CHAPTER 30

Serving the Public Interest Through Educational Evaluation

Salvaging Democracy by Rejecting Neoliberalism

Sandra Mathison

The public's interest is manifest in many contexts, but one that touches almost every member of society is schooling. Therefore, the institution of schooling is a key context for serving the public's interest, but schools are a complex and contested venue for both special and public interests. Simultaneously, schools are opportunities to create passive, docile workers *and* to create critical, independent-minded thinkers and doers.

The public interest (or common good, as it is sometimes called) is something most people agree is worthy, but the devil is in the details. For some, the public interest is served when there is the greatest good for the greatest number, a classic utilitarian perspective. For some, the public interest is served when all individuals are free to do as they please as long as they do no harm to others, a libertarian perspective. For some, the public interest is the elevation of every single member of a community or group.

In this chapter, I argue that globalism, and neoliberalism particularly, is a primary influence on conceptualizations of schooling and education; as a consequence, it influences what we consider to be quality schooling and education, including the means we employ to discern quality in education. This chapter analyzes the impact of globalism on education and thus on the evaluation of education. I suggest how evaluation of and in schooling might alternatively challenge or resist the values inherent in neoliberal conceptions of schooling and thus promote democratic values, including education for the benefit of all and a governmental role that arises from the interests of the people rather than capital.

Understanding the Global Context

When we speak of global contexts, we may invoke a common worldwide concern with say the environment—a common collective concern that we care for the natural resources on earth. This is what Nye (2002) referred to as globalism—that is, that the world is characterized by economic, military, environmental, and social connections across nations and continents. These connections are manifest in mutual relationships such as, for example, the dependence on low labor costs in Asia to provide affordable goods for U.S. and European markets. Often the term *globalism* connotes a mutually satisfying interconnected relationship—Asian workers are gainfully employed, and a desire for affordable consumer goods in other parts of the world is satisfied.

The political theory underlying economic globalism is neoliberalism. To bring stability to the unstable and conflict-riddled world after the Second World War, organizations with global reach (the United Nations, the International Monetary Fund, and the World Bank) were created. Many nations experienced affluence and economic growth during this period, but a global increase in unemployment and inflation during the 1970s disrupted this sense of progress resulting in increases in political power for socialist and communist parties in Europe and even the United States (Harvey, 2005). It is this populist threat to the economic elites in both capitalist and developing countries that is commonly understood to have ushered in neoliberalism, a theory of political economic practices that promotes individual entrepreneurial freedom, frees capital to move across time and space by eliminating regulations, and assigns the state the role of facilitating competitiveness and privatization (Harvey, 2005). The cover of David Harvey's book, *A Brief History of Neoliberalism*, shows portraits of Ronald Reagan, Deng Xiaoping, Augusto Pinochet, and Margaret Thatcher. This book cover signals the worldwide reach of neoliberalism, a set of economic practices that have simultaneously taken hold in a wide range of contexts, including developed and imperialist countries such as Britain and the United States, emerging democracies such as Chile, and communist countries such as China.

Neoliberalism is suspicious of democracy, in either a majoritarian or populist sense. In either sense of democracy, the collective good is viewed as potentially inconsistent with individual rights and liberties, and thus neoliberals favor governance by experts and elites, usually elites with capital. In practice, neoliberalism looks different in different regions of the world and is in large part not effective at achieving its outcomes—both the U.S. and Chinese governments have had to use deficit financing to support their militarism, consumerism, and infrastructure development (Harvey, 2005). Neoliberalists would claim that neoliberalism has been imperfectly applied, but others might suggest that the free market ideology has run up against the resistance of other ideologies. Still, the rhetoric of neoliberalism underlies many educational reforms around the world, including how we determine whether education is effective and efficacious.

Educational Reforms Driven by Neoliberalism

Although this chapter deals most especially with the manifestation of globalization on evaluation in elementary and secondary education in the United States, there are similar educational reforms around the world that are provoked by neoliberalism and have seen a focusing of evaluation on single measures, standardization, and breaking down of the sanctity of national boundaries for the sake of the global economy (Ross & Gibson, 2007). The following examples illustrate the common thread of neoliberalism in educational reform around the world and across educational levels.

