

How Schools Change Reforms: Redefining Reform Success and Failure

LARRY CUBAN

Stanford University

Because schools change reforms as much as reforms change schools, judging an innovation's success or failure has been, and is, no easy task. I identify three common criteria used by policymaking elites (effectiveness, popularity, and fidelity) and two less common ones used by practitioners (adaptability and longevity) and apply them to the two-decade-old Effective Schools school reform. What emerged as crucial in evaluating school reforms is what criteria are being used to make judgments, whose criteria they are, and how schools change reforms as they are implemented. The example of Effective Schools challenges policymakers and researchers to become more explicit about which criteria they employ to judge success and to understand how the journey of school reform is a story of constant adaptation that ultimately undermines the common criteria generally used to judge success and failure.

In 1906 in a town built by U.S. Steel on the shores of Lake Michigan, a new superintendent introduced an educational innovation that hundreds of school districts adopted in the next decade. Visitors traveled thousands of miles to talk to Superintendent William Wirt, sit in classrooms of the cheerfully decorated Froebel and Pestalozzi Schools, and marvel at how children of immigrants learned during the day while their non-English speaking parents attended classes at night. Even though U.S. Steel owned the property and employees largely ran the town, the educational experiment converged with company interests in providing what observers called a productive education for both white-collar and blue-collar employees. Progressives of the day, imbued with the pedagogically revolutionary ideas of John Dewey and Frederick Taylor's scientific management, wrote articles and books praising the combination of work and play, of school and community, of efficiency and civic-mindedness, that put the name of Gary, Indiana, on the early twentieth century map of school reform (Cohen, 1990; Cremin, 1961).

The Platoon School (or Gary Plan), as the innovation was labeled, was introduced in a remodeled elementary school holding children from kindergarten through the twelfth grade. Administrators divided the student body into two groups or " platoons." One platoon would be in the

classrooms or auditorium while the other would be in the basement where there were woodworking, printing, and other shops; upstairs in music, art, and play rooms; or outside on the playground. During the day, each platoon would change places, Giving each child academic, practical, recreational, and aesthetic experiences while thoroughly using the entire facility (Cohen & Mohl, 1979; Cremin, 1961).

While most elementary school children in cities at the turn of the century stayed the entire school day of five or six hours in a self-contained classroom with one teacher, Gary pupils worked with many teachers during an eight-hour day, even receiving released time for religious instruction. Moreover, because Superintendent William Wirt believed in tying the larger community of Gary to the school, adults (many of whom were recent immigrants working in the steel mills) would attend evening and weekend classes to learn to read and write English, hear lectures, and use various shops to learn industrial skills. Such a work-study-play-community school arrangement—a revolutionary shift in school organization and curriculum—made it possible to have many more students attend school since the schedule permitted all available space to be used by students during the day and adults at night. The Gary innovation spread swiftly across the nation. Educational pundits of the day applauded its success (Dewey & Dewey, 1915).

In 1918, however, two educational experts completed a foundation-funded study of the Gary schools. It praised some aspects of the platoon plan but raised serious questions about the quality of academic work and weak student performance on achievement tests. As a consequence, national interest in the Gary Plan ebbed considerably in subsequent years. By the mid-1920s, the innovation had receded and virtually disappeared from the national scene. In Gary, it lasted in some form or another into the 1940s (Cohen, 1990; Cremin, 1961).

Today Platoon Schools are largely forgotten. Yet the core notions of using buildings fully; offering a diversified curriculum combining academic subjects, practical tasks, and play in which students move to various parts of the school building; and having the school as an educational, social, and recreational center for adults have become mainstream features of elementary schooling. The Platoon School foreshadowed the modern elementary school (Cohen & Mohl, 1979).

Was the Platoon School a success, then, because it contained the embryo of the modern elementary school? Or was it a success because it became popular in the media and spread swiftly to hundreds of school districts? Or was it successful because it lasted for over four decades in Gary? Or was the reform a failure? After all, the Gary Plan soared in popularity, matured, and then vanished from the national scene—even after decades in the city

of its birth; few present-day school reformers would recognize the name or remember the program. The historical example suggests the puzzling ambiguity of, if not confusion in, evaluating the “success” of school reforms.

In this article, I will argue that most highly touted school reforms today are like the Platoon School. They are adopted and, as they are implemented, undergo changes that transform them in ways that few of the designers of the original reform could predict, or even claim ownership. Because schools change reforms as much as reforms change schools, judging an innovation’s success or failure is no easy task. Such doubts, however, have hardly prevented policy elites (then and now) from rushing to judgment by employing preferred standards to judge success. Allies in the media amplify elite opinions, often framing the making and unmaking of reforms as a political contest—who won and who lost (Broder, 1987; Fallows, 1996). Few media judgments of success are made about either the substance of the reform or its gradual metamorphosis. As a result, some promising reforms that evolve too slowly for impatient policymakers and media pundits are aborted while others that are earmarked as winners by opinion-setters in the horserace for public attention often fade and disappear. I argue that what becomes crucial in evaluating reforms within schools is identifying what criteria are used to make judgments, whose criteria they are, and how schools change reforms as they are implemented.¹

This article, then, is written primarily for an audience of policymakers, administrators, and researchers who seek school reforms and are in a position to pass judgment on their worth. Increased awareness of the dominant (and historically defined) framework of thinking about school reform as a Super Bowl where winners get glory and losers anonymity can reveal which criteria are prized over others and how such choices serve particular interests in judging winners and losers in the contest of school reform.

Why this audience? I am a rationalist at heart. I believe that lay and professional elites who shape educational policy in this country act as closet historians; they have a picture in their heads of the past, an interpretation that forms the basis for creating a better present. They use history whether they know it or not. If they can come to know the inexorable and enduring process of how schools change reforms historically and, further, come to understand clearly the factors that shape how judgments of success and failure have been (and are) made in America, they may be politically smarter in designing, implementing, and evaluating programs that will result in sustained improvements for children.

Suppose, for example, that, as time passed, an innovation that was once labeled a failure overcame early obstacles in its implementation and desired results came later than expected. Or suppose that the designers of a school reform judged the results to be unsuccessful because teachers

failed to adhere to the plan. Yet these very same teachers took the design, adapted portions of it to their classrooms, secured favorable results from their students, and privately called the reform a winner. How can a school reform be judged successful by one group and tossed out as a miserable defeat by another?

To advance my rationalist credo and to begin answering these questions, I examine first the implicit but dominant standards that policymakers and others have used to determine “success” and “failure.”² Next, I inquire into two other standards that have been frequently neglected by policy elites and media, and then apply all of them to a two-decade old reform initially aimed at big-city elementary schools that eventually was transformed into a national crusade called Effective Schools. For a country that has been undergoing unrelenting school reform for the last two decades and has embarked on a novel experiment involving the establishment of national goals, curricular standards, and tests, such contemporary and historical inquiries into the anatomy of judging success and failure may counter the persistent forgetfulness that so commonly afflicts American reformers (Knapp, 1997; Tyack & Cuban, 1995).

