school officer initiated new programs within the state department of education.

This process resulted in sets of reforms that were predictably diverse. Some states borrowed heavily from the recommendations of the national reports; others focused on local concerns. Usually elected officials played a prominent leadership role, sometimes with advice from educators, but rarely at their behest. Usually reforms entailed increased levels of state spending and were therefore likely to need approval by the state legislature. It should be remembered that any attempt to characterize the substance of reforms is a matter of synthesizing diverse policies.

3) Main features of reform or ad hoc measures

With these provisos, one can nonetheless say the reforms cluster around five issues. Policies have matured in each area, initially emphasizing mandates for higher standards, evolving towards more complex and long-term policy initiatives.

(1) attempts to raise standards for student learning through such means as a) raising requirements for high school graduation and public university entrance,

b) graduation exit tests, c) increased testing to monitor mastery of the curriculum, d) recognition of students for outstanding performance, e) alignment of the curriculum and testing programs, sometimes with accreditation as well, f) increased services for gifted and talented students, and g) increased policy consideration of ways to improve higher order thinking and problem-solving skills;

(2) attempts to improve teaching through such means as a) tests of new teachers, b) higher requirements for entrance into teacher training programs, c) greater academic course emphasis within pre-service teacher training, d) alternate teacher certification facilitating talent from other fields entering teaching, d) raising salaries, e) attempting to differentiate roles between novice and expert teachers through career ladders, f) merit pay, g) various attempts to "empower" teachers, improve their working conditions, and enhance the "professionalism" of teaching;

(3) attempts to help students who might otherwise be "at risk" through such means as a) increased early childhood and preschool programs, including increased funding for Head Start, b) attempts to increase parental involvement

in education, c) a variety of dropout prevention efforts, d) a few health and wellness programs aimed to prevent or minimize drug addiction, pregnancy, drunk driving, suicide, and other health threats, and e) some consideration of schools' role as partners in rendering comprehensive social services to meet the needs of the students and families at greatest risk;

(4) attempts of states to hold districts and schools accountable for their performance by collecting, publicly reporting, and rewarding performance data through such means as a) definition and collection of data for a set of "indicators" of the educational performance of schools, and increased attention to test results; b) systematic reports to the public of these school performance data (such as average test scores or other assessment results, daily average attendance, graduation and dropout rates, and other data), c) creation of policy consequences to school performance data, such as allocation of technical assistance, requirements for improvement planning, rewards such as recognition or cash awards for high performance, and sanctions such as increased monitoring or actual state intervention and "educational bankruptcy" proceedings for low school performance; and d) most prominently, the work

evolving from the first education summit ever of the President of the United States and the State Governors to define national goals for education (announced in the State-of-the-Union address in January and at the National Governors' Association meeting in February, 1990) and associated attempts to measure progress and hold themselves accountable for helping to meet them; and

(5) attempts to "restructure" education by altering governance and operations through such means as a) state and locally sponsored experiments in enhanced school-site management (devolving some local school district control over budget, personnel and curriculum to individual school buildings), b) experiments in shared decision-making among principals and teachers within a school, some promoted and sponsored by teacher unions, c) permitting application for waivers from state regulations for schools or districts performing well, pursuing a plan to restructure, or pursuing some other thoughtful school improvement plan, (and associated initiatives to deregulate some federal education programs); d) permitting parents to chose which public school their children will attend, and e) other attempts to modify instruction and coordinate human service delivery at schools.

These reforms are diffuse. Many states have adopted policies for higher student and teacher standards but are at early stages of work to restructure schools. Because they were undertaken by states or sometimes localities, not a central authority, there is variation from state to state in which reforms were adopted, with what relative emphasis, and for what purpose.

4) Implementing strategies

Implementation is pervasive, if hard to assess. Almost all states have adopted policies in what is called the "first wave" of reform since 1983. Typically these are ambitious, multifaceted, and perceived and funded by state authorities as "comprehensive" reform packages.

Many of these reforms take the form of new state regulations required by state legislation or rules which are supposed to be observed by local school districts and schools. The state departments of education oversee the process of implementation. Teachers and principals may only encounter "reform" in the form transmitted from the state department of education through a central district office. They may have only a vague idea of the original rationale and purpose of some reform policies.