The Reinvention of European (and Latin American) Higher Education

Higher education in Europe and Central and South America is being reinvented to be more competitive in the global market (in both attracting students and producing graduates) and to increase the mobility of labor across national boundaries. The Bologna Declaration of 1999 was signed by the European ministers of education to facilitate this intent and called for the development of "comparable criteria and methodologies" to promote quality assurance. The objectives of these higher education reforms are: "1) to facilitate the speedy entrance of educated professionals into the job market through shortened degrees; 2) to enhance the cross-border mobility of students and job seekers; and 3) to increase the competitiveness of European higher education internationally" (Sedgewick, 2003). Two key features of the Bologna Process are the creation of a common two-tiered degree structure (bachelor's and master's degrees) already typical in Britain, the United States, Australia, and Canada, as well as the introduction of tuition fees, the latter change contributing significantly to a conception of education as a private good (Altbach, 2008).

This sort of universal accreditation movement promotes individual accomplishment, mobility, and economic benefit, the student and graduate as commodities that can be attracted for the mutual benefit of individuals, regions, and nations.

International Comparisons of Achievement

All evaluation involves comparisons, and so the fascination with international comparisons of educational achievement makes some inherent sense. Studies that make such comparisons (e.g., Program of International Student Assessment [PISA], Trends in International Mathematics and Science Study [TIMSS], Progress in International Reading Literacy Study [PIRLS], and the International Assessment of Educational Progress [IAEP]) are also motivated by the perception that schooling plays a critical role in global economic competitiveness, and so we need these indicators to take the pulse of national economic competitiveness (Bracey, 2008).

The neoliberalist claim is that the availability of data about schooling practices and outcomes around the world will result in the adoption of approaches from around the globe that "work" (Puryear, 1999). Such cross-national comparisons are meant to motivate political action at the national level:

> Convinced that poor-quality schools are a major bottleneck to economic growth and social advancement, [heads of state] are charging ministers of education with reform agendas and providing them with political support. Often, they are aided by technocrats from sectors of government other than education, particularly ministries of finance and planning, or from nongovernmental think tanks, whose views of educational policy are based firmly on modern economic theory. (Puryear, 1999)

No Child Left Behind

No Child Left Behind (NCLB), the manifesto for educational reform in the United States, is always described as bipartisan, but it is a substantial collaboration among politicians, at the federal and state levels, and coalitions representing corporate interests (Mathison, Ross, & Vinson, 2006). The rhetoric of NCLB suggests that schools will finally be held accountable for the success of each and every child, but the legislation also supports privatization (through tutoring provided by for-profit businesses and by encouraging parents to take their children elsewhere if there school is "failing") and the

standardization of teaching and curriculum. This 2001 reauthorization of the previous *Elementary and Secondary Education Act* focuses on a number of required outcome measures, including testing of all third through eighth graders in reading and mathematics and the disaggregation of test scores by subgroups of students (e.g., by ethnicity, race, gender, and special needs). For the first time, federal funding to local educational authorities is tied to participation in the mandates of NCLB and clear demonstrations of academic progress.

Neoliberalism and the State's New Evaluative Role

It is important to understand that neoliberalism defines a role for government—it is not an ideology that rejects governmental intervention. Whereas classical liberalism simply rejects the state, neoliberalism accepts and fosters a role for the state.[1] Corporate CEOs, politicians, and bureaucrats work together to promote and sustain the ideology's core values.

