Educational Evaluation and Policy Analysis
Fall, 1991, Vol. 13, No. 3, pp. 309–321

The World's New Political Economy Is Politicizing Educational Evaluation

James W. Guthrie
University of California, Berkeley

Educational evaluation conventionally has concentrated on measuring individual student achievement, appraising instructional methods and materials, and assessing program performance. Major issues in the field have been scholarly and methodological. The central career orientation of educational evaluators has been toward academic colleagues and practicing educators. However, contemporary education reform efforts aimed at using schooling to enhance national economic development are altering this conventional orientation. Managerial expectations are replacing professional relations as the prime orientation of the enterprise, and the broader environment in which evaluation takes place is becoming intensely politicized. This essay (a) explains the evolving human capital imperative and its relation to education, (b) links these schooling changes to economic development, (c) summarizes the historic orientation of the education evaluation field, (d) suggests the evaluation dynamics which develop when governments reshape schooling systems in order to enhance national economic growth, and (e) outlines an alternative model for educational appraisal.

Educational development increasingly is a function of economic change, and, conversely, educational change increasingly is intended to foster national economic development.

An economic or investment model of schooling has been ascending in significance since the 1965 enactment of the Elementary and Secondary Education Act. Educated intelligence, human capital, is rapidly emerging as a strategic economic resource, and a "human capital imperative" may be shaping a global model of schooling. Increasingly, Western Bloc nations expect lower and higher education systems to contribute forcefully to economic growth and are enacting reform policies to accomplish this objective. As the purposes of education become economically more determined, the practices of pedagogy seem to become ever more predictable. A common set of schooling components appears to be emerging internationally.

This newly developing, economically motivated education systems model for nations alters the assessment environment and thus provokes tensions between conventional evaluation norms and procedures and the evolving expectations of the political system. This essay suggests a means to restore equilibrium through creation of a new institutional perspective for evaluation.

Education's Evolving External Environment

The dominant justification for schooling is shifting. Many industrialized nations are attempting to enhance their economic position through the development of human capital, and, therefore, policymakers increasingly escalate their expectations for the performance of education systems. Although it is not always as obvious, renewed emphases upon the economic purposes of schooling are also substantially reshaping expectations for and the operating environment of educational evaluation.

Powerful military forces, possession of strategically significant geographic locations, access to valuable raw materials, and widespread capacity in basic industries were once the mainstays of national power and international hegemony. Such is less the case today. Conventional military might may still be important for political purposes, but it is declining in economic significance, as is geography. The "new" strategic raw material, upon which economic productivity is ever more dependent, is human capital.[1]

Throughout history, technological innovations have redistributed power, enabled a tribe, a people, or a nation to vie for and gain dominance. Fire, ferrous metal, and farming are historic discoveries which transformed nations and transferred power. Modern examples include internal combustion engines, interchangeable parts, electrical energy, and electronic components. The list is longer, but the point is the same: Significant technological revolutions formerly were founded upon sporadic discoveries. Increasingly, such shifts are crucially dependent upon systematic scientific inventions.

Rapid communication, expanding information, and modern organizational arrangements are transforming Western economies. They are now global in their competitive outlook; internationally interdependent; insatiable in their quest for technological innovation; possessed of a fluid ability to move capital, techniques, and personnel across national borders; and crucially dependent upon the availability of human talent. Reliance upon a narrow number of intellectual elite is increasingly outmoded. Modern manufacturing and service industry techniques demand an entire labor force capable of continually adjusting to new technologies and making informed decisions. Educated and highly skilled human intelligence is coming to be viewed as a nation's primary economic resource, and it is needed in large amounts.

Modern economics, however, is not simply boosting or gently influencing an already initiated notion that education systems should enhance a nation's human capital resources. Rather, international economic forces are rapidly and intensely reshaping the forms of schooling across national boundaries. This globalization of education is occurring because nations no longer can easily protect their domestic economies from international economic forces. Failure to respond quickly to technological and organizational inventions can rapidly jeopardize a nation's standard of living and a government's political future. Increasingly, even Eastern Bloc nations find that they can no longer isolate themselves from the rapid ebb and flow of developments in international trade, technology, and finance.

The following quotation from an October 1989 *Atlantic* article (Morris, 1989) on economic development crystallizes the complex, intertwined, and rapidly evolving nature of international manufacturing and services industries.