IMPLICIT STANDARDS IN DEFINING SUCCESS AND FAILURE IN SCHOOL REFORM

The primary standard used by most policymakers, media editors, administrators, and researchers is to ask whether a program’s intended goals were achieved. Have you done what you said you were going to do and can you prove it? In a society where “bottom lines,” Dow Jones averages, sports statistics, and vote-counts matter, quantifiable results often determine success. This is the *effectiveness standard*. For the last quarter-century the effectiveness standard, an outgrowth of a strong belief in professional expertise and technical rationality applied to organizations, has been used for schools to examine what students have learned in school and do after graduation by using proxy measures for both such as student test scores, rates of college attendance, and other numerical indicators (Majone, 1989; Schön, 1983).³

THOSE IN AUTHORITY SUBJECTIVELY SET THE DESIRED GOALS FOR REFORM AND THE MEASURES THAT WILL BE USED TO DETERMINE SUCCESS. FOR EXAMPLE, NATIONAL AND STATE POLICYMAKERS CONCLUDED BY THE LATE 1970S THAT AMERICAN PUBLIC SCHOOLS HAD DECLINED BECAUSE SCHOLASTIC APTITUDES TEST SCORES HAD PLUNGED DOWNWARD. THIS WIDELY REPORTED VIEW THAT SAT SCORES WERE RELIABLE MEASURES OF SCHOOL PERFORMANCE—EVEN THOUGH

TEST-MAKERS REPEATEDLY STATED THAT SUCH CLAIMS WERE FALSE—BECAME THE POPULAR WISDOM OF THE MOMENT, HELPING TO FUEL PUBLIC SUPPORT FOR STATES’ RAISING ACADEMIC REQUIREMENTS IN THE 1980S. IT DID NOT MATTER THAT THE TEST-MAKERS CALLED SUCH USE OF SCORES EXCESSIVE; WHAT MATTERED MORE TO PUBLIC OFFICIALS SEEKING POLITICAL SUPPORT AND THE MEDIA SEEKING AUDIENCES WERE QUANTITATIVE MEASURES THAT COULD BE USED TO ESTABLISH RANKINGS OF SCHOOLS, THEREBY CREATING EASILY IDENTIFIABLE WINNERS AND LOSERS (NATIONAL COMMISSION ON EXCELLENCE IN EDUCATION, 1983; U.S. CONGRESS, 1992).

Yet even here, test results proved ambiguous measures of a reform’s success. Just as the 1918 evaluation of the Platoon School slowed down considerably the spread of the Gary Plan, early evaluations of Title I of the Elementary and Secondary Education Act (ESEA) in the mid-1960s revealed so little improvement in poor children’s academic performance as to jeopardize congressional renewal of the program. Such evidence gave media critics and national policymakers hostile to federal intervention a reason to brand the War on Poverty programs as failures. Yet unpromising test scores were insufficient to overcome the program’s political attractiveness with constituents and legislators. Each successive president and Congress used that popularity as a basis for allocating eagerly sought funds to needy students in schools across the nation (Hoff, 1997; Kennedy, Jung, & Orland, 1986; McLaughlin, 1975).

Popularity, then, is a second standard that public officials use in evaluating success. The spread of an innovation, and its hold on the imagination of voters and professional educators, becomes an important criterion since fashionableness, as documented by opinion polls and media reports, easily translates into political support for top policymakers endorsing the reform. The rapid diffusion of special education, bilingual education, whole language, new math, and desktop computers in schools since the 1970s offers an obvious example of innovations that swept across the nation capturing both district and state school boards and practitioners’ attention. Few educators or public officials initially questioned the accelerating outlays of public funds for new programs because they were viewed as ways of coping with important unmet educational needs of children. The popularity of these reforms became evidence to support media editors’ and policymakers’ judgments that they were, at least at first, resounding successes (*Education Week*, 1993).

A third common standard policymakers and administrators use to judge

success is assessing how well the innovations mirrored what reformers intended. This *fidelity standard* aims at assessing the fit between the initial design, the formal policy, the subsequent program it spawns, and its implementation. How, they ask, can an effectiveness standard ultimately be applied if the reform departs from the blueprint? If the federal government, for example, funds a bilingual program that it wants to be implemented because it has proved to be effective elsewhere, state and local implementers must follow the original design of the program as they put it into practice or else the desired outcomes will not be achieved.

The fidelity standard places great importance on implementers, who invariably are teachers and principals, following the designer's blueprint. When practitioners add, adapt, or even omit features of the original design, then policymakers, heeding this standard, say that the policy and program cannot be determined effective because of these changes (Emrick, Peterson, & Agarwala-Rogers, 1977; Emshoff et al., 1987; Snyder, Bolin, & Zumwalt, 1990).

From where do these interrelated and dominant criteria of *effectiveness*, *popularity*, and *fidelity* arise? Policymakers derive the criteria of effectiveness and fidelity from a view of organizations as instruments for solving problems through top-down authority, formal structures, clearly specified roles, and technical expertise to fulfill one primary task: achieve desired goals set by organizational decision makers. Within organizations where rational decision making and control are prized, two linked questions are asked: Have the prescribed procedures in performing the task been followed (fidelity) and have the goals been achieved (effectiveness)? Hence, in judging reforms, those who carry out the changes must be faithful to the design before the cardinal standard of effectiveness in achieving goals is invoked (Scott, 1987, pp. 57-78).

But how have these beliefs about rational organizations become embedded in these standards? The growth of professional expertise, or what Donald Schön calls "technical rationality" and Giandomenico Majone names "decisionism," have come to be anchored in university-credentialed knowledge in the twentieth century, especially since World War II. This expertise, often grounded in the natural, physical, and social sciences, is located in professional training programs at universities. Rather than favoring practitioner expertise anchored in schools and classrooms, public officials, administrators, and researchers use this scientifically grounded knowledge to evaluate whether reforms have succeeded (Majone, 1989, pp. 12-20; Scheffler, 1984; Schön, 1983, pp. 21-36).

Popularity as a measure of success, contrary to the effectiveness and fidelity standards, derives from the political nature of public institutions and the astute use of symbols to convey values. Schools, for example, are totally dependent on the financial and political support of local communi-

ties and the state. Taxpayer support or opposition for bond referenda or program initiatives taken by school boards is often converted into political capital or deficits at election time. The media play an important role in defining a school reform as news, especially in the slant editors take in portraying the innovation. Potent symbols (e.g., school prayer, the computer) become shorthand terms for highly prized values to mobilize supporters. Whether an innovation spreads and captures public and practitioner attention becomes a strong basis for evaluating its success (Kliebard, 1990; Pfeffer & Salancik, 1978; Wirt & Kirst, 1989).

Although political and rational considerations dominate policy talk on school reform, they, or the primary symbols, are seldom made explicit. Hence, these mainstream standards are infrequently debated publicly, much less questioned. Unexamined acceptance of effectiveness, fidelity, and popularity thus avoids inquiring into the slippery issue of whose standards are being used to judge success. In asking the question of whose standards will be used, neglected alternatives for judging success and failure emerge.

WHOSE STANDARDS ARE USED?

When federal, state, and local policymakers (and media reporters following their lead) talk about reforms and use these criteria to determine success, their judgments carry much weight because elected state and district policymakers are authorized to act in behalf of the community. This political legitimacy influences voters.⁴

But what about practitioners who are the foot-soldiers of every reform aimed at improving student outcomes? Their criteria diverge considerably from what policymakers, administrators, and researchers would prize. Practitioners bring moral and service values inherent to teaching that differ from the technical and scientific values that policy elites possess. Practitioners accumulate expert knowledge about students and how to teach skills and subject matter that few researchers or policymakers plumb. From these values and expertise emerge standards for judging success and failure that diverge from those of policymakers and researchers. Of course teachers seek improvement in students' performance and attitudes but what teachers count as significant results are seldom test scores but attitudes, values, and actual behavior on academic and nonacademic tasks in and out of the classroom. What becomes especially important to teachers is how they can put their personal signature on the mandated reform and make it work for their students and themselves (Berman & McLaughlin, 1978; Cohen & Spillane, 1992; Cuban, 1993; Elmore & McLaughlin, 1988; McLaughlin, 1978; Wolcott, 1977; Welker, 1992).⁵

To reformers, however, alterations in their design and variations in prac-

tice become evidence of decay and failure; to teachers, the very same modifications are viewed as healthy signs of inventiveness, active problem solving, and a precondition for determining effectiveness—as they would define it. To one, the end-product is everything; to the other, getting to the end-product is as important as the outcome. This practitioner-derived *standard of adaptiveness* (the flip side of the fidelity standard) becomes an essential prior condition for other criteria to be applied.