This "top-down" pattern sometimes leads to ambiguous

situations, particularly with regard to reforms designed to nurture local creativity at the "bottom." Some states try to empower teachers and permit schools to restructure in order to improve student learning only to find local educators do not yet have a clear vision of how they want to use their new flexibility. Lay policy makers have wanted to increase teacher professionalism only to find that merit pay and career ladders are not popular with teachers. State leaders have been more eager to grant increased flexibility and waivers of state regulations than schools have been eager to apply for them. As implementation progresses, educators become much more involved, and give meaning to reform in the local context. This initial ambiguity is often handled by making extra state money and waivers available to a limited number of schools and districts applying for special projects, so if they are successful, the policy or funding can later become state-wide.

Thus lay sponsors of reform generally trust that with the right combination of incentives existing education institutions will implement reform. Skepticism that inside, "normal channels" will not successfully implement reform is an element in the creation of new accountability mechanisms, but one that leads only occasionally to the creation of new oversight institutions, and very rarely leads to mandates for evaluation of specific reform programs.

The United States is currently spending about $353 billion on education, of which just over half goes for elementary or secondary schools. This represents a significant increase (roughly 29%) since 1980, making the U.S. among the top nations in the world on per pupil spending. While there is considerable state and local variation in per pupil funding, there is not a one-on-one connection between these variations in funding and equally evident variations in educational quality. Most reformers recognize that there are start-up costs of undertaking major educational restructuring, but disagree on whether major funding increases are essential to making schools more productive.

II. EVALUATION AND MONITORING OF POLICIES AND REFORM

Traditionally, oversight of education in the United States has operated though democratically elected local school boards, rather than through monitoring and evaluation by professionals in a central authority. Some other countries have central ministries that not only set policy for school operations, but have explicit responsibility for monitoring the implementation and evaluating the results of those policies. Central school inspectorates have extensive experience at visiting, observing, and recommending improvements for school and teacher performance. Ministries in such countries have a rich history and expertise upon which to draw

in answering the set of questions in this part of the OECD report. The United States, in contrast, does things differently.

In the US the federal government has a limited, almost nonexistent role in monitoring and evaluating anything but the operation of federally funded programs in local schools. Universal public education in the US began based on local oversight. The first federal education agency was established to collect and report national educational statistics. When federal funds became available to fund local educational programs in 1965, associated efforts at program evaluation developed. They were intended to evaluate the effectiveness of specific, new, federally funded programs, and not the effectiveness of normal, local school operations.

The first federally funded effort to assess the quality of what the average American student knew or could do was undertaken in 1969 by the National Assessment of Educational Progress (NAEP). NAEP was carefully crafted to report on nationally and regionally representative samples. Its findings supplied background information about national educational performance to states and local authorities with actual governing responsibility.

Those local school boards and state boards of education did not

rely heavily upon technical evaluation data in order to monitor school performance. The authority of local school boards was based on their political election, which in turn was grounded on their knowledge of the community, rather than on a scientific and professional command of evaluation. State authorities, like the federal government, were distant from the actual operations of schools, but commonly exercised their oversight by establishing minimum standards for school accreditation. These standards tended to focus on educational "inputs" relating to student health and safety and the minimum qualifications of teacher and facilities. They did not originally relate to evaluations of the educational performance or "outputs" of schools.

In the context of recent reforms, these arrangements are changing. State policy makers are shifting accreditation towards output indicators of performance and challenging the traditional authority of local school boards by creating state-operated data systems which, implicitly or explicitly, have monitoring and evaluation purposes. These data systems gauge and report the educational performance of schools, especially their student outcomes. Most states have recently adopted such enhanced accountability policies. These build upon state testing programs and other state data to issue public reports, sometimes called school-by-school "report cards," and to reward and monitor schools on the basis of such performance.

1) Quality evaluation policies and their methods

OECD sensibly asks to know "how the quality policy adopted and introduced affects evaluation." This reasonable question is difficult to answer for the United States. States are putting considerable effort into programs to gauge and report on overall local school performance, but these programs are only loosely connected to more formal or traditional evaluation. Evaluation per se has been allocated meager staff or funds.

States increasingly try to measure and report school performance on some set of indicators for purposes of public accountability. These indicators emphasize student outcomes as the "bottom line" measure of school performance. They also try to contextualize those outcomes. State systems try to include data on both educational policies (that can be set and altered by responsible authorities) and on social and demographic data about the students or communities which the school and governing authority must serve. Thus there is simultaneously emphasis on both student outcomes (the results obtained by pupils) and the quality of resources and strategies used.