> [N]eo-liberalism has come to represent a positive conception of the state's role in creating the appropriate market by providing the conditions, laws and institutions necessary for its operation.
>
> In the shift from classical liberalism to neo-liberalism, then, there is a further element added, for such a shift involves a change in subject position from "homo economicus," who naturally behaves out of self-interest and is relatively detached from the state, to "manipulatable man," who is created by the state and who is continually encouraged to be "perpetually responsive." It is not that the conception of the self-interested subject is replaced or done away with by the new ideals of "neo-liberalism," but that in an age of universal welfare, the perceived possibilities of slothful indolence create necessities for new forms of vigilance, surveillance, "performance appraisal" and of forms of control generally. In this model the state has taken it upon itself to keep us all up to the mark. The state will see to it that each one makes a "continual enterprise of ourselves" . . . in what seems to be a process of "governing without governing." (Olssen, 1996, p. 340)

The New Evaluative Role of the State

In part, the state is assigned the responsibility of constructing and sustaining the rhetoric that fosters the neoagenda. DeJarnatt's (2003) analysis of the rhetoric of school reform in Philadelphia illustrates how a set of values that suggest limited government intervention can best be manifest with specific and strong government intervention:

> The reform forces use the rhetoric of choice and parental empowerment dear to the authoritarian populists and the privateers but the reforms themselves have been imposed with minimal choice or input by parents, students, or teachers and the market has been imposed by the state not chosen by any parent or student. Instead the changes have been dictated by the neoconservative[2] state bureaucracy, guided by an unquestioned belief in the value of uniformity and high-stakes standardized testing. (p. 33)

Gutmann (1999) defines three models of the state's role in education: a "family state," model where the state controls education; a "state of families" model, where parents are vested with responsibility for education; and a "state of individuals" model, where individual choices are made possible without prejudice for any perspective. Although the neoliberal rhetoric of educational reform suggests a "state of families" perspective, for example, with the claim that parents are a child's first teacher and will be empowered by accountability, in reality the

implementation of this agenda employs a "family state" theory through the construction of an image of failing schools that can only be saved by state-controlled accountability primarily based on student assessment.

Education may be a local prerogative, but often the regional or federal government is uniquely positioned to demand compliance through particular approaches to evaluation by threatening to withhold funding or support from those who do not comply and succeed. In the United States, the withholding of Title I money (i.e., money meant to support those who are academically disadvantaged because of poverty) is the federal government's leverage. The neoliberal agenda depends on this state intervention to support the rhetoric of choice and quality. Without the governmental power to demand common indicators for making choices, the neoliberal preference for private, charter, and other school choice options simply does not work.

Additionally, the state, through its power of surveillance and specialized knowledge, has taken on a role of providing evaluative information to the public about what is good and right. This is sometimes confined to programs the government funds, such as the U.S. Office of Management and Budget's system for evaluating and publicizing whether government-funded programs work. *Expect More* (http://www.expectmore.gov) offers to tell the public which programs are performing effectively and ineffectively, as well as those about which the jury is still out. Other government resources reach beyond government-funded programs to let the citizenry know what works. The best example of this is the *What Works Clearinghouse* (http://ies.ed.gov/ncee/wwc/), created in 2002 by the U.S. Department of Education's Institute for Educational Science "to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education." In both cases, the government assumes the role of telling the public what the best choices are.

There are, of course, private nongovernmental agencies that offer similar services to the public, such as SchoolMatters, a product of Standard and Poors, which is owned by McGraw-Hill Companies—one of the biggest producers of educational tests. "SchoolMatters gives policymakers, educators, and parents the tools they need to make better-informed decisions that improve student performance. SchoolMatters will educate, empower, and engage education stakeholders." In Canada, the Fraser Institute publishes school rankings for half of the country's provinces based on provincially mandated student achievement tests based on their contention that, "An educational market, one in which parents choose their children's schools and schools compete more freely for students, will produce better educational results for more students." CanWest Global Communications, the company that owns many Canadian daily newspapers, implicitly supports this contention by publishing newspaper inserts with these "reports" of the quality of schools prepared by the Fraser Institute and relegating alternative views to the op-ed pages.

The New Role of Evaluator as Technician

This government as evaluator role coheres with neoliberalism's governance by experts and Chomsky's (1997) "spectator democracy," a society in which a specialized class of experts identifies what is good and manufactures consent for the populace. Educational evaluators become technicians in this environment, carrying out the tasks associated with managing, administering, and reporting student assessment data. For some evaluators this is a positive sign of focus and unity, whereas for others it is stepping away from true evaluation—that is, using multiple criteria, indicators of performance, with an eye to all consequences of education (Mathison & Muñoz, 2008).