> Ford, with one third of its sales from outside the United States, owns 25 percent of Mazda. Mazda makes cars in America for Ford; Ford will reciprocate by making trucks for Mazda; and the two companies trade parts. Each owns a piece of Korea's Kia Motors, which produces the Ford Festiva for export to the United States. Ford and Nissan, Japan's No. 2, swap vehicles in Australia and are planning a joint minivan program in America. Ford and Volkswagen have merged into a single company in Latin America, which exports trucks to the United States.
>
> General Motors holds a 41.6 percent stake in Isuzu, which is starting a joint venture in America with Subaru, which is partly owned by Nissan. GM also owns half of Daewoo Motors, Hyundai's major competitor in Korea. Daewoo makes Nissan cars for Japan and Pontiacs for America; soon it will be selling cars that were primarily designed by GM-Europe to Isuzu in Japan. GM has also teamed with Japan's No. 1, Toyota, to produce cars under both companies' labels in America and Australia. (pp. 53–54)

As complicated as the above-described description is, it probably represents the economic future for the industrialized world. Consequently, traditional educational values and institutions are being crowded by political officials who, in response to these developing economic imperatives, believe that new educational policies are necessary for their nations to become or remain vital. Ex-

pansion of the populations served by schools and colleges, centralized curricula expectations, national educational objectives, expanded uses of standardized tests, growing dependence upon government agencies to collect and analyze school performance data, intensified efforts to link colleges and industry, and altered expectations for educational evaluation are among the predictable outcomes of this globalization movement.

It also should be added quickly that international economic conditions are by no means the only forces shaping modern educational systems. The picture is made more complicated by the existence of national education reform efforts which stem from purely domestic political purposes, are perhaps motivated by a nation's internal ideological dynamics, and are probably shaped by deep-rooted historical conditions, religious beliefs, and idiosyncratic practices.

Thus, given the multiplicity of national motives and the complexity of national conditions, one cannot help but be struck by the remarkable resemblance among internationally emerging educational policies. This convergence is particularly apparent in Western Bloc nations, and such is the focus of this essay. However, the probability appears great that, in time, global economic conditions will propel Eastern Bloc and nonaligned nations in similar directions.

Specific educational reform tactics may differ from nation to nation, depending upon historic development patterns, contemporary politics, current resource levels, and specific operating structures. In national systems emphasizing an elitist schooling model—for example, France, England, and historic members of the British Commonwealth—the clear long-run education reform goal is to expand the numbers of individuals eligible for and interested in seeking higher levels of schooling. In egalitarian-oriented systems, such as the United States, the long-run goal of education reform is to elevate achievement standards in such a way that there are larger numbers of well-educated workers.

Regardless of the variety of national tactics, the strategic objective is the same. The long-run goal is to utilize educated intellect as a strategic means for a nation to gain or retain an economically competitive position in the global marketplace. Pursuit of this objective is the principal stimulus for current widespread national efforts at education reform. Also, in tandem with the converging international model of schooling, this economic development strategy is altering policymakers' expectations for educational evaluation. The environment for evaluators and policy analysts is becoming intensely more complicated and politically involved as a consequence.

Tracing the Link to National Economic Development

Not all emerging education system similarities are aimed at enhancing national economic development. For example, higher and lower education policies regarding "choice" and "privatization" and the devolution of lower education management to school sites are generally intended more to enhance the productivity and efficiency of schooling systems themselves than to enable education to aid a nation's economy. However, many of the common reforms now being undertaken by industrial nations can be logically linked to national economic development. The purpose of this section is to illustrate that connection.

Downward Extension of Schooling

Preschool and child-care services are clearly a policy system response to growing two-earner household demands for assistance with child care. In addition, the desire to utilize the institutional time to assist youngsters in preparing for, or deriving greater benefit from, schooling is, arguably, a policy intended to enhance human capital. If children can arrive at school with a greater readiness for formal instruction, then schools themselves can be more productive. Teaching more subject matter earlier should enable more students to progress through higher levels of education. Thus, to the extent to which preschool and child-care services are more than simply custodial experiences for children, these policies should result in higher academic achievement, less wastage, and more school persistence.

Centralized Curricula

Governments increasingly are unwilling to permit decentralized curriculum decisions because they can less afford the risk of "important" school subjects being omitted or being granted insufficient coverage. This is particularly true for mathematics, science, technology, and foreign language, which are needed especially for economic development. It is also true for subjects such as history and government, which can convey a sense of national identity and common purpose.

Standardized Performance Appraisal

This reform is the enforcement tactic for the previously described centralized curriculum. Particularly where school site management is being implemented, a mechanism is needed for central governments to ensure compliance with national purposes. Standardized examinations, compiled, administered, and corrected by independent parties, are a strong lever for gaining such compliance. At the least, such a tactic can detect lack of compliance. In addition, standardized appraisal systems can be used as accountability strategies. Test scores can be used to expose low-performing units to unfavorable publicity and other negative sanctions, as well as for rewarding high-performing units.

Higher Education Reforms

The expansion of higher education, the aggregation of independent units into systems, the linking of higher education to the private sector, and shifting costs to consumers all have economic productivity linkages. Expansion and cost shifting are intended to provide places for larger student numbers, a move consistent with enhancing a nation's supply of human capital.