In operation, there is unexpressed tension among varying constituencies between the relative emphasis placed on the quality

of student outcomes and the quality of resources and strategies used. Politicians, policymakers, and business leaders tend to look at student outcomes as the "bottom line." (For example, when in 1989 the US Secretary of Education released his annual "Wall Chart" of state comparisons of educational indicators, he concluded that reform had stalled because student performance measures showed so little improvement.) On the other hand, school administrators and teachers tend to put greater emphasis on the importance of indicators marking the quality of resources made available to them. Recognizing the variety of students they serve, teachers feel responsible for following currently accepted norms of professional practice, and do not feel in control of the variety of educational outcomes that result from following those practices. Some superintendents say that if they follow accepted educational practices, they do not need to know student outcomes, presuming that positive outcomes will tend to follow, and that if they do not, the educators cannot be held responsible.

Different constituencies emphasize different indicators of performance. Performance is reported on comprehensive sets of educational indicators, logically divided into "outcomes" (of pupil and school performance), "context" (of population characteristics of demographics and wealth beyond policy control), and "policy" or resource inputs (of school practices that can be altered and

controlled). Politicians, policymakers, most parents, and business see student outcomes as direct indicators of good performance, and context and policy indicators as elaborations upon the setting in which those results were obtained. The National Governors' Association (NGA) 1987 report Time for Results presumed that improved student outcomes were the results being sought. Governors will therefore, like the US Secretary of Education, look at student outcome indicators (such as test scores) as the most important indicators of whether or not education is improving.

In contrast, educators pay more attention to the quality of resources and strategies used to secure improvement. Their perspective is influenced by the troubling issues of the impact of student background, and they therefore attend to indicators of policy or resource inputs (such as funding for early childhood programs, smaller class size, higher teacher salaries, and the like) as the most pertinent predictors of a system's educational performance. Many educators, like parents and advocates for poor, minority, and disadvantaged students, want those students to succeed, and want to produce the same results for those students as for all others. Some say explicitly and often that "All children can learn." Nonetheless, it remains difficult to know when and under what conditions it is a reflection on the performance of educators themselves that disadvantaged students perform badly on standardized measures of performance.

It can happen, t..en, that lay persons see little improvement in student performance and are concerned about the system's performance, while, when examining the same set of indicators, educators feel lay persons have not given them the resources and political support they need. To date, many explain away low performance on indicators systems as either 1) a reflection of the imperfections of the indicators, or 2) a reflection that someone else in the system has not done their part. Many believe there is a need to redefine the roles and reaffirm mutual responsibilities among the many partners involved in improving America's e ucational performance.

2) Evaluation of the quality of schools

A recent trend in the US and elsewhere is to identify the school building as the basic unit of educational service delivery and therefore the basic unit of educational improvement. Commonly, data on educational performance will be aggregated and publicly reported at the school level. A survey conducted by the Council of Chief State School Officers in 1987 indicated that 23 states publicly report performance data at the school-site level (compared with 37 doing so at the local district level, and 43 doing so for state averages), and more report on individual schools each year.

The data in these reports are dominantly school-site aggregates of individual student data. These may be test scores, dropout and graduation rates, attendance rates, or course enrollment. They rarely include data such as course or extracurricular offerings that pertain only to the school level that cannot be aggregated up or down. Likewise, they may not include per pupil spending, teacher salaries, class size, curriculum, and discipline because such policies are normally set at the local district level.

Among the incentives created for school improvement by both the national and state governments are programs to recognize outstanding performance of individual schools. While only about a dozen states automatically recognize outstanding school performance based on the regular state testing program, 33 states recognize outstanding schools when additional nomination and application processes are included. The US Department of Education bases its prestigious school recognition program upon a process of application and peer review which draws heavily upon criteria deriving from what is known as the "effective schools research."

Parents evaluate schools informally. Independent of the technical information issued by governing authorities, parents learn the scholastic and athletic reputation of individual schools, and are likely to use this information in deciding where to purchase a new home.