This technician role carries over to what are accepted as appropriate evaluation methodologies. Although there is much debate, the U.S. government has indicated a preference for funding randomized clinical trials or regression

discontinuity designs. This expectation is outlined in the *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide*, published by the U.S. Department of Education.

> Well-designed and implemented randomized controlled trials are considered the "gold standard" for evaluating an intervention's effectiveness, in fields such as medicine, welfare and employment policy, and psychology. This section discusses what a randomized controlled trial is, and outlines evidence indicating that such trials should play a similar role in education.

Few, if any, educational evaluations have been of the sort suggested by the U.S. government, and indeed much of the theoretical and practical work in educational evaluation since the 1960s has been directed to creating different evaluation methods and models of evaluative inquiry (not just borrowed research methods) that answer evaluative questions—questions about feasibility, practicability, needs, costs, intended and unintended outcomes, ethics, and justifiability (Mathison, 2008). The methodological mandates described earlier put aside the unique contributions evaluation can make and demand compliance with particular methodologies driven by particular epistemologies.

The professional evaluation community is not of a single mind about the claim that randomized clinical trials are the gold standard for educational evaluation (Donaldson & Christie, 2005). Although this is simplistically portrayed as revival of the quantitative-qualitative debates in evaluation, the situation is more complex. The methodology of choice is a reflection of underlying values (House & Mathison, 1984). Just as progressivism was the value context up to the late 1970s and even early 1980s, neoliberalism has been the value context that brings educational evaluation to where we are today in the United States and, indeed, in most parts of the world. Schools are a business, education is a product, products should be created efficiently, and one should look to the bottom line in making decisions. Implicit in this neoliberal perspective are values (and rhetoric) that motivate action. The most obvious of these values is that accountability is good, that simple parsimonious means for holding schools accountable are also good, that choice or competition will increase quality, and that it is morally superior to seek employability over other purposes of education. Econometrics drives thinking about what these simple parsimonious means are—thus the appeal of single indicators such as standardized tests.

Neoliberalism relies on specialized knowledge and silencing or at least muting the voices of the populace. Unlike many approaches to evaluation that are built on the inclusion of stakeholders in directing and conducting the evaluation, experimental design is controlled by experts and stakeholders (especially service providers and recipients) are conceived of more as anonymous subjects and less as moral, sociopolitical actors.

Educational Evaluation Simplified to Student Assessment

The educational reform package promoted within a globalized world is large and complex, but often evaluation of education is simplified to a single outcome and a single measure, often the assessment of student learning in a few academic areas. This key element is controlled by the government and thus is a means for controlling the institution of schooling. Tests are therefore a government-mandated intervention that creates, even if primarily rhetorically, market conditions and focuses the content of schooling on certain knowledge and values.

But the question is: How does the assessment of students cohere with neoliberal ideology? The idea of assessment-based accountability is deceptively simple:

> Students take tests that measure their academic performance in various subject areas. The results trigger certain consequences for students and schools—rewards, in the case of high performance, and sanctions for poor performance. . . . If students, teachers, or schools are chronically low performing, presumably something more must be done: students must be denied diplomas or held back a grade; teachers or principals must be sanctioned or dismissed; and failing schools must be fixed or simply closed. (Elmore, 2002)

The assumption is that the threat of failure will motivate students to learn more, teachers to teach better, educational institutions to be better, and the level of achievement to continue to rise.

Neoliberalism, more generally, is

> the desire to intensify and expand the market, by increasing the number, frequency, repeatability, and formalization of transactions. Market forces are also intensified by intensifying assessment, a development especially visible on the labor market. The use of specialized software in call centers has provided some extreme examples: the time employees spend at the toilet is measured in seconds: this information is used to pressure the employee to spend less time away from the terminal. Firms with contracts are also increasingly subject to continuous assessment procedures, made possible by information technology. For instance, courier services use tracking software and GPS technology, to allow customers to locate their packages in transit. (Treanor, n.d.)