Creating systems is simultaneously a controlling, cost-cutting, and economy-building move. Control is intended to place the institutions more under the authority of central government so that they can better be used for economic purposes. For example, in this way it is easier to regulate admission standards, curriculum offerings, and degree requirements. In addition, through standardization and economies of scale, there is a hope of reducing operating costs.

Links to the private sector are intended to render higher education economically more practical, more in touch with the market for labor, and more in touch with the practical applications of basic research and technology transfer. Also, by having private sector representatives participating in higher education governance, operational efficiency is thought to be enhanced.

A Component Conspicuous by Its Absence

Given the almost universal desire to enhance the productivity of educational services, it is interesting to note that virtually no central government has undertaken a serious contemporary effort to improve the effectiveness of its education system, higher or K–12, through substantial investments in basic educational research or development of educational technology.

The growing dependence of national economies upon the productivity of education systems, and the increasing proportions of national resources allocated to education, intensify political pressures to measure schooling and appraise school performance. The obvious implication is that evaluation and analytic activities assume added overall significance. However, growing attention to schooling's economic purposes, and the particular converging components of an international model of schooling, are also reshaping the environment for educational evaluation, expanding expectations held for professional evaluators, and altering the circumstances under which evaluation efforts are viewed as useful and credible. Explaining conventional evaluation orientations and the evolving evaluation dynamics is the subject of the following section.

Educational Evaluation, Conventionally

There has long been a link between education systems, at least between mass education systems, and a nation's economy. Schools traditionally have responded to economic change, teaching courses and training students in new skills as work places evolved. Also, at the individual level, students have long been exhorted to go to school in order to enhance their personal economic well-being.

However, putting aside their growing contemporary connection with the economy, schools have traditionally been expected to fulfill a substantial range of additional functions, both for society and for the individuals and households involved. Acculturating new citizens; promoting religious, linguistic, and political indoctrination; inculcating government principles; ensuring social cohesion and civic order; preparing a citizenry for military participation; facilitating social mobility; and developing artistic and aesthetic tastes are among the other-than-economic functions variously expected of schools. Many of these other purposes are being subordinated to national economic development in the last quarter of the twentieth century. Nevertheless, it is important to be mindful that other functions existed, still exist (even if currently diluted), and undoubtedly will persist. Educational evaluation, thus, has had masters to serve other than or in addition to national economic development.

Roots of Evaluation

Education evaluation is a subfield comprising different disciplines and multiple beginnings.[2] The deepest roots probably stem from late nineteenth- and early twentieth-century efforts by psychologists to make individual testing scientific. Plumbing intellectual capacity, academic achievement, personality, aptitudes, and attitudes all gave rise to technical efforts to construct a science of testing (see Clifford, 1984). Psychologists were the pioneers, and their efforts became increasingly systematic and quantitative. A growing measurement field provoked the label *psychometry.*

Multitudes of pencil-and-paper and other kinds of small-apparatus tests were devised. The long-run implications of these scientific efforts were intended to apply to schools, work places, criminal justice systems, the military, and other segments of society. However, initial development efforts were strongly concentrated on recording, quantifying, categorizing, and predicting individual human traits and abilities. The primary reference groups of the prominent researchers engaged in these efforts were other psychologists and scientists. If there were any links to other segments of society, they were likely to be with those interested in the application of these new measurements techniques to areas such as industrial production, medicine, and mental health. Mental measurement research was certainly not tightly tied initially to government, policy-making, or resource allocation.

By the second quarter of the twentieth century, paper-and-pencil tests were ever more widely used to make judgments and predictions about individual performance. Standardized examinations were employed by selective institutions to determine admission eligibility and placement. Private-sector firms were developing tests which could be used by schools to assess students' learning. These were generally diagnostic. Testing and evaluation were given major boosts by the military's reliance upon individual testing in World War I and, particularly, World War II. However, regardless of the growing technical sophistication involved, the major development focus continued to be upon assessing individual knowledge or traits, not the performance of a classroom, a school, or an educational system.

Following World War II, methodological sophistication about education and school research increased substantially. The publication of a classic description of experimental and quasi-experimental design considerations by Donald Campbell and Julian Stanley (1963) illustrated a turning point. Educational evaluation increasingly included social scientists. It was no longer to be a field dominated exclusively by psychologists. This expansion was enhanced in the 1950s and 1960s by the increasing participation of the federal government in sponsoring specialized educational programs in America's public schools. In fact, former United States Senator Robert Kennedy (D, New York) exacted, as the price of passage, a proviso that 1965 Elementary and Secondary Education Act programs be systematically evaluated to determine their effectiveness (Guthrie, 1968). The infant field of educational program evaluation was thus given an enormous boost.