3) Evaluation of educational quality as measured by pupil performance

Individual pupil achievement represents the backbone of most indicators of educational performance. Newspapers, policymakers, and educators alike look to student scores on recognized national tests (such as SAT and ACT college admissions tests) as evidence of the performance of the educational system.

Traditionally pupil performance is measured by teachers and local districts on the basis of performance on other standardized and locally developed tests. State testing programs have evolved which also combine state-developed and nationally normed standardized tests. Local and state authorities have used these to inform their own decision making, and to report to the public that student performance was acceptable. There has been considerable recent concern that outdated norms on nationally standardized tests have generated inflated test results (see page 34) and thus inaccurate pictures of true student performance. In addition, these state and local tests are not been used to compare schools or to judge national performance and national trends over time.

Some reformers see the most commonly used, nationally standardized tests as an imperfect system of measurement. Many

feel that these tests of pupil achievement overemphasize basic skills and mastery of isolated items of knowledge, and do not reflect recent research on cognitive development and the nature of student learning. There is therefore interest in developing new and better assessments of students' deep understanding of subject matter. The U.S. Department of Education will soon fund a new National Center for Research on Assessment, Testing, and Evaluation to study and help develop such improved measures.

Many in the US are interested in emerging trends towards "authentic" assessment of pupil performance. Developmental efforts such as a National Science Foundation grant to the Connecticut Department of Education parallel those in the United Kingdom and the Netherlands. Connecticut is working to develop improved pupil assessments with a consortium of 6 other states and the Coalition of Effective Schools, which, under the leadership of Theodore Sizer, is working to restructure schools to be more sensitive to students' leaning. New York State already uses a performance-oriented approach in its elementary science test.

Many believe that what gets measured gets done. Reform leaders hope that by focusing assessment on significant concepts and students' deep understanding, and by involving teachers in judging students' responses, these new assessments will create tests worth

teaching to, tests likely to help focus teachers and schools on significant pupil outcomes.

4) Evaluation of teacher quality

Teacher quality is a central object of educational reform, and significant changes are underway in how teachers should be evaluated. These are largely separate efforts, however. There are political roots to attempts to improve teachers and teaching through testing new teachers, alternative teacher certification, higher teacher salaries, career ladders, and other regulations mandated by states. There are independent, professional roots to efforts of the National Board for Professional Teaching Standards to shape how individual teachers, once on the job, are evaluated.

There has been increased interest recently in trying to connect teacher evaluation to student learning. Many fear, however, that this would inappropriately hold teachers accountable for student results they cannot control, results that clearly have multiple causes. Teacher evaluations, therefore, have focused on teacher behavior, rather than student outcomes.

Improvements in teacher evaluation system are being addressed by the National Board for Professional Teaching Standards.

Historically teacher evaluations have been haphazard and weak. They have too often been performed on the basis of checklists of the presence or absence of certain teacher behaviors. The new board, comprised of a majority of teachers, is working to develop authentic measures of teacher performance, which pay increased attention to both the purpose and context of instruction. In addition, a Research and Development Center in Teacher Evaluation and Educational Accountability has been mandated by the US Congress and should soon be in operation.

5) Evaluation of curriculum quality

In February 1990 the President and Governors set as a national goal student mastery of "challenging subject matter." They challenged America to make their students' achievement in math and science "first in the world" by the year 2000. Most states since 1983 had already increased the number of academic courses required for high school graduation. Those new requirements allowed states to conform more nearly to the Nation at Risk recommendation that during the four years of high school, students study 4 years of English, 3 years each of mathematics, science, and social studies, one half year of computer science, and 2 years of foreign language study for students preparing for college. Once these regulations were on the state books, it remained to define the content and meaning behind course titles and new requirements.

There has been ambiguity and mutual mistrust among the parties who might define curriculum content. The political sensitivity of curriculum decisions has divided liberals and conservatives and inhibited government efforts to improve curriculum quality. By law the federal government cannot prescribe a national curriculum, although it has served an important exhortatory role. Local attempts to monitor curriculum for religious and political acceptability leave many leery of unfettered local autonomy. This history, and the perceived failure of national curriculum reform following Sputnik in the 1960's, has made it difficult to discuss and mount a government effort to improve curriculum content.