To make rational choices, as one is presumed to do within the neoliberal framework, it is necessary to have comparative indices about the performance of alternatives. Because the focus is on individual choice, it is critical to have individual-level performance data. Although the matrix sampling procedure used by the National Assessment of Educational Progress (NAEP), and in the past by some state departments of education, provides sound evidence about the performance at an aggregate level, these approaches provide no data at the individual student, teacher, school, or even district level. The assessment promoted by the neoliberal agenda emphasizes individual choice and therefore requires data at every unit of analysis, from the individual student to whole schooling systems. So, as suggested by Treanor the intensification of assessment in schools reflects neoliberalism's increased surveillance.

Indeed the intensification of assessment takes a number of forms: standardized assessment at legislated grade levels and the narrowing of teaching and learning to conform to the test content and format. The census assessment of, for example, every third- through eighth-grade student in the United States is an obvious form of intensification. In many classrooms, what is taught and learned at all grade levels reflects the substance and form of knowledge on standardized assessments. Research demonstrates that teachers adopt generic forms of content and presentation, develop a test-based curriculum, separate content for the test from "real" content, and fragment knowledge even more than is already the case in schools (Mathison & Freeman, 2003; McNeil, 2000).

The benefit of this intensified assessment is portrayed as necessary to provide information about the quality of schools that parents need to make rational choices about the best educational options for their children. If a school "fails" to educate all of its students, parents are advised to take their business elsewhere. This rhetoric communicates a promise that there are high-quality educational experiences available for all children, and parents must actively make choices to move their children into succeeding schools. That this rhetoric is hollow only minimally diminishes the suggestion that a market of educational options exists and can be chosen if individuals just select the options that produce the highest test scores.

The value of standardized test scores within the neoliberal rhetoric is fortified by values of modernity. There is a predominant view that we truly know something is valuable when it can be objectively measured and statistically manipulated. This view conspires with preoccupations with individualism, competitiveness, the indispensable role of hard work to success, and the equating of equity with sameness.

The Business of Assessment

Neoliberalism's free market exchange encourages the creation of derivative professions and submarkets. In the United States, assessment, especially high-stakes assessment, has created roles for tutors and testing coaches to increase the likelihood of success. The demand for assessment is big business. The U.S. General Accounting Office estimates that, between 2002 and 2008, states spent somewhere between $1.9 and $5.3 billion to develop tests and then score and prepare reports of the test results (Metcalf, 2002). Private, multinational companies that are not accountable to the public are major players in the accountability promises of the neoagendas. The connections between at least McGraw-Hill and the current government are not secret. Indeed, the Bush and McGraw families have had close ties for three generations (Metcalf, 2002).

In addition, these same companies score standardized tests, warehouse and report test results, provide tutoring services, create test preparation materials, and sell a plethora of services that promise to increase test scores. For example, NCLB fosters the expansion of tutoring businesses by at least initially requiring that failing schools use federal funds to contract out extraschool tutoring for students. The need for tutoring has therefore created businesses that assist with the development of tutoring businesses. As one website suggests,

> Tutoring is great for entrepreneurs and home based business opportunity seekers! College students find tutoring to be an excellent part time, flexible job! Homeschooling moms and other homeschoolers are naturals as tutors! Dads can tutor too! Teachers often do tutoring on the side! Tutoring is the perfect job for stay-at-home moms who want to work from home and earn extra money to add to the family income without daycare hassles. (see http://www.cleverapple.com/)

In addition to providing help with setting up one's business, this quote illustrates a coming together of this business opportunity with other values that are often consistent with neoliberalism, such as reinforcing stay-at-home moms, emphasizing the importance of employment, portraying child-care issues as hassles, and supporting nonpublic school options such as home schooling.

The Rhetoric of Assessment-Based Reform

The rhetoric of accountability—leaving no child behind, closing the achievement gap, high expectations, personal effort, and so on seems sensible. The idea of education as a commodity has been naturalized, and the concomitant business language of control, regulation, bottom lines, profits, quality control, and maximized benefits also come to seem sensible. This is especially true for politicians, but many school and university administrators, teachers, parents, and even students are drawn into and live this rhetoric.