Another U.S. federal government study vastly influenced the course of education evaluation. In 1964, the U.S. Congress en-

acted a major civil rights act. Among other provisions, this bill mandated the nationwide assessment of educational opportunity. The strong suspicion among members of the executive and legislative branches desiring this study was that despite the racial desegregation decisions beginning in 1954, American school systems, particularly ones in the southern region of the United States, were still discriminating against Black students. A prominent sociologist, James S. Coleman (1966), was commissioned to lead a social science team which, in December of 1966, released the study *Equality of Educational Opportunity.*

This landmark report contained numerous influential facets. However, two features are of particular significance for this essay. First, the so-called "Coleman Report" relied upon a spectrum of behavioral and social science disciplines, psychologists, sociologists, statisticians, educators, and so forth. This study was a milestone of multidisciplinary cooperation and marked an important, highly visible blending of disciplines into an educational evaluation effort. Second, in that the study was commissioned by government, was focused on an entire national educational system, and had results which readily could be used for political purposes, it represents a milestone in the political evolution of educational evaluation.

Economics, Education, and the Evolution of Evaluation

Economic aspirations for education and the practical manifestation of this strategy in national schooling reform efforts appear to be influencing evaluation in two fundamental ways. First, evaluation increasingly is undertaken to comply with managerial directives and less often to fulfill scholarly purposes. Second, the context in which evaluation is conducted and performance results are reported is increasingly politicized.

These fundamental shifts have numerous practical expressions of which the following conditions are illustrative:

1. An increasing expectation that educational evaluation will be conducted for and results made understandable to generalists and laypersons, not simply educational professionals and technical specialists;
2. Expanded requests that educational performance reports be made available to a spectrum of popular audiences;
3. Evaluator involvement with greater ranges of, and politically more influential special interest groups and stakeholders directly interested in, evaluation outcomes;
4. An expansion of high-level general government agencies interested in, and perhaps responsible for, the conduct of educational evaluation and schooling appraisals, and greater linkages to government planning efforts and the policy-making process;
5. Skepticism regarding evaluations undertaken by educators and a consequent greater reliance upon evaluation generalists;
6. Higher likelihood of competing and adversarial evaluations;
7. Intensified risk of advocacy imitating analysis;
8. Policymakers requiring systemic evaluation strategies;
9. Greater reliance upon eclectic analytic tactics and expanded measurements
10. Added dependence upon institutional and international performance comparisons.

Managerial and Policy Relevance

The audience for evaluation has expanded and changed because of the wider national interests now involved. The evaluation community increasingly is pressed by public sector managers and policymakers for reports and assessments of education system productivity. What was once a narrow audience of a relatively few technical specialists has grown to encompass high-level government officials responsible for coordinating and making decisions across a spectrum of public sector services. Educational evaluation increasingly is expected to issue popularly understandable appraisals and public interpretations of complicated findings.

The new audience now includes managerial generalists, individuals who are neither educators nor evaluation experts. Individual program evaluations and technical appraisals intended for professional educators may continue to be in demand. However, the overwhelming growth is in requests for professional assessments that are understandable to lay generalists and the results of which can

be used for managing entire education systems.

Expanded Popular Audiences

As schooling outcomes increasingly shape an individual's and a nation's economic well-being, the populace expresses a commensurate interest in school performance. Students and their households desire specific information regarding an individual's performance, particularly his or her school performance relative to some comparison group. In addition, as education systems concentrate upon national economic development, they portend more spillovers, outcomes affecting an entire society. Hence, households increasingly also desire information regarding the education system collectively. When the entire economy is at stake, educational productivity becomes a matter of concern to "stockholders," not simply conventional audiences, that is, workers, managers, and members of the board of directors. Evaluators are increasingly expected to provide education-system performance information to this expanded popular audience (Savage, 1989).

The quest for "pop" evaluation information extends in an additional direction. Evaluation is a sufficiently technical undertaking that many professional educators, teachers, and administrators, who understand the significance of evaluation information but possess little preparation as evaluation specialists, also increasingly desire more easily understood interpretations and syntheses.

Evaluators conventionally have had little experience in translating technical findings into lay terms. The media can only partially fulfill the need. Hence, there is a growing demand for a cadre of translators possessing sufficient scholarly understanding to bridge technical specialists with policymakers and the media. Government agencies and private organizations desire to capitalize upon these skills as they increasingly are called upon to issue easily understood reports intended to enhance public and policymaker understanding of education system performance.

Stakeholder Complexity

Economically oriented education systems not only have expanded social consequences but also have attracted the personal interests of a wider range of special interest stakeholders. Assessment outcomes can influence the allocation of massive resources; they can reduce or reallocate finances, jobs, status and prestige, school attendance opportunities, college admissions, facilities location, and so forth. On each of these dimensions, there are one or more organized interest groups or interested individuals who rather quickly can coalesce into a group and mobilize for political action.