Another standard occasionally used by practitioners and social scientists to judge success or failure is to ask whether the reforms have endured (Kirst & Meister, 1985). This *longevity criterion* is a plausible standard. After all, in public schools, where so many innovations last no longer than warm breath on a cool window, a program that survives is a signal achievement to which officials can point with pride. The comprehensive high school, vocational education, junior high schools, kindergartens, and (a more prosaic example) overhead projectors were once reforms and are now pervasive after almost a century. They were not written in snow. Durability, then, is another standard by which success or failure can be determined.

As alternative criteria, adaptability and longevity derive from a set of assumptions about organizations different from the ones dominating mainstream policy thinking. These assumptions find organizations less rational, less tightly coupled, and shaped more by external political, social, and demographic factors as well as behaviors of the people within them. Unpredictable external factors and a web of larger social beliefs intersect with an informal side of the organization—who talks to whom, the stories that people tell—to fashion workplace behavior often producing unintended consequences for rationally inspired policies. While formal goals are important, they are easily displaced by an abiding concern for accommodating to societal beliefs and organizational survival. Adaptations in roles and structures occur naturally and often as people within an organization cope with the inexorable range of external and internal problems that ebb and flow. Organizations, then, evolve in order to survive (Meyer, 1992; Meyer & Rowan, 1978; Scott, 1987, pp. 79-101).

But why are these standards of adaptiveness and survival, ones that practitioners would prize in a school reform, seldom invoked by policy entrepreneurs? The question boils down to one of power and status: Whose standards count? When national, state, and district policymakers, holding implicit views of how organizations work, using research findings to bolster their decisions, and having access to media, place their weight behind reforms, legitimacy in making changes rests with those at the top of the organization, not those at the bottom whose organizational views, expertise, and values might differ. Without the cachet of scientific expertise, access to top officials, or easy entree to reporters, individual teachers are

stuck. Collectively, teachers have organized into unions and, more recently, asserted their political clout through taking explicit positions on school reform. Yet in making policy and judging success, unions play a limited role. Thus, when individual teachers do choose to make adaptations, they do so unobtrusively or, in some cases, engage in guerrilla warfare with administrators. Organizational legitimacy, use of scientifically derived data, and power often determine whose criteria are used to judge success and failure. Finding out whose standards are being used becomes an important initial step in making judgments about reforms.⁶

Thus far, I have argued that assessing school reforms goes beyond the policy elites' using unarticulated standards of effectiveness, popularity, and fidelity. I have identified alternative standards of adaptiveness and longevity. Applying all of these standards to a reform that began in the late 1970s offers a chance to examine how reforms are transformed as they are implemented in schools and whose standards of judging success and failure come into play. I first describe the Effective Schools movement itself, the shifting environment that gave shape to the movement, and the degree to which its ideology and programs were put into practice.

EFFECTIVE SCHOOLS RESEARCH AND REFORM

In the mid-1970s, a small number of researchers began working to disprove the then mainstream wisdom that what largely determines students' academic performance—as measured by standardized achievement tests—is family background. Research studies on the inability of public schools to overcome the effects of poverty and race had led national policymakers to call for reduced federal funding of programs (Coleman et al., 1966; Jencks et al., 1972; Ravitch, 1983).

Within this social milieu of pessimism about the inadequacy of public schools to make a difference in the lives of poor minority children, the Effective Schools movement was born. Believing deeply in the value of equity and expecting that urban schoolchildren would be especially harmed were such a prevailing consensus of opinion among policymakers to persist, this small band of activist researchers led by Ron Edmonds identified a handful of big-city schools enrolling large numbers of low-income minority children who scored higher on standardized achievement tests than would have been predicted by the socioeconomic status (Brookover & Lezotte, 1979; Edmonds, 1979; Firestone, 1991; Weber, 1971).

These researchers-cum-reformers extracted from these schools certain factors (e.g., clearly stated academic goals, principal's instructional leadership, concentration on basic academic skills, strong emphasis on maintaining order in school, frequent monitoring of academic achievement, con-

necting what is taught to what is tested, etc.) that they linked to the students' higher-than-expected academic performance on standardized tests. By the early 1980s, lists of essential factors had been consolidated into various Effective School models that swiftly spread to many big cities across the country. In creating the Effective Schools' ideology and model programs, Ron Edmonds and others prized four values: All children, regardless of background, can learn and achieve results that mirror ability, not socioeconomic status; top-down decisions wedded to scientifically derived expertise can improve individual schools; measurable results count; and the school is the basic unit of reform (Cohen, 1982; Edmonds, 1979; Firestone, 1991; Levine & Lezotte, 1990; Purkey & Smith, 1983; Ralph & Fennessy, 1983; Stedman, 1987).

At about the same time, another national impulse for school reform had emerged with a barrage of reports about failing American schools. U.S. presidents, corporate leaders, and critics blasted public schools for contributing to a globally less competitive economy, sinking productivity, and jobs lost to other nations. The United States, as *A Nation at Risk* put it, had educationally disarmed itself in a hostile economic war. "If only to keep and improve on the slim competitive edge we still retain in world markets," the report said, "we must dedicate ourselves to the reform of our educational system." States raised graduation standards, created curriculum frameworks, initiated accountability plans, and focused on teacher performance. Excellence, not equity, became the slogan of the day (National Commission on Excellence in Education, 1983, p. 7).

In the wake of this charge of mediocrity and the resurgent belief, propagated by policy elites, that public schooling did indeed make a difference to the nation's economic health, Effective Schools models focusing on academic achievement and accountability in minority-populated schools became popular. Suburban districts and entire states hastily installed programs that identified five, seven, or fourteen factors of effectiveness and laid out careful designs for schools to follow in implementing different models in such diverse places as Georgia, Minnesota, and California. Amendments to the Elementary and Secondary School Act in 1988 included specific reference to Effective Schools research. Federal officials directed Title 1 schools to consider that research when designing their programs and told state agencies receiving federal funds that they had to set aside monies to help schools establish programs based on the factors that Ron Edmonds and his colleagues had identified (Bullard & Taylor, 1993; Firestone, 1991; Murphy, Hallinger, & Mesa, 1985; Silver, 1994).

Even with this energetic state and federal activity in bending the Effective Schools movement to the newly hatched economic imperative, by the end of the 1980s, policymakers spoke publicly less and less of "effective schools" and

far more about "restructuring," "site-based management," "professional development of teachers," and "systemic reform." Other beliefs about how school reform should be implemented were emerging among federal and state policymakers, who viewed the reform of schooling as a far larger project involving national goals, curriculum, and testing rather than a local school-by-school venture (Cohen & Spillane, 1992; Murphy, 1992; National Governors' Association, 1991; Ogawa, 1994; Smith & O'Day, 1990).

Many local and national reformers—believing deeply in the value of equity—still pursued the same goal of improving the academic fortunes of low-income minorities but used different approaches consistent with the new reform surge. For example, university professors such as James Comer, Robert Slavin, and Henry Levin launched national programs of school-by-school reform in the late 1980s and early 1990s that had targets similar to those of the Effective Schools movement and even a similar focus—using the school site—but with different strategies for improvement (Chenoweth, 1992; Comer, 1988; Levin, 1989; Slavin, Madden, Karweit, Levermore, & Dolan, 1990). In doing so, they borrowed heavily from an earlier and heretofore neglected tradition of school improvement through planned change that was heavily dependent on faculties, principals, and parents' making managerial and curricular decisions at each school site. Also, state and district efforts to use vouchers for private schools and parental involvement in directing schools leaned on Effective Schools research to improve big city schools (Chubb & Moe, 1990; Firestone, 1991; Fullan 1991; General Accounting Office [GAO], 1989).⁷

If Edmonds's work in the late 1970s spawned a cottage industry of Effective Schools aimed at ensuring equity for low-income minority students, the linkage of public schools to the economy in the early 1980s, in effect, nationalized the Effective Schools movement while dropping the brand name. Federal and state policymakers, believing in education as the engine for the economy and using the same Effective Schools research, sought a broader and speedier impact on the nation's schools than the slower school-by-school approach. They called for national goals, curriculum, and tests.