However, the rhetoric of assessment and accountability hides a deep-seated advantage of privilege and agendas that undermine educating the populace. By centering the success and failure of schools in children's test scores, attention is diverted away from such issues as reasonable and adequate resources for education and equitable and perhaps free access to educational opportunities. By adopting market perspectives on schooling, families and schools that value positive social development and happiness over competitiveness and progress over achievement

find themselves displaced, floundering to sustain an alternate rhetoric, one that is not shared by those with privilege and capital. Understanding how neoliberalism works against strong schools for all is a necessary step in developing an alternative rhetoric that recaptures schooling as a necessity in a democratic society, one where there are common public interests that will not be served by private interests.

Educational Evaluation in the Public Interest

> What the best and wisest parent wants for his own child, that must the community want for all of its children. Any other ideal for our schools is narrow and unlovely; acted upon, it destroys our democracy. (Dewey, 1907)

What Is the Purpose of Schooling?

Vocationalism and democratic citizenship have long competed as the main purpose of schooling. Taylorism came to U.S. schools early in the 20th century, an approach to schooling that emphasized efficiency of production but developed alongside progressivism's focus on the effectiveness of schools to promote democratic principles. A rationalized system of degrees and credentials has come much later to European universities. Education has always been conceived as an institution that serves the public interest by preparing young people for work and citizenship, promoting a common culture (especially in nations of immigrants), and reducing race, ethnic, and class inequalities. What is critical about these purposes is whether they are conceived in the interest of individuals or the collective, public interest. The current emphasis is on the private and economic benefits (vocational purpose or schooling for the market that serves individual and corporate economic interests), rather than the public benefits (schooling for democratic citizenship with attention to mediating special interests for the common good). (For a discussion of the purpose of public schools, see the Center on Education Policy report, *Why We Still Need Public Schools: Public Education for the Common Good* (Kober, 2007). Neoliberal values currently capture the public attention regarding the purpose of schooling and, consequently, how education is evaluated. As politicians, corporate CEOs, and free marketeers continue to dominant the public rhetoric about the quality (or lack of quality) of schools, so too will the strategies for educational evaluation reflect those values. There is, of course, a debate about whether in the United States, for example, the quality of schools has actually declined (Berliner & Biddle, 1996; Jennings & Hamilton, 2004).

Evaluation as a Democratizing Force in Schooling

The practice of evaluation is typically conceived of as following the lead of the interventions it is meant to judge. Evaluators are seen and see themselves as serving decision-making masters. Evaluators have special methodological expertise, which is employed to answer the questions set by those who control the intervention. For educational evaluation to serve the public interest, this relationship and purpose of evaluation must be disrupted. Such a disruption can occur when evaluators work with all stakeholders in schools and communities, with much greater attention paid to parents, children, and local community perspectives on what counts as good schooling.

An example that illustrates a different rhetoric and methodology for educational evaluation is the Massachusetts Coalition for the Authentic Reform in Education (MassCARE) platform. MassCARE is a statewide organization of primarily parents, but also educators, students, and researchers, who through grassroots organizing are pushing back on the neoliberal agenda for schooling and its purposes. Their goal is to replace the single high-stakes indicator

(the Massachusetts Comprehensive Assessment System or MCAS as it is more commonly called) with an accountability system that defines the quality of schools in fairer and necessarily more complex ways. The MassCARE plan consists of four integrated components:

1. Local authentic assessments that are gateways to graduation, approved by regional boards and based on the Common Core of Learning and a streamlined set of competencies
2. A school quality review model to assess the effectiveness of school practices, based on the models in Britain, Boston Pilot Schools, Rhode Island, and Massachusetts' own process for reviewing charter schools
3. Standardized testing solely in literacy and numeracy, to provide one method for tracking progress of schools from year to year
4. Required annual local reporting by schools to their communities, using a defined set of indicators that also focuses on equal opportunity and access to knowledge for all students. (MassCARE, n.d.)