Modern education stakeholders generally are divided into four broad categories—education producers, patrons, policymakers, and the public. These classifications are neither neat nor mutually exclusive. It is possible on occasion for group interests to be aligned. For example, in some instances the interests of teachers and school managers coalesce. Similarly, school managers and policymakers periodically may have interests in common. However, on other occasions these and other groups may fragment badly. Educational interest group alignments are complicated and fluid, depending upon issues and surrounding events. The range, complicated and shifting alliances, and personal interests characterizing these stakeholders suggest that evaluators increasingly operate in and, to be effective, must be sensitive to intensely politicized interest group environments.

Expanding Evaluation and Links to Government Planning and Policymaking

The desire to assess education's contribution to national economic development motivates high-level government officials to monitor education more carefully. This, in turn, triggers expansion of general government evaluation efforts and a tendency to rely less upon professional educators and more upon evaluation generalists. Existing education bureaus probably will maintain, perhaps even intensify, their own evaluation efforts. Nevertheless, general government will create or expand its own independent efforts to evaluate school system performance.

Governments increasingly view evaluation as a feedback loop providing information useful for shaping the next round or cycle of

policy activities and resource allocations—hence, they desire that evaluation be regular, cyclical, and tactically consistent. Results need to be not only reliable but also timely in order to mesh with systematically recurring activities such as executive branch budget submissions and legislative sessions.

Also, in order to be maximally useful for planning purposes, evaluation information must be tactically consistent. For example, frequently altered achievement test formats impede longitudinal performance comparisons. Redefining "administrator" dilutes the ability to judge efficiency or administrative overhead data. Thus, the greater the reliance of system planners upon evaluation data, the greater the expectation that evaluators will undertake their efforts in a timely and consistent manner.

Skepticism About Professional Educators

When schooling assumes intensified national purposes and receives massive proportions of national resources, then professionals managing the service are increasingly viewed as instrumental to a policy objective. As such they are subject to added scrutiny, regulation, and suspicion. Educators may always have been viewed by policymakers as advocates for their own narrow professional endeavors and specialized agencies. However, this advocacy becomes more intensely scrutinized and possibly distrusted in the specialized realm of evaluation. Policymakers desire sources of analysis that provide at least the appearance of objectivity, independence, and dispassion. Hence, there is heavier reliance upon evaluation generalists in place of education specialists.

Intensified Advocacy and Adversarial Intensity

The complexity of the political environment, as well as the massive resources at risk, frequently renders stakeholder groups ambivalent regarding evaluation. Special interests satisfied with the status quo sometimes would rather avoid an appraisal. There is little to gain, and, if evaluations suggest less than ideal conditions or outcomes, much can be at risk if change ensues. For opposite reasons, those dissatisfied with existing conditions may well desire an evaluation. These conditions have prompted the political adage, "Where you stand on an issue may depend upon where you sit."

The above-described conditions can trigger political conflict among advocacy groups and special interests, even before the results of an evaluation are known. The conflict can span a range of issues all the way from whether the appraisal should be undertaken at all, to what assessment questions are appropriate to be asked of the program or operating unit, to who or what group should be authorized to conduct the evaluation, to efforts to control or refute the results.

Advocacy as Analysis

The magnitude of resources involved in education, and differences in stakeholder preconceptions, frequently lead interest groups to engage in advocacy evaluation. This occurs through two major avenues: (a) influencing the procedures or outcomes of what otherwise might be an objective, disinterested assessment or (b) controlling an evaluation by conducting it directly or arranging for the services of a sympathetically predisposed evaluator. Such advocacy evaluation efforts may well invoke, or at least attempt to invoke, the trappings of technical validity and scientific neutrality or objectivity in an effort to persuade decision makers of the study's legitimacy.

Such conflicts of interests in evaluation are camouflaged with varying degrees of success. If sufficient resources are perceived to be at stake, an adversarial interest group may commission or conduct a counter assessment or at least attempt to dilute damaging findings stemming from the initial study. Depending upon the credibility of the parties involved and the technical complexity of the evaluation design and procedures, conflicting findings and opinions can be confusing to the public and to policymakers. Under conditions of evaluation uncertainty, status quo advocates frequently prevail.

Appraising Entire Education Systems

Education for economic development encourages decision makers to expand or alter their views regarding the nature of useful

evaluation. Increasingly, evaluation is expected to appraise an entire national or statewide system, not simply the achievement of students or the performance of a specific program or particular operating unit. Policymakers seldom can gain leverage over smaller operating components such as classrooms, schools, local education authorities, or regions. Hence, they desire information based upon the governmental aggregate over which as a governing unit they exercise authority or influence. This can be a functional aggregate such as elementary, secondary, or tertiary schooling, or a government unit or subunit, a nation or a state. Regardless of the specific unit, the point is that policymakers desire information regarding systemic, not necessarily operating unit, performance.