Throughout the 1980s, U.S. Secretaries of Education William Lamar Alexander and William Bennett talked about "good schools" and "effective schools" in the same breath. Assistant Secretary of Education Chester Finn spearheaded the production of a popular pamphlet called *What Works* that drew directly from the effective schools research. President Bush and his policy advisers organized the nation's governors to endorse six national goals in 1989. A movement toward national goals, curriculum, and tests received the stamp of approval from a Republican president who styled himself the "Education President" (U.S. Department of Education, 1987; Wirt & Kirst, 1989).

When administrations changed, top policy advisers to Democrats also drew from the same well of Effective Schools ideology and research. Within an article that became a script for national and state policy elites, for example, Marshall Smith and Jennifer O'Day (1990) made clear that the ancestry of "systemic reform" was in the Effective Schools movement. "The most effective schools," they said, "maintain a schoolwide vision or mission, and common instructional goals which tie the content, structure, and resources of the school together into an effective unified whole." Moreover,

the school mission provides the criteria and rationale for the selection of curriculum materials; the purposes and the nature of school-based professional development, and the interpretation and use of student assessment. (p. 235)

Smith's passion for both equity and excellence in schooling, his earlier work that synthesized Effective Schools research, and his experience in Washington policy circles made him an ideal candidate to serve the new U.S. Secretary of Education, Richard Riley, in 1992. As Undersecretary of Education and later Deputy Secretary, Smith drafted many of the Clinton administration's bills on national goals and testing and accelerated the shift toward nationalization of the Effective Schools movement without once using the phrase *effective schools*. The bipartisan reliance of policy elites on Effective Schools research renewed attention to the earlier point I raised of whose standards for evaluating success are invoked.⁸

Whose standards for success are being used? Although three different groups have vied for attention in judging the success and failure of effective schools, one group's standards easily dominated the others. Federal, state, and district policymakers (including their accompanying foundation- and corporate-supported analysts and evaluators), amplified by the media, are the opinion setters. These policy entrepreneurs see school organizations as goal-driven political entities with formal structures and leaders exerting top-down authority. These entrepreneurs juggle highly prized values of equity, efficiency, excellence, and getting reelected. They are also especially sensitive to strong public expectations for school accountability and test scores. Hence, this policy elite favors the standards of effectiveness, fidelity, and popularity—even when they conflict with one another. Because the world they inhabit is one of running organizations or advising those who do, their authority and access to the media give them the leverage to spread their views about how to judge success (Darling-Hammond, 1993; Kahne, 1996; Ogawa, 1994; Rotberg & Harvey, 1993).

Although effectiveness, fidelity, and popularity criteria govern mainstream policy formation, analysis, and evaluation, as applied to public schooling these standards are seldom applied to other social and political

goals that schools are expected to achieve. Since the early 1980s, the dominant view among national policy elites that education is the engine driving the economy has overshadowed other common purposes of public schools in a democracy such as building citizens, seeking social justice, and providing a psychologically and emotionally healthy experience for children. Instead, "effectiveness" and "popularity" of a school reform in public schools have often been shrunk to quantifiable student outcomes and even those student outcomes have been further miniaturized to mean higher scores on standardized achievement tests (Kahne, 1996).⁹

Since the initial Effective Schools research in the early 1970s relied on standardized achievement test scores to identify low-socioeconomic-status schools that exceeded academic expectations, using test results seemed appropriate. Moreover, for the loose alliance of federal, state, and local policymakers and administrators, these numbers were useful tools to display to an increasingly skeptical public that schools can succeed and be held accountable for their performance. With newspapers publishing school-by-school standardized test scores since the late 1960s, the popular notion that schools were engaged in a contest with clear winners—those at the top of the list—and losers—those at the bottom—became an easy segue for a public that eagerly followed the World Series and national rankings of football teams.

Yet the widespread notion of improved test scores as a valid measure of school improvement remained slippery. Researchers and psychometricians have often explained to policymakers and administrators the technical difficulties of using standardized achievement test scores to judge a school's performance. They have pointed out how even the lowest-performing schools can show improvement simply by shifting the metric used, changing the definition of what constitutes improvement, and preparing children for tests. They have pointed out how reliance on test scores can steer a district's programs in undesired directions (Brophy & Good, 1986; Cuban, 1984; Purkey & Smith, 1983; Ralph & Fennessy, 1983; Rowan, Bossert, & Dwyer, 1983).

The world that policy elites inhabit, however, is one driven by values and incentives that differ from the worlds that administrators and practitioners inhabit. Policymakers respond to signals and events that anticipate reelection or reappointment and media coverage. They consider the standards of effectiveness, fidelity, and popularity rock-hard fixtures of their policy world (Elmore & McLaughlin, 1988; Kahne, 1996).

Do practitioners share the same standards? Some urban educators who remain committed to the dreams of the founders of Effective Schools and the importance of equity for low-income minority children still adhere to the design where five or more factors are considered to be correlates of

Effective Schools. For these true believers, the fidelity standard is the first step toward the effectiveness standard of higher test scores (Bullard & Taylor, 1993; Levine & Lezotte, 1990).

Although most big-city teachers and principals have expressed initial support for Effective School models because of the attention and occasional extra resources they received after years of neglect, many teachers and principals have also expressed strong skepticism about the validity of test scores as a measure of either their effects on children or the importance of their work (Herndon, 1968; Kohl, 1968; Wise & Darling-Hammond, 1985)

Such practitioners use different criteria to determine their success. They are just as interested in student outcomes as are policymakers, but the outcomes differ. What skills, content, and attitudes have my students learned beyond what is tested? Are our students decent people? Are the parents working with us rather than against us? To what extent is the life lived in our classrooms and schools healthy, democratic, and caring? Furthermore, they ask: Can these innovations be bent to our purposes? They judge their success by figuring out how to adapt mandated materials they are given for their classrooms to their daily routines with students. Incentives that drive practitioners are located in the satisfactions they gain from working with students and one another. These often-ignored questions are seldom heard (Holt, 1964; Kahne, 1996; Sizer, 1984, 1993; Wigginton, 1985).

A third set of standards comes from researchers. Researchers judge success by the quality of the theory, research design, methodologies, and usefulness of their findings. In most cases, by the mid-1980s, researchers had pronounced the Effective Schools research seriously flawed in theory, design, and methodology, although a few did point out redeeming qualities in that body of literature. These researchers' standards have been selectively used by both policy elites and practitioners in making judgments about establishing Effective Schools programs or any of the subsequent incarnations in the late 1980s and early 1990s (Brophy & Good, 1986; Levine & Lezotte, 1990; Purkey & Smith, 1983; Rowan, Bossert, & Dwyer, 1983; Stedman, 1987; Teddlie & Stringfield, 1993).

So at least three sets of standards for judging success are available. Practitioner- and researcher- derived standards have occasionally surfaced and received erratic attention from policy elites. But it is this strong alliance of state and federal policymakers and associates in the corporate, foundation, and media worlds with their implicit standards of effectiveness, fidelity, and popularity that continues to dominate both public debate and school reform agendas.

Before we apply the dominant but unarticulated standards of effectiveness, fidelity, and popularity and the alternative ones of adaptability and

longevity to Effective Schools programs, an essential question must be asked: Were the Effective Schools ideology and program actually implemented? A Government Accounting Office report found in 1988 that two out of every five districts in the nation had installed some version of an Effective Schools program and among the fifty largest city and county systems, thirty-nine had done so (GAO, 1989). "Some version" means that local administrators and teachers adapted these programs to local circumstances and that the Effective Schools concepts themselves underwent changes as they became institutionalized. Reforms may change schools but schools also change reforms. So within a decade, some form of Effective Schools ideology and program had spread beyond big cities to suburbs and rural districts, mostly, however, in elementary rather than secondary schools.