Yet in fact, there is something like a process for setting national curriculum. Policy has been left in state departments of education and in the hands of local authorities. States have evolved inoffensive and all-inclusive lists of topics to be "covered." These are adapted by the textbook publishing industry into texts, which shape day to day instruction in the classroom. Some say this constitutes a de facto national curriculum. The largest among the states which approve a limited list of textbooks from which local districts can choose, California and Texas, have considerable national influence because they represent large markets.

Textbook selection thus becomes decisive in shaping actual curriculum quality. Although there are not formal, national mechanisms for evaluating texts, knowledgeable leaders have criticized textbooks as having been "dummied down" by emphasizing easy "readability" more than coherence and clear explanation. New editions of textbooks have tended to use short words and sentences, even when more complex words and sentences are needed to explain ideas clearly.

In California, the state department of education has undertaken serious curriculum reform. It has produced what many consider to be very good guidelines and given them power by aligning them with the state testing program and the state textbook adoption policies. California further exerted pressure on the private textbook publishing industry by one year refusing to approve any of the currently available 8th grade math texts until a new version was produced which met the new state standards. Such state action, however, is not typical nor always effective.

Increasing commitment exists to improve curriculum quality. The most significant and persistent efforts are led by professional associations in the areas of math and science. The President and Governors' commitment to raise American student achievement by the year 2000 in challenging subject matter and especially in math and science will undoubtedly enhance these efforts.

The National Council of Teachers of Mathematics issued a report, Curriculum and Evaluation Standards for School Mathematics, in 1989, and others now call for a parallel statement of desired standards for the teaching of science. The National Research Council of the National Academy of Science issued a major report, Everybody Counts: A Report to the Nation on the Future of Mathematics Education, in 1989, and another called Reshaping School Mathematics: A Philosophy and Framework for Curriculum in 1990, both of which promote improvements in curriculum. In addition, the US Department of Education has funded efforts to study instruction in math, science, social studies, literature, and language arts at their national Research and Development Centers. These efforts, combined with the work of the National Assessment of Educational Progress and its governing board, will create further incentives to improve curriculum.

6) Evaluation of the quality of other educational resources

The quality of parental involvement and community support of a school are generally considered to be important element of good schools. Yet, polls indicate that many teachers feel disappointed by inadequate parental support for their children and schools, and that parents likewise are skeptical about the responsiveness and job performance of teachers.

This apparent decline in the mutual trust of teachers and parents occurs in the context of demographic changes in the student body. Some communities have experienced large increases in the number of poor, minority, non-English speaking, or foreign-born students whose families traditionally have not felt at ease in their children's schools. Most schools see significant increases in children from single-parent homes, children whose mothers work outside the home, and children affected by divorce. These changing demographics are well documented and discussed by public policy makers.

Changes in social mores are harder to observe and document, but trend-watchers also see a discernible tendency towards less parental involvement in child-rearing itself. Data here are very "soft," but it appears fewer families have dinner together on a regular basis, parents and communities have fewer rules and norms setting expectations for children's behavior, and children themselves spend more time in after-school jobs, watching TV, with friends, or in other activities that do not involve either school or family contact. The possibility exists that what James Coleman has called "social capital" is declining, and that its decline erodes some of the social underpinnings supportive of effective schools and educational quality.

7) Changes in the monitoring and supervisory mechanisms

The chief mechanisms for monitoring education are state systems for collecting and reporting data on school performance over time. Research and evaluation of the effects of individual reforms are largely missing. State legislators (often funding education generously in the face of budget constraints) have allocated money to programs they think will work. They have not usually set aside time, money, or staff for program evaluation, and if they had, new methods for evaluating the effects of multiple, simultaneous reform initiatives would have to be worked out. It therefore currently falls to the newly emerging performance reports to monitor the overall operation of schools and districts.

In some cases the claim is made that state performance measurement systems described above will tell whether reforms are working, but this is only true when judging the impact of reforms collectively. Ironically, scientific study monitoring the effects of discrete reforms is not funded, despite the fact that one of the major substantive elements of the reform movement is the increased monitoring of performance of the educational system as a whole.

State performance reports are new, and Americans are just discovering their advantages and disadvantages. They have potential for building public confidence. They help tell legislators,

parents, the public, and the business community how well, overall, schools are performing, and whether things are getting better or worse. They also can be used to improve schools by offering new information or perspective to educators on the comparative performance of their schools. And they can be used to direct technical assistance or additional resources and attention to areas that need improvement.