Eclectic Appraisal Tactics and Expanded Measurement Portfolios

Public officials are called upon to balance three popularly held value preferences—equality, efficiency, and liberty (choice).[3] This constitutes a spectrum of criteria against which to assess institutional and program performance. Because of the complexity of this spectrum, evaluators increasingly are expected to employ multiple appraisal measures oriented toward educational contexts, inputs, processes, and outcomes.[4]

This comprehensive tactical approach is necessary for several reasons. First, even those who desire to transform the education system into an engine for economic development also desire efficiency of resource utilization. Progress on this value dimension can be appraised only if measures of both input and output exist. Additionally, government seldom has an opportunity to pursue a single policy objective exclusively. Consequently, education appraisal systems must also be comprehensive in order to satisfy not only the expectations of efficiency proponents but also the expectations of those who espouse egalitarian and libertarian objectives. Measurement of resource distribution, racial and gender access to schooling, and time to degree attainment are illustrative of the appraisal tactics taken to provide information beyond simply system outputs.

The necessity to diffuse advocate criticism also stimulates the use of expanded evaluation measures. The managerial approach to appraisal frequently emphasizes quantifiable outcome measures and performance indicators. Dissatisfied advocates and other critics of an evaluation report or finding frequently search for bases upon which to level their counteroffensive. One basis of criticism is inappropriate, ill-defined, or ill-conceived research tactics.

For example, it is sometimes alleged that quantifiable outcome measures contribute to organizational goal displacement, encouraging concentration upon the trivial because it is measurable and neglect of the meaningful because it is complicated. Another assertion is that many education outcomes are lagged; that is, they can be determined only after a sufficient period of gestation. Civic responsibility serves as an example. Students may complete secondary school prior to being eligible to vote. To meet objections such as these, evaluators are motivated to rely upon an extended quantitative portfolio of measurement tactics, surveys, metaanalyses, achievement tests, longitudinal studies, unobtrusive measures, systematized observations, and so forth.

Dependence Upon Comparisons

Education systems frequently are self-contained. It is difficult for them to establish internally valid and objective performance standards. At what age should children learn to read? How much geography, biology, or chemistry should an eighth grader know? Can 15-year-olds reasonably be expected to learn calculus? Are students learning less now than they used to? Are teachers' salaries sufficient? How many administrators should a school system employ?

These and similar questions for which there are few absolute answers illustrate the attractiveness of comparisons in educational evaluation. Comparisons can be of two types, vertical and horizontal. Vertical analyses involve comparing an institution with itself over time. How high were test scores in the polity 5, 10, and 20 years ago? What is the trend? In addition to longitudinal studies, increasing stock is placed in horizontal comparisons. How do reading, mathematics, or science scores in the polity compare with an

economic rival? What is the trend line? Indeed, this horizontal comparative tactic has proved sufficiently popular to spawn a vastly expanded evaluation industry in the form of studies of international educational performance.

The reshaping of national schooling systems to comply with a human capital imperative and the consequent alteration in expectations for educational evaluation frequently result in unproductive and unresolved policy sector tensions. The above-described alterations in the environment of educational evaluation result in many proposed changes. Existing evaluation agencies and actors attempt to adapt to evolving conditions, and new agencies are formed. However, it is unlikely that all the significant tensions resulting from this policy transformation can be resolved by incremental changes and additions by existing institutions.

Unresolved Evaluation Issues

Neither special interest groups, professional associations, nor governments typically have reconceptualized and reconstructed education evaluation systems sufficiently to smooth the practical problems and political tensions accompanying the shift to managerially oriented, politically motivated expectations for educational appraisal. It is difficult for conventional evaluation operations to alter their stance and adapt to the expanded set of managerial expectations. Similarly, newly formed government or advocacy group evaluation agencies are unlikely to avoid the pressures of politicization.

A new set of institutional arrangements may be necessary to resolve these tensions, restore a dynamic equilibrium, and promote or preserve public regard for professional evaluation. National education inspectorates provide a useful example of the difficulties involved in adapting traditional institutions to meet the new expectations and conflicting conditions.

National Inspectorates

In many nations, particularly in Western Europe, these professional education support agencies have a proud history. Through the recruitment of able professional educators and by promulgating high standards of pedagogical performance for schools, they have frequently earned the respect of their practicing educational colleagues. Generally, they are thought to be simultaneously rigorous in their judgements and helpful in their advice to practicing educators.

Inspectorates are expected to perform a range of functions which differ from nation to nation.[5] Frequently, however, their general emphasis has been upon the enhancement of individual school effectiveness and elevation of professional practice. Only in a few nations have they conventionally been oriented toward system-wide educational appraisal.

Under the pressures of altered policy sector expectations, inspectorates will predictably attempt to modify or expand their system-wide appraisal charters and practical assessment tactics. To meet the evolving requests for managerially useful evaluations described in the preceding section, they may expand their outlook to encompass the assessment of systems, enlarge their portfolios to include a greater number of tactical measures, attempt to issue less technical and more readable reports, contribute consistent and timely information for policy planning purposes, and cater to broader popular and professional audiences. In short, they may attempt to transform their operations to comply with new managerial assessment expectations, even if this means paying less attention to the conventional professional purposes of evaluation.