The very nature of secondary schools made the grafting of Effective Schools research findings—largely anchored in studies of elementary schools—cumbersome. Secondary schools' size, multiple goals, differentiated curriculum, and faculty's subject-matter training made the five (or six or ten) factors of school effectiveness an awkward fit (Firestone & Herriott, 1982). For urban and suburban junior high schools, the reemerging middle-schools movement in the early 1980s (after its initial appearance in the mid-1960s) has competed with Effective Schools ideology for a version of what constitutes a "good" school. Middle-school beliefs in team teaching of core subjects, teachers acting as advisers, a pedagogy grounded in active learning and small-group instruction, and large time blocks set aside for classes are anchored in a neoprogressive philosophy that was, at best, skeptical and, at worst, hostile to some of the Effective Schools factors and standardized achievement testing. Thus, middle schools imbued with this neoprogressive ideology, especially those in affluent suburban districts, made a dry, inhospitable seed-bed for Effective Schools (Carnegie Council on Adolescent Development, 1989; Lewis, 1990).

Although determined efforts were made to graft lists of Effective Schools factors onto high schools in the mid-1980s, the graft seldom took because of competing ideologies among reformers about the social and political purposes of public schools. No consensus among policy elites, high school practitioners, and researchers yet exists about what constitutes a "good" or "effective" high school (Boyer, 1984; Coleman, Kilgore, & Hoffer, 1982; Corcoran & Wilson, 1985; Lightfoot, 1983; Muncey & McQuillan, 1996; Sizer, 1984, 1992).¹⁰

Few of that hardy band of Effective Schools reformers in the early 1970s could have foreseen that the basic values they professed about all children learning, the importance of academic achievement, and accountability would be seized a decade later by corporate leaders, foundation officials, governors, and presidents. Effective Schools reformers had fought the then

“wisdom” of policy elites that schools made little difference in the lives of big-city children. Now in the late 1990s, national goals, curricular standards, and tests dominate talk of school reform. Test scores remain the coin of the national realm. Coordinating district, state, and national goals and curriculum, texts, and tests has become the challenge of the decade. These reformers two decades ago could not have foreseen how fears of decline in economic competitiveness would have driven policy elites to link improved schools to an improved economy and thus have nationalized the ideology of Effective Schools.

After almost two decades of Effective Schools research and implementation, the legacies are plain to see in the policy announcements and dominant beliefs embedded in state and national agendas for improving schools: Regardless of background, all children can learn; the school site, rather than the classroom, is the unit of reform; administrative leadership is essential for improvement; clear and measurable goals are critical; top-down decisions and expertise make a difference in school improvement; academic improvement in schools requires linking basic skills instruction to the school’s goals, curriculum, daily lessons, and tests; and finally, what counts as success is students’ scores on standardized achievement tests (Murphy, 1992; Ogawa, 1994; Shipps, 1987).

APPLYING THE STANDARDS

In applying the standards most commonly used by federal, state, and local policymakers (effectiveness, fidelity, and popularity) to Effective Schools reforms, the judgments are mixed. There are individual urban and rural elementary schools that have clearly met the effectiveness criterion for years (Levine & Ornstein, 1993; Teddlie & Stringfield, 1993). Depending on the evidence used, Effective Schools reforms may have also played a role in the steady increases in basic skills as measured by the National Assessment of Educational Progress over the last twenty years, although the gap in scores between minority youth and whites on these tests remains as wide as it was two decades ago (Hoff, 1997). The persistent presence of urban schools scoring in low ranges on these tests, the instability of test scores over time in schools, and the growing appeal of vouchers in Milwaukee and Cleveland and of private firms running public schools suggest that any judgment of success by the effectiveness standard depends heavily on the setting and is, therefore, spotty (Bullard & Taylor, 1993; Levine & Lezotte, 1990).

Fidelity as a standard of judging success disappeared early in the movement for Effective Schools as districts and staffs swiftly tailored the different factors of effectiveness to their setting. Moreover, other reforms blended nicely with Effective Schools. Rodney Ogawa (1994) documented

how key players who promoted school-based management saw it easily meshing with parts of school effectiveness initiatives. Similarly, the federal government’s enthusiasm for selected pieces of Effective Schools research and practice (e.g., Title I amendments; the notion of national goals tied to national assessments) further suggests how little quality control the designers exerted over the reform. Were fidelity-to-design the standard for judging success, mark it a failure.

On the measure of popularity, score Effective Schools a success. Once researchers and reformers identified essential factors correlated with increased academic achievement, the innovation spread swiftly in urban districts. Since the early 1980s, when national reports, presidents and governors, and top educational policymakers adopted the idea that schools’ productivity as measured by test scores was instrumental to a competitive global economy, Effective Schools ideology and essential features were drafted into a national crusade for excellence.

Similarly, were adaptiveness to be used as a criterion, then one could easily argue that the Effective Schools program has been strikingly successful. The ideology of the movement—its central commitment to all children learning, its focus on clear goals and the school as the natural site for improvement, and its embrace of standardized test scores as a proper measure of academic performance—has become mainstream thinking among policymakers and popular rhetoric among administrators and practitioners. That small-town, suburban, and state administrators (e.g., Beaverton, Oregon; Frederick, Maryland; Casper, Wyoming; Minnesota Effective Schools Program) embraced the school reform also suggests its plasticity with pupil enrollments very different from the ones in which the movement began (Bullard & Taylor, 1993). Adaptiveness also means that school districts and individual schools alter the reform as it is implemented. Applying the standard of adaptability, mark Effective Schools a resounding success.

What about longevity? Is almost twenty years sufficient time to say that Effective Schools reform has lasted? Certainly, it has outlasted such once-popular innovations as science and math curriculum reforms and instructional television (ITV) in the 1960s and language laboratories of the early 1970s. What has given Effective Schools longer staying power, ironically, is its loss of name and focus on low-income minorities. Just as the Platoon School lost its name and attachment to the Gary Plan yet survived for decades in varied districts, the central beliefs and selected features of Effective Schools programs, shorn of identifying attachments to big-city schools in the late 1970s, have continued to permeate reforms in the mid-to-late 1990s.

Applying, then, the standard of effectiveness, there is some evidence of partial success (e.g., individual schools that have performed consistently

above expectations; test-score evidence of gains in basic skills for urban children) but no clear long-term trend of student improvement in academic performance. For popularity and adaptiveness, there is no question that both have been in full display. Effective Schools programs have been tailored to meet school settings different from those for which they were originally conceived. If some Effective Schools reformers disliked the constant modifications and dilution of their correlates of effectiveness, other administrators and practitioners enjoyed the reform's flexibility. Its resiliency and popularity have given the ideology and program a remarkable reach. However, such plasticity and popularity—a reform for all seasons—mean that whatever ideological and programmatic bite it contained softened considerably as it spread to small towns, suburbs, states, and the embrace of the federal government.

Hence, as Effective Schools became a generic program of improvement, even losing its brand name, its potential to meet the standard of effectiveness lessened considerably. Schools modified the innovation. It is, of course, precisely that capacity for adaptiveness and enduring popularity that has permitted Effective Schools to survive for over twenty years. The striking success of Effective Schools on both criteria may suggest to reformers that change within established institutions such as schools occurs far more than informed opinion suggests.

So a mix of partial successes and failures for the Effective Schools reform emerges from this analysis—depending on what vantage point reformers take. For those Effective Schools promoters who prize fidelity, departures from the original design are marks of failure. For pragmatists among such committed reformers, however, the partial success in meeting the standard of effectiveness and total success in being adaptive, popular, and surviving might leave them feeling proud of what has occurred since the late 1970s and optimistic for the future since such changes must to be counted in decades, not years.