On the other hand, these reports have disadvantages. They report on the performance of people who are ambivalent about reform to begin with, and who often have much to lose depending upon what they report to the state. Overreliance on narrow educational measures like standardized tests creates incentives to focus on merely "looking good" - by trying to raise test scores (on weak tests), rather than by improving student learning.

The shortcomings of standardized tests are well known. Often they are poorly aligned with the curriculum of the school and with higher order skills. Recently a few states have begun collecting student writing samples, and Vermont, in November 1988, announced its plan to collect portfolios of student writing and math work as measures of student of achievement. The art of combining this kind of student achievement information with other indicators is at only an early stage. Many states do report student (and teacher)

attendance, graduation and dropout data. Some collect information from public universities and employers on the success of students after graduation. But when cash awards or incentives are linked to performance, the performance measure tends to revert to the "hard-measure" test scores, even where multiple other indicators are publicly reported.

Increasingly states report these kinds of data on educational "results" in the context of data on the social background of the students and schools. Americans are still struggling with how to make fair comparisons among schools serving students with varying backgrounds. Some states are developing ways to compare student achievement among demographically similar schools in "comparison bands," but much remains to be done. The development of new tests and measures must be undertaken before data will be available that realistically capture success at what educators, the public and others can agree are the important goals of education.

The widely acknowledged mismatch between existing measurement technology and what the public and profession judge to be important aspects of education has various consequences. Educators resist accountability measures because of the limitations of existing tests. Business leaders and elected officials, while recognizing the limitations of existing tests, insist that educational

performance data be made available to the public while the technology is being improved. State by state, each jurisdiction attempts to create accountability mechanisms that will be seen as fair, potent, and politically palatable. The unmet need for improved measurement technology is being felt by the nation and states attempting to create systems that are clear and fair.

The federal government initiated the move to compare educational performance among the states. In 1984, the Secretary of Education began publishing the "Wall Chart," a summary of multiple indicators of school performance. It now includes average college entrance (SAT and ACT) test scores, school completion data, and other information about educational resources and reforms by states in ways that demonstrate comparisons. These comparisons were recognized as crude since the states had not collected data using common definitions or processes. The political climate was such that instead of discrediting all attempts to make such comparisons, the states instead moved to control the process. Acting together in their own voluntary national organization called the Council of Chief State School Officers (CCSSO), the states established their own assessment project to compare multiple education indicators among the states. This voluntary effort is in the vanguard of attempts to get "good" data that will produce fairer and more meaningful comparisons among states.

At the national level, it has been decided to expand state-by-state administration of the National Assessment of Educational Progress, NAEP, the national test of student academic achievement. Originally designed to sample students only nationally and regionally, recent plans allow state samples to be tested and produce comparable state data. Thirty seven states have agreed to participate in a 1990 pilot test of 8th grade mathematics. The National Center for Educational Statistics has received increased funding to maintain and expand its work, which includes longitudinal studies of specific age cohorts (NELS), other surveys of teachers and schools (Schools and Staffing Survey), the recently redesigned Common Core of Data on elementary and secondary education, international comparisons of student achievement, and more. These data are then available, though in no formal way linked to the supervisory work of state and local decision makers.

States vary in their policies, but described above is the strong tendency to define, collect, and annually publish data on some emerging set of indicators. Increasingly these data may be linked to some policy of rewards or sanctions, but generally even these remain advisory to local authorities.

Local districts vary even more widely than states in the resources and expertise they have available to evaluate school

performance. Some large systems in major cities and wealthy suburban districts have budget and staff for quite sophisticated research and evaluation activities. These districts often pay higher salaries and hire better expertise than the state department of education which is nominally regulating them. Poor, small, or rural districts seldom have staff or data for evaluation. The strong research departments in richer systems frequently collect a great deal of data and do special evaluation studies. These data pertain to student testing, enrollment and attendance, school finance and spending, program monitoring, compliance with pertinent regulations, and a great deal more. Aggressive superintendents use these data to monitor school performance closely, sometimes linking them to principals' job ratings and many specific program decisions.