The MassCARE plan suggests what professional evaluators would typically agree is good evaluation practice: the use of multiple and agreed-on indicators, attention to context, genuine involvement of stakeholders, focusing on improvement given the fundamental formative nature of educational evaluation, and appropriate reporting to schools' stakeholders. This approach to school evaluation is meant to place judgments about the quality of schools and plans for improvement in the hands of local communities, with the state playing an oversight and technical support role, a plan that MassCARE asserts focuses on high standards and allows local innovation and improvement. This is a decidedly New England plan, but it suggests key features of how educational evaluation can serve the public interest. If educational evaluation is to make a contribution in the public interest, then evaluation must be done in ways that permit education's various publics to be involved in determining what is right and good and what is not. Participatory, collaborative approaches to evaluation suit this intention best (Mathison, 2000). An approach, such as the MassCARE plan, has a particular formative purpose, genuinely involves multiple stakeholders, and does so through democratic deliberative processes.

Current educational evaluation is in some senses a backlash against what is seen as too much focus on process, a softhearted approach that looks primarily at intentions but not results. While looking only at what are often called input variables, one certainly has only a partial basis for determining the quality of an educational intervention. The pendulum has, however, swung too far in the other direction, with a focus on a single outcome (student scores on state-mandated achievement tests or international assessments of student achievement). Rethinking educational evaluation as a fundamentally formative task does not diminish the importance of accountability but focuses on forms of accountability other than bureaucratic outcomes-based accountability. The fundamental ameliorative intent of evaluation may be most palpable in public institutions like schooling, where the moral obligation to get things as right as possible is simply what it is about. Getting things as right as possible in such a context is making a commitment to doing evaluation in the public service (i.e., evaluation that contributes the best information possible about how things are working and how to make them better).

Such evaluation requires evaluators to assume leadership roles, ones in keeping with the AEA guiding principle of "responsibility to the general and public welfare." MacNeil (2002) suggests that evaluators should be stewards of citizen deliberation, and Greene (1996) calls on evaluators to be scientific citizens who accept "the assumption of public accountability and social responsibility for the political, moral and value consequences of one's work as a scientist" (p. 278). There are two contexts in which

adherence to this guiding principle must occur—in the doing of a specific evaluation and as a professional evaluation community.

In a particular educational evaluation, evaluators cannot be technicians serving the interests of decision makers, but instead they must accept responsibility for creating an evaluation process that is in the public interest. House and Howe's (1999) deliberative democratic evaluation provides a set of principles for this process: inclusion (considering the interests, values, and perspectives of stakeholders), dialogue (with and among stakeholders), and deliberation (publicly engaged reasoning toward evaluative conclusions). Any participatory, collaborative approach to evaluation will suggest procedures for how to do educational evaluation in the public interest, but Cousins and Whitmore's (1998) transformative participatory evaluation and Lincoln and Guba's (1989) fourth-generation evaluation are two good examples. The extent to which educational evaluation is done in the public service might be judged by using the Deliberative Democratic Evaluation Checklist developed by House and Howe (2000).

But in addition to doing evaluation as a public service, collectively, as a profession, evaluators also need to act like MacNeil's stewards and Greene's scientific citizens. How educational evaluation is done is a matter of both local practice and public policy, and, as such, educational evaluators' participation in the public discourse about the matter is an obligation. The American Evaluation Association (AEA) offers some examples of this, including their statements on high-stakes testing and on educational accountability. This is not a straightforward matter (as illustrated by the dissension within AEA over its response to the U.S. Department of Education's endorsement of randomized clinical trials as the gold standard for educational evaluation; see Donaldson & Christie, 2005) and will inevitably create conflict and discomfort. But as Linda Mabry (personal communication, January 31, 2007) characterized evaluator's responsibility: "I doubt we would deserve more arrows for doing something that might excite debate than we would deserve for fearful disengagement." Such is an essential stance for educational evaluators working in the public interest.