SO WHAT?

Now what is gained by exploring the standards by which judgments about success and failure are determined and applying them to the Effective Schools movement? In other words, so what? I have argued that it is crucial to evaluate school reforms by identifying the criteria used to make judgments, whose criteria they are, and how schools change reforms as they are implemented. What determined which standards came into play and whose standards were applied in evaluating success and failure were noted in the example of the Effective Schools movement. I have asserted that the choice of standards depended on informal networks of policy

entrepreneurs, using symbols and scientific expertise, and having access to media. These policy elites, then and now, ultimately judge the success of school reforms. I argued that the standards they use were too often unarticulated and should be made explicit for public debate.

Why? Because the lay and professional elites, with their political savvy, use particular criteria for evaluating success and failure to frame the agenda of public problems. In doing so, they dominate which directions educational policy and funds take, yet they are largely uninformed about past reforms. For policy elites to know that the journey of previous school reforms is a story of constant adaptation and the power of established institutions to bend innovations to fit their contours could lead amply funded policymakers to anticipate and encourage adaptations of their plans. They can design hybrids that meld the new and old in ways that leave room for modifications (Tyack & Cuban, 1995).

If hybrid reforms encourage adaptations, then policy elites, especially the experts they hire to evaluate reforms, can move away from the usual thumbs-up or thumbs-down verdict on a reform, and point to improvements in practice. Policy entrepreneurs must come to understand the inevitable variations in school-by-school change. If reforms change schools—as some have—so do schools clearly change innovations. As long as that fact remains either a mystery or a design failure to policy entrepreneurs it remains a serious problem that their cadres of analysts and advisers must face.

The political stakes for students, parents, and practitioners are high when presidents of the United States, U.S. secretaries of education, governors, legislators, academics, and foundation and media officials seek to reform the nation's schools. If elite reformers can come to know how schools have inexorably changed reforms in the past and, further, come to understand clearly the factors that shape how judgments of success and failure have been (and are) made in America, they may not only get smarter in designing, implementing, and evaluating programs but become more accountable for what happens in schools.

Notes

1 By "policy elite" I mean loose but intersecting networks of public officials, corporate executives, foundation officers, and academic experts who use both public and private funds to run projects and write papers consistent with their versions of reforms. As policy entrepreneurs, they have ready access to media and the capacity to frame problems and set a public agenda for their discussion (Kingdon, 1984). As advisers to top public officials, corporate leaders, and foundation executives, they are analysts whose values reflect their patrons. Political party labels do not define these elites, although there are clearly Republican and Democratic members who carry their affiliation on their sleeve and, when administrations change, move in and out of office. I do not use the word as a populist jab at appointed officials. Nor do I suggest conspiratorial groups secretly meeting and designing

scripts for action (Safire, 1997). I suggest only that these overlapping networks of like-minded and decent individuals share many of the same values and tastes and seek school improvements aligned to those values. They convene frequently in different forums, speak the same policy talk, and are connected closely to sources of public and private influence in governments, media, businesses, academia, and foundations. They help create a climate of opinion that hovers over no more than a few hundred influentials in policymaking. Familiar with the ways of the media, these policy elites extend and shape that climate of opinion by closely working with journalists who report what they say, write, think, and do (Fallows, 1996; Reich, 1992). Few members of these loosely connected policy elites, however, have had direct or sustained experience with either principals or teachers, much less the act of teaching. Yet their recommended mandates touch the daily lives of both educators and children.

Earlier elites in education have been examined to determine their influence on policy formation. For a polemical discussion that contains inferences more than direct evidence of educational elites that dominated schools before and after World War II (see Bestor, 1953). For a balanced and well-documented examination of the “interlocking directorates” among educational progressives in the early decades of this century and their influence on both policy and practice, see Tyack and Hansot, 1982.

Thus far, with one exception, the interlaced networks of academics, foundation executives, corporate leaders, and national and state educational policymakers in the 1980s and 1990s await their chroniclers and analysts. The one exception is Rodney Ogawa’s (1994) map of how the reform of school-based management was initiated, spread, and became adopted by networks of policymakers, teacher union officials, and academics beginning in the mid-1980s. He does not use the word *elite* in his analysis but his careful and systematic examination of the documents, actors, and linkages powerfully illustrates how a small group of policy entrepreneurs promoted a school reform officially and saw its institutionalization in the nation’s political agenda for school improvement.

2 The quotation marks around these words are to underscore their shifting, ambiguous, and historically based character.

3 Using numerical measures to determine organizational effectiveness has a long history, of course. I offered one early and unusual example in the report on the Gary, Indiana, schools in 1918. The accelerated incidence of “indicators” for school effectiveness since the 1960s is nicely picked up in Murnane (1987, pp. 102-103). Having some set of quantitative measures to assess the degree to which the goals of a school, district, state, and nation have been achieved has become pervasive in American education in the last quarter century. In California, for example, every school reports to parents the test scores, rates of attendance, suspensions, and other data on an annual basis. The State Department of Education has an annual report card issued to the public recording different measures of effectiveness.

Schools are not alone in this use of numbers by social scientists to measure organizational effectiveness. Announcements of leading economic indicators often lead the nightly television news programs. For hospitals the efficiency of nonprofit and for-profit institutions has been measured by numbers of employees, empty beds, and other indicators. One popular measure of hospital effectiveness (rather than efficiency) has been mortality rates (see Lewin, 1987, p. 29; Tolchin, 1990, p. A11). In the airline industry, on-time arrivals has become a measure of an airline’s effectiveness. Other indicators have been used and combined into an index that rates each company’s performance (see Goodman, 1992, p. A5). And, of course, in the auto industry, miles-per-gallon has been a standard measure of efficiency that still gets full-page displays in most of the nation’s newspapers when annual figures for each car manufacturer are released (see, for example, *San Jose Mercury*, 1986, p. 8A).

I offer these examples across U.S. institutions to suggest the pervasiveness of metrics in the service of determining organizational effectiveness. Schools, then, are just another social institution where social scientists have applied a dominant standard of quantitative measures to assess whether the organization is reaching its goals.

4 The ambiguous but nonetheless crucial role of the media in publicizing official definitions of success underscores the question of whose standards will dominate what is labeled a success. One clear example of this arose during the Gulf War in 1991. Howard Kurtz (1993) describes how Pentagon officials dealt with the press on the success of bombing raid: “Reporters were upset about the military’s initial claim that 80 percent of its bombing raids had been successful. Officials later admitted that the number of ‘sorties’ had included reconnaissance, refueling and other support missions, and that ‘success’ had been defined as the pilot releasing his payload and returning to base, regardless of whether the target was destroyed.” Kurtz comments: “We had been snookered again” (p. 227).

5 For a clear instance of the constant struggles of teachers to modify district and state reforms intended to alter classroom instruction, see Cohen and Ball, 1990, pp. 249-256. The entire issue of this journal deals with the studies of a group of Michigan State University researchers who examined what a group of elementary school teachers who taught math had done with the new California math curricular framework.

6 Note that in the instance of the Platoon School, the foundation-funded evaluation of the Gary Plan used experts to collect data on all aspects of the innovation but it was the test data that garnered media attention in 1918 and ultimately diminished its popularity. See Wolcott (1977), for the overt and covert ways that teachers acted in trying to reshape an administrative design to plan all curriculum, instruction, and budgeting for a district. The instance cited above of falling SAT scores being used by federal and state officials as evidence of U.S. schools’ decline revealed how scientific evidence is used by policymakers to seek ends that test experts found distasteful. Such expert opinion was dismissed as U.S. secretary after secretary of education in the 1980s used SAT scores on their “wall charts” that were reproduced in newspapers and magazines as evidence of American success or failure as scores rose and fell.