However, unless an inspectorate already has a substantial history of meeting national, policy system, evaluation needs, the probability is great that newly evolved, policy-sector motivated evaluation efforts will bear little sustained fruit. This will be true almost regardless of whatever practical changes inspectorates undertake. In part, professionally oriented national inspectorates will be trapped by their history and the lingering expectations of their education professional constituents. In addition to whatever practical problems they may have in making a transition to a different assessment paradigm, inspectorate members continually will perceive an obligation to practicing school professionals, and this loyalty will tarnish policymakers' perceptions of their objectivity.

Inspectorates risk becoming marginal institutions, abandoning a historic constituency to serve a new one which will not easily believe that the organization can switch its loyalty. The result may be that education inspectorates end up with no loves; having given up a partner of long standing to pursue a younger one, they eventually find themselves unwanted by either.

A New Perspective for Appraising Conditions of Education

New professional institutions for evaluators may be necessary to meet evolving management and policy sector expectations and simultaneously to cope with new political tensions. What appears to be needed are evaluation agencies and actors that are apolitical, independent of advocacy groups and special alignments, devoid of intense personal involvement and private conflicts of interests, institutionally credible, knowledgeable about both education and government, methodologically competent, absent inappropriate affiliations and debilitating past controversies, and capable of understanding, synthesizing, and brokering complicated appraisal information among technical specialists, professional educators, government generalists, the media, and the public.

This may seem to be a complicated and unrealistic set of expectations and qualifications. However, at least in the United States, and perhaps in other nations as well, examples of such new evaluation agencies are emerging. In at least two dozen states, policy analysis organizations have been formed that strive to comply with the above-listed set of credentials. These organizations are in various states of formation. Probably the longest standing, Policy Analysis for California Education (PACE), has existed only since 1983. Thus, no one of them yet occupies a dominant position in its state. Each still exists in a fragile environment in which it must continually secure financial resources, talent, and attention. Despite their primitive and precarious status, sufficient experience exists to offer several preliminary generalizations regarding the characteristics and conditions necessary for their success and survival.

New Formations

Old organizations frequently carry too much historic freight to meet effectively the complex mixture of new managerial and political expectations for evaluation. Remaking an existing agency in hopes of serving new audiences with new approaches still saddles the enterprise with old perceptions. It is unlikely that existing agencies will disappear, and doubtless many of them will do all they reasonably can to adapt. However, the probability is slender that an institutional skin graft or organizational facelift will suffice. New evaluation agencies, unfettered of past perceptions and obligations, appear needed.

"Old" People

Of course, the new agency need not really be staffed by old people, but its eventual success crucially depends upon a threshold level of experience among key actors. Brokering education evaluation to a variety of lay and popular audiences is not a set of quickly acquired skills. The necessary attributes can be learned, and new agencies should take as one of their responsibilities the preparation of future generations of policy and evaluation brokers. However, a new evaluation agency is unlikely ever to be successful unless it is initially staffed by a critical mass of experienced policy brokers.

Political Independence

For a new evaluation agency to be viewed as Republican or Democratic, Tory or Liberal, Christian or Socialist, and so forth, is a veritable kiss of organizational death. The organization must strive mightily to retain its political independence when many outside forces will attempt to coopt it to their point of view. Independence, however, is insufficient. The new agency must also fend off even the slightest perception that it is partisan or otherwise aligned with an interest or advocacy group.

Every conceivable punctilio of political impartiality must be observed. Efforts to provide technical assistance to one interested party or a policy briefing for one candidate must be followed with efforts to provide similar services to those holding opposite views or running as an opposing candidate. If polit-

ical donations are an exchange that provides needed access to office holders, then every effort must be exerted to ensure that all candidates for an office are assisted equally. There is a long list of practical steps, but they sum to the same significance: Impartiality in practice and perception is a *sine qua non* of evaluation effectiveness.

Credibility: Institutional, Organizational, and Personal

Credibility can be carefully nurtured over time. It is also possible to obtain a line of credibility credit almost instantly through appropriate institutional affiliation and organizational arrangements. In many nations this can be achieved by alignment with or sponsorship by one or a consortium of major universities. This is the pattern among new policy analysis agencies in the United States, where the overwhelming proportion of higher education institutions strive to maintain rigorous standards of scholarship independent of political influence. In nations where there is a long-standing perception that the universities and colleges are politically partisan, that is not an effective alignment. An alternative is necessary.