As already mentioned, serving the public interest through educational evaluation needs to recognize that education has multiple stakeholders and that all need to be included in authentic ways. The MassCARE plan illustrates the many stakeholder groups should be involved in the evaluation of schools. The refrain of parent, teacher, student, guardian, and grandparent voices must be as loud as, or louder than, those of the state, politicians, and corporate CEOs. All of these stakeholders are necessary for a fair and democratic evaluation, but the power of money (such as in the corporate-backed NCLB legislation) will not satisfy the need to work in the public interest. The interests of "Voltaire's bastards," as Saul (1992) called the ruling elite so contemptuous of the citizenry, must be balanced by the interests of all.

Incorporating multiple stakeholders in a dialogue will inevitably lead away from a single criterion (such as,student academic achievement in limited areas) and a single indicator (such as, state-mandated test scores). It is not common sense that schools or the quality of the education provided should be judged based on test scores alone. Student achievement in literacy and numeracy is important, but not at the expense of other indicators, such as the academic, social, and physical opportunities and achievements of students; the adequacy of curricular and instructional resources; the opportunities offered to teachers; safety; tolerance; and so on. Whatever the criteria, and given the likelihood many of the same criteria will be invoked in whatever context, there is still the need to respect and include the perspectives of all stakeholders.

Dialogue and deliberation are key for identifying education's stakeholders, engaging them in a process that identifies what the good making attributes of schools are, identifying how best to capture those attributes, making sense of information, and concluding with a forward-looking deliberation about how to make things better.

Conclusion

Participatory, collaborative evaluation approaches that rely on deliberation are fundamental to educational evaluation in the public interest. The current neoliberalism reflects elite, corporate interests that are disdainful of the populace's ability to understand and promote their own interests. Importantly, the kind of educational evaluation suggested here does not guarantee a clear resolution, and indeed many aspects of schooling have been contentious from the common school era to the present (Mathison & Ross, 2008). The role of evaluation is not to settle a matter, make a decision, or take an action that is definitive and immutable. To serve the public interest, we must see evaluation as a continuous process of assessing the particulars to move toward betterment with the implicit expectation that an ideal state or a single solution cannot be attained and does not exist. Evaluation, in this deliberative way, contributes to the heart of democracy because it is the means by which a democratic community maintains its intent and identity, given an indeterminate future.

Notes

1. The term *state* here refers to a geographically bounded, economically organized, and dependent entity that has the power to provide social (e.g., education and police services) and material (e.g., transportation and currency) services and goods for its polity.

2. I have avoided the use of the term *neoconservatism* in this discussion because this term means somewhat different things in different parts of the world. In general, neoconservatism is a political philosophy that rejects liberalism and implies certain roles for the state. In the United States, this means a governmental focus on foreign policy and maintenance of superpower status and the implicit adoption of certain values such as Christianity and heterosexuality. In the Middle East, this political philosophy focuses on the legitimacy of religious authority for governments. Neoconservatism and neoliberalism are not the same, but they often co-exist, and this is especially so in the United States.

References

Altbach, P. (2008). Globalization and the forces of change in higher education. *International Higher Education*, *50*. Retrieved on March 3, 2008, from http://www.bc.edu/bc_org/avp/soe/cihe/newsletter/Number50/p2_Altbach.htm.

Berliner, D., & Biddle, B. (1996). *The manufactured crisis*: *Myths, fraud, and attack on America's public schools*. Addison Wesley.

Bracey, G. (2008). International comparisons: Worth the cost? In S. Mathison & E. W. Ross (Eds.), *The nature and limits of standards-based reform and assessment* (pp. 35–48). New York: Teacher's College Press.

Chomsky, N. (1997). *Media control*. New York: Seven Stories Press.

Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, *80*, 5–23.

DeJarnatt, S. (2003). *The Philadelphia story: The rhetoric of school reform* (ExpressO Preprint Series, Paper No. 71). Available at http://law.bepress.com/expresso/eps/71

Dewey, J. (1907). *School and society*. Carbondale, IL: Southern Illinois University Press.

Donaldson, S., & Christie, C. (2005). Determining causality in program evaluation and applied research: Should experimental evidence be the gold standard? *Journal of Multidisciplinary Evaluation*, *3*. Retrieved on March 3, 2007, from http://www.wmich.edu/evalctr/jmde/content/JMDE%