7 What makes this puzzling initially is that academic researchers have continually pointed out the methodological and design problems in the Effective Schools research and argued consistently that there is little scientific basis for constructing such programs. Although occasional scholarly studies are exempt from these criticisms (see the decade-long study on school effects in Louisiana completed by Teddlie & Stringfield, 1993), nonetheless, few of the practitioner-based reformers have listened to the academics because they have largely endorsed the central tenets of the reforms regardless of design and methodological flaws. In this case, practitioner endorsement gave the movement a legitimacy that academic research could not. I explain this as a rare alliance (albeit a silent one) between practitioners and policymakers and another exercise in policymakers’ influence in having their standards of judging success adopted.

8 Another connection between a top policy adviser in the Clinton administration and the Effective Schools movement is Michael Cohen, who served Secretary Riley before moving to the White House. In the late 1970s, Cohen worked in the National Institute of Education monitoring federal research grants to Ron Edmonds. In the early 1980s, he wrote articles based on findings from the research on effective schools. As a policy analyst working in the Department of Education in the 1980s and early 1990s, Cohen worked closely with Marshall Smith, who had left the government to be a professor at the University of Wisconsin and, later, Stanford before returning to Washington in 1993.

9 For an example of deep concern over using test scores to judge the success of Title I programs (see Rotberg and Harvey, 1993). They point out all of the shortcomings to using test scores, as reported by teachers, principals, researchers, and policymakers, and go on to propose alternative measures that broaden the indicators used to determine whether the program is successful. These measures remain within the effectiveness standard for obvious reasons since the original legislation required evaluations that relied on measuring student academic performance.

10 By the early 1990s, the Coalition of Essential Schools (CES) had produced an alternative definition of a “good” high school. With its nine “essential principles,” ones that are largely antagonistic to implementing a five-factor formula dependent on scores on standardized achievement tests, there is focus on developing the student’s mind, teaching as coaching—no teacher should have more than eighty students—and democratic decision making. Graduation depends on how well each student displays mastery of certain skills and knowledge. Hence, there is no need for age-grading of students or awarding units for completion of courses since what the student performs in actual work is important. Although CES members and networking schools have swiftly spread in the last five years, less than 1 percent of U.S. high schools belong to CES (Sizer, 1992).

References

- Berman, P., & McLaughlin, M. (1978). *Federal programs supporting educational change: Vol. IV. The findings in review*. Santa Monica, CA: RAND Corporation.
- Bestor, A. (1953). *Educational wastelands*. Urbana: University of Illinois Press.
- Boyer, E. (1984). *High school*. New York: Harper & Row.
- Broder, D. (1987). *Behind the front page*. New York: Simon & Schuster.
- Brookover, W., & Lezotte, L. (1979). *Changes in school characteristics coincident with changes in student achievement*. East Lansing: Michigan State University, College of Urban Development.
- Brophy, J., & Good, T. (1986). School effects. In M. Wittrock (Ed.), *Third handbook of research on teaching* (pp. 570-602). New York: Macmillan.
- Bullard P., & Taylor, B. (1993). *Making school reform happen*. Boston: Allyn & Bacon.
- Carnegie Council on Adolescent Development. (1989) *Turning points: Preparing American youth for the 21st century*. New York: Carnegie Council.
- Chenoweth, T. (1992). Emerging national models of schooling for at-risk students. *International Journal of Educational Reform*, 1(3), 255-269.
- Chubb, J., & Moe, T. (1990) *Politics, markets and America’s schools*. Washington, DC: Brookings Institution.
- Cohen D., & Ball, D. (1990) Relations between policy and practice. *Educational Evaluation and Policy Analysis*, 12(3), 249-256.
- Cohen, D., & Spillane, J. (1992). Policy and practice: The relations between governance and instruction. *Review of Research in Education*, 18, 3-50.
- Cohen, M. (1982). Effective schools: accumulating research findings. *American Education*, 18(1), 14-15.
- Cohen. R. (1990). *Children of the mill: Schooling and society in Gary, Indiana, 1906-1960*. Bloomington: Indiana University Press.
- Cohen, R., & Mohl, R. (1979). *The paradox of progressive education: The Gary Plan and urban schooling*. Port Washington, NY: Kennikat.
- Coleman, J., et al. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Office of Education, National Center for Educational Statistics.
- Coleman, J., Kilgore, S., & Hoffer, T. (1982). *High school achievement: Public, Catholic, and private schools compared*. New York: Basic Books.
- Comer, J. (1988). Educating poor minority children. *Scientific American* 259(5), 28-34.
- Corcoran, T., & Wilson, B. (1985). *The search for successful secondary schools*. Washington, DC: Government Printing Office.
- Cremin, L. A. (1961). *The transformation of the school*. New York: Vintage Press.
- Cuban, L. (1984). Transforming the frog into a prince: Effective schools research, policy, and practice at the district level. *Harvard Educational Review*, 54(2), 129-151.
- Cuban, L. (1993). *How teachers taught* (2nd ed.). New York: Teachers College Press.
- Darling-Hammond, L. (1993). *Federal policy options for chapter 1: An equity agenda for school restructuring*. New York: National Center for Restructuring Education, Schools, and Teaching, Teachers College, Columbia University.
- Dewey, J., & Dewey, E. (1915) *Schools of tomorrow*. New York: Dutton.
- Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37(10), 18-24.
- Education Week*. (1993, February 10). Charting a course for reform: A chronology, pp. S1-S20.
- Elmore, R., & McLaughlin, M. (1988). *Steady work*. Santa Monica: RAND Corporation.
- Emrick, A., Peterson, S., & Agarwala-Rogers, R. (1977). *Evaluation of the national diffusion network, vol. 1: Findings and recommendations* (SRI Project, 4385). Washington, DC: US Office of Education.
- Emshoff, J., et al. (1987). Innovation in education and criminal justice: Measuring fidelity of implementation and program effectiveness, *Educational Evaluation and Policy Analysis*, 9(3), 300-311.
- Fallows, J. (1996). *Breaking the news*. New York: Random House.
- Firestone, W. (1991). *Rethinking effective schools*. Englewood Cliffs, NJ: Prentice-Hall
- Firestone, W., & Herriott, R. (1982). Prescriptions for effective elementary schools don’t fit secondary schools. *Educational Leadership*, 40, 51-53.
- Fullan, M. (1991). *The new meaning of educational change*. New York: Teachers College Press.
- General Accounting Office. (1989). *Effective schools programs: Their extent and characteristics* (GAO Publication No. HRD-89-132BR). Washington, DC: Government Printing Office.
- Goodman, J. (1992, April 20). Airline watchdog is fly-by-night (and day) scholar. *Chronicle of Higher Education*, p. A5.
- Herndon, J. (1968). *The way it spozed to be*. New York: Simon & Schuster.
- Hoff, D. (1997, April 16). Chapter 1 study documents impact of poverty, *Education Week*, pp. 22-23.
- Holt, J. (1964) *How children fail*. New York: Pitman.
- Jencks, C., et al. (1972) *Inequality: A reassessment of the effects of family and schooling in America*. New York: Basic Books.
- Kahne, J. (1996). *Reframing educational policy*. New York: Teachers College Press.
- Kennedy, M., Jung, R., & Orland, M. (1986). *Poverty achievement and the distribution of compensatory education services* (First interim report of the National Assessment of Chapter I). Washington, DC: U.S. Department of Education.
- Kingdon, J. (1984). *Agendas, alternatives and public policies*. Boston: Little, Brown.
- Kirst, M., & Meister, G. (1985). Turbulence in American secondary schools: What reforms last? *Curriculum Inquiry*, 15(2), 169-186.
- Kliebard, H. (1990). Vocational education as symbolic action: Connecting schooling with the workplace. *American Educational Research Journal*, 27(1), 9-26.
- Knapp, M. (1997). Between systemic reforms and the mathematics and science classroom: The dynamics of innovation, implementation, and professional learning. *Review of Research*, 67(2), 227-266.
- Kohl, H. (1968). *36 Children*. New York: New American Library.
- Kurtz, H. (1993) *Media circus*. New York: Random House.
- Levin, H. (1989). *Accelerated schools: A new strategy for at-risk students* (Policy Bulletin No. 6). Bloomington: Consortium on Educational Policy Studies, School of Education, Indiana University.
- Levine, D., & Lezotte, L. (1990). *Unusually effective schools: A review and analysis of research and practice*. Madison, WI: National Center for Effective Schools Research and Development.
- Levine, D., & Ornstein, A. (1993). School effectiveness and national reform. *Journal of*