Institutional credibility can be enhanced, or its absence mitigated, by appointment of a prestigious advisory or governing board. The obvious strategy is to wrap the new evaluation agency in the mantle of legitimacy worn by a number of notable and neutral individuals. Careful attention to the selection of such an advisory body has added benefits. Not only is it possible to operate in the penumbra of their individual and collective luminescence, but also they are frequently individuals of a caliber which enables them to provide valuable advice in the process. Quite obviously, organizational credibility is enhanced if one or more of the principal staff members of a new evaluation agency possess a favorable reputation. The ideal situation is to align all three sources—institutional, organizational, and reputational.

Professional and Technical Competence

This encompassing characteristic is a another *sine qua non.* The excitement of a new agency, the value of experience, the patina of impartiality, and the mantle of credibility may carry a new evaluation agency for a short time. However, the quality of its "products" eventually determines its future. Competence must encompass an unusual range of knowledge and ability. An agency must possess individuals informed about educational practices, institutions, and issues, knowledgeable regarding government structures and procedures, sophisticated about political issues and actors, trained in research procedures, and experienced in translating technicalities to laypersons. These are all critical capabilities. Sometimes, though rarely, they can be found in a single person. More generally, it takes a team.

Resource Independence

To become successful—that is, to enjoy sustained high regard—new evaluation agencies must possess a measure of resource independence. Independent from what? The answer is insulation from the possibility of undue influence, independence from the agencies, programs, interest groups, and institutions with and about which a new evaluation organization must interact and constantly cast judgment.

Resource independence serves an additional function. An evaluation organization constantly dependent upon contracts for its financial survival seldom has sufficient resource slack to concentrate on issues for which there is no immediate client or to stand back from immediate issues and offer a long-range and comprehensive view of an education system. "Speaking truth to power" is a capacity that new evaluation organizations must possess to perform successfully. Inappropriate resource dependence erodes this possibility

Insulation can be achieved through several means, for example, philanthropic foundation support, gifts, endowments, subscriptions, university grants, and contracts. It is possible to accept contracts from government agencies, as well as from other organizations, over which a new evaluation agency must periodically cast judgment. It is crucial, however, not to cross a threshold of overdependence where any single funding source, with a potential conflict of interest, is critical for the organization's survival.

Conclusion

It is possible that existing evaluation agencies may alter their *modus operandi* sufficiently to achieve renewed credibility in a world of altered expectations and political pressures. It is also possible that a newly formed evaluation agency of the type described above might fail even if it met all the previously listed threshold criteria. However, policy officials, professional educators, and members of the interested public are more likely to support the new over the old, given the complexity of the evolving evaluation environment.

Notes

Portions of this article were presented at the Hong Kong Education Research Association's annual meeting, November 11–12, 1989.

[1]This is a term which began to gain widespread use as a consequence of the research of Nobel Laureate economist, Theodore Schultz (1971).

[2]For added detail regarding education evaluation, see Ernest R. House (1990).

[3]The significance of these three value preferences for public policy formation is explained in "Ready, AIM, Reform: Building a Model of Education Reform and High Politics," by James W. Guthrie and Julia E. Koppich, a draft chapter in a forthcoming volume edited by Hedley Beare and William Lowe Boyd and titled *Restructuring Schools: An International Perspective on the Movement to Transform the Control and Performance of Schools.* (Falmer Press).

[4]These variable dimensions are explained in greater detail by Jaap Scheerens in a paper commissioned by the OECD Secretariat titled "International Education Indicators: Process-Indicators of School Functioning—A Selection Based on the Research Literature on School Effectiveness," OECD, Center for Educational Research and Innovation, Paris, July 29, 1989.

[5]An April 1989 paper by Guy Neave resulting from the OECD-sponsored international conference on "The Role of Central Inspectorate in Defining, Assessing and Reporting on Quality in Education" describes the spectrum of activities in which inspectorates are expected to engage. It also describes the evolution through which a number of national inspectorates are moving in keeping with national schooling reform efforts.

References

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Boston: Houghton Mifflin.

Clifford, G. J. (1984). *The sane positivist.* Middletown, CT: Wesleyan University Press.

Coleman, J. S. et al. (1966). *Equality of educational opportunity.* Washington, DC: U.S. Government Printing Office.

Guthrie, J. W. (1968). A political case history: Passage of the ESEA. *Phi Delta Kappan, 49*(6), 302–306.

House, E. R. (1990). Trends in evaluation. *Educational Researcher, 19*(3), 24–28.

Morris, C. R. (1989). The coming global boom. *Atlantic, 264*(4), 53–54.

Savage, D. (1989, August). *The press and educational research: Why one ignores the other.* Paper presented at the Colloquium on the Interdependence of Educational Research, Educational Policy, and the Press at the Commonwealth Center for the Education of Teachers, University of Virginia, Charlottesville.

Schultz, T. W. (1971). *Investment in human capital.* New York: Free Press.

Author

JAMES W. GUTHRIE, Professor of Education and Co-Director of Policy Analysis for California Education (PACE), School of Education, University of California, Berkeley, California 94720. *Specialization:* educational administration.