III. FIRST RESULTS OF EVALUATION AND MONITORING

It's hard to know whether either reform or new monitoring mechanisms are working. The political desire to get results is consistently complicated by difficulties in measurement. National data are not encouraging. NAEP results indicate improved performance of young and minority students, but little improvement for most. SAT scores have stopped their long-term decline but have

not demonstrated significant improvement. The proportion of students scoring at the top on the SAT seems to have actually declined. Few disagreed with the Secretary of the US Department of Education when he expressed concern at how little improvement could be demonstrated by NAEP or the Wall Chart.

To many, the most promising results of reform activity have not been evaluation results or measured changes in student achievement but the sustained commitment and expanded constituency for educational improvement.

Educators and local citizens are slowly becoming more involved in these activities. At the local level there is a paradox in how Americans assess education. Polls indicate that the public believe the performance of local schools is significantly better than the national average. They may readily acknowledge the national picture of mediocrity portrayed in A Nation at Risk and demonstrated by international comparisons of student performance, while believing at the same time that their own local schools are good.

It has recently been shown that this paradoxical view is fueled by misleading test results commonly reported by states and local districts. Test scores on private, nationally standardized

achievement tests have been shown to demonstrate that, based on technically outdated norms, almost all states and local school districts are "above average." This mathematically puzzling result has come to be called the "Lake Wobegon" effect. (Lake Wobegon is the name of a mythical small town in an American radio show in which "all the children are above average.") Even where these test results have been released without the intent to mislead the public, they have insulated parents and the public from more realistic evaluations of student performance.

In this context, international comparisons of student achievement fill a real need for Americans. There is a pervasive interest in the United States in comparative educational performance. The national government compares the states (the Wall Chart and NAEP); many states compare districts and schools (in their "report cards" to the public); some districts compare the performance not only of schools, but also of principals or teachers. US policy debate is routinely informed by international comparisons. In education, Americans are interested to know how our students perform compared to those of other economically developed nations.

Math achievement seems most free of cultural bias and therefore has been a special topic of interest. The International Assessment

of Mathematics and Science report, A World of Differences (1989), gave sobering indications that on an originally American instrument, American students performed poorly, even though they believed they were good at the subject. Such information provides an additional perspective from which to try to measure educational quality.

It is a perspective which teachers, politicians, parents and policymakers all see as relevant, appropriate, and important in our attempts to evaluate and improve educational quality.

BIBLIOGRAPHY

The National Commission on Excellence in Education, A Nation at Risk: The Imperative for Educational Reform, US Government Printing Office, Washington, D.C., (April) 1983.

Theodore R. Sizer, Horace's Compromise: The Dilemma of the American High School, Houghton Miflin Co., Boston, 1984.
and
Education Commission of the States, The Re:Learning Project, What is ReLearning?, 1987 and annually thereafter; also Re:Leaning Handbook, 1989.

Carnegie Forum on Education and the Economy, A Nation Prepared: Teachers for the 21st Century (The report of the Task Force on Teaching as a Profession), Washington, DC, 1986.

US Department of Education (OERI Study Group), Creating Responsible and Responsive Accountability Systems, and Measuring Up: Questions and Answers about State Roles in Educational Accountability, Government Printing Office, Washington, D.C., 1988

National Center on Education and the Economy, To Secure Out Future: The Federal Role in Education. Rochester, NY, 1989.

National Governors' Association, Time for Results, 1987, and Results in Education: The Governors' 1991 Report on Education, 1989

Educational Testing Service, A World of Differences: An International Assessment of Mathematics and Science, 1989.
and
Educational Testing Service, The National Assessment of Educational Progress, The Writing Report Card, 1984-88: Findings from the Nation's Report Card; and The Reading Report Card, 1971-88: Trends from the Nations' Report Card; and The Geography Learning of High-School Seniors, all 1990.

National Research Council of the National Academy of Sciences, Everybody Counts: A Report to the Nation on the Future of Mathematics Education, 1989, and Reshaping School Mathematics: A Philosophy and Framework for Curriculum, 1990, both National Academy Press, Washington, D.C.

Stephan S. Kagan and Richard J. Coley, State Education Indicators: Measured Strides, Missing Steps, Center for Policy Research in Education and Policy Information Center, Educational Testing Service, 1989.

US Department of Education, The Progress of Education in the United States of America 1983-89, Government Printing Office, Washington, D.C., 1990.

Joan Baron and Pascal Forgione, Better Measures: Improving Tests of Student Achievement in Connecticut, Government Printing Office, Washington. D.C. (forthcoming)