- Teacher Education*, 44(5), 335-345.
- Lewin, T. (1987, April 2). A sharp debate on hospitals. *New York Times*, p. 29.
- Lewis, A. (1990). *Making it in the middle: The why and how of excellent schools for young urban adolescents*. New York: The Edna McConnell Clark Foundation.
- Lightfoot, S. L. (1983). *The good high school*. New York: Basic Books.
- McLaughlin, M. (1975). *Evaluation and reform: The Elementary and Secondary Education Act of 1965, Title I*. Cambridge: Ballinger.
- McLaughlin, M. (1978). Implementation as mutual adaptation in classroom organizations. In D. Mann (Ed.), *Making change happen* (pp. 19-31). New York: Teachers College Press.
- Majone, G. (1989). *Evidence, argument, and persuasion in the policy process*. New Haven: Yale University Press.
- Meyer, J. (1992). Innovation and knowledge use in American public education. In J. Meyer and R. Scott (Eds.), *Organizational environments: Ritual and rationality* (pp. 233-260). Newbury Park, CA: Sage.
- Meyer, J., & Rowan, B. (1978). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83(2), 340-363.
- Muncey, D., & McQuillan, P. (1996). *Reform and resistance in schools and classrooms*. New Haven: Yale University Press.
- Murnane, R. (1987). Improving education indicators and economic indicators: The same problems? *Educational Evaluation and Policy Analysis*, 9(2), 101-116.
- Murphy, J. (1992, January 3). *School effectiveness and school restructuring: Contributions to school improvement*. Plenary Address to International Congress for School Effectiveness and School Improvement, London.
- Murphy, J., Hallinger, P., & Mesa R. (1985). School effectiveness: Checking progress and assumptions and developing a role for state and federal government. *Teachers College Record*, 86(4), 615-641.
- National Commission on Excellence in Education. (1983). *A nation at risk*. Washington, DC: Government Printing Office.
- National Governors' Association. (1991). *From rhetoric to action: State progress in restructuring the education system*. Washington, DC: National Governors' Association.
- Ogawa, R. (1994). The institutional sources of educational reform: The case of school-based management. *American Educational Research Journal*, 31(3), 519-548.
- Pfeffer, J., & Salancik, G. (1978). *External control of organizations*. New York: Harper & Row.
- Purkey, S., & Smith, M. (1983). Effective schools: A review. *Elementary School Journal*, 83(4), 427-452.
- Ralph, J. H., & Fennessey, J. (1983). Science or reform: Some questions about the effective schools model. *Phi Delta Kappan*, 64(10), 689-694.
- Ravitch, D. (1983). *The troubled crusade*. New York: Basic Books.
- Reich, R. (1992). *The work of nations*. New York: Alfred A. Knopf.
- Rotberg, I., & Harvey, J. (1993). *Federal policy options for improving the education of low-income students*. Santa Monica: RAND Corporation.
- Rowan, B., Bossert, S., & Dwyer, D. (1983). Research on effective schools. *Educational Researcher*, 12(4), 24-31.
- Safire, W. (1997, May 18). Elite establishment egghead eupatrids. *New York Times Magazine*, p. 16.
- San Jose Mercury*. (1986, September 22). EPA mileage ratings for 1987 model vehicles, p. 8A.
- Scheffler, I. (1984). On the education of policymakers. *Harvard Educational Review*, 54(2), 152-165.
- Schön, D. (1983). *The reflective practitioner*. New York: Basic Books.
- Shipp, D. (1997). The invisible hand: Big business and Chicago school reform. *Teachers College Record*, 99(1), 73-166.
- Scott, R. (1987). *Organizations: Rational, natural, and open systems*. Englewood Cliffs, NJ: Prentice-Hall.
- Silver, H. (1994). *Good schools, effective schools*. New York: Cassell.
- Sizer, T. (1984). *Horace's compromise*. Boston: Houghton Mifflin.
- Sizer, T. (1992). *Horace's school*. Boston: Houghton Mifflin.
- Slavin, R., Madden, N., Karweit, N., Livermore, B., & Dolan, L. (1990). Success for all: First year outcomes of a comprehensive plan for reforming urban education. *American Educational Research Journal*, 20(2), 255-278.
- Smith, M., & O'Day, J. (1990). *Systemic school reform. Politics of education association yearbook*. London: Taylor & Francis.
- Snyder, J., Bolin, F., & Zumwalt, K. (1990). Curriculum implementation. In P. Jackson (Ed.), *Handbook of research on curriculum* (pp. 402-435). New York: Macmillan.
- Stedman, L. (1987). It's time we changed the effective schools formula. *Phi Delta Kappan*, 69(3), 215-224.
- Teddlie, C., & Stringfield, S. (1993). *Schools make a difference: Lessons learned from a 10-year study of school effects*. New York: Teachers College Press.
- Tolchin, M. (1990, January 12). Hospital death rate report challenged by researchers. *New York Times*, p. A11.
- Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia*. Cambridge: Harvard University Press.
- Tyack, D., & Hansot, E. (1982). *Managers of virtue*. New York: Basic Books.
- U. S. Congress, Office of Technology Assessment. (1992). *Testing in American schools: Asking the right questions*. Washington, DC: Government Printing Office.
- U.S. Department of Education. (1987). *What works: Schools that work*. Washington, DC: Author.
- Weber, G. (1971). *Inner-city children can be taught to read: Four successful schools*. Washington, DC: Council for Basic Education.
- Welker, R. (1992). *The teacher as expert*. Albany: State University of New York Press.
- Wigginton, E. (1986). *Sometimes a shining moment*. Garden City, NY: Doubleday.
- Wirt, F., & Kirst, M. (1989). *Schools in conflict*. Berkeley: McCutchan.
- Wise, A., & Darling-Hammond, L. (1985). Beyond standardization: state standards and school improvement. *The Elementary School Journal*, 85(3), 315-336.
- Wolcott, H. (1977). *Teachers and technocrats*. Eugene: Center for Educational Policy and Management, University of Oregon.

How schools change reforms: Redefining reform success and failure. Article. Mar 1998. Teach coll rec. L. Cuban. Because schools change reforms as much as reforms change schools, judging an innovation's success or failure has been, and is, no easy task. What emerged as crucial in evaluating school reforms is what criteria are being used to make judgments, whose criteria they are, and how schools change reforms as they are implemented. The example of Effective Schools challenges policymakers and researchers to become more explicit about which criteria they employ to judge success and to understand how the journey of school reform is a story of constant adaptation that ultimately undermines the common criteria generally used to judge success and failure. Beginning with the assumption that schools change reforms as much as reforms change schools, this paper argues that the judgment of a reform's success depends heavily on the criteria used to make the judgment. Criteria used by policymakers (effectiveness, popularity, and fidelity) and practitioners (adaptability and longevity) are identified and applied to effective schools reform. (SM). How schools change reforms: Redefining reform success and failure. Teachers College Record, 99(3), 453-477. Google Scholar. Darling-Hammond, L. (1997). The human side of school change: Reform, resistance, and the real-life problem of innovation. San Francisco: Jossey-Bass. Google Scholar. Ferguson, K. E. (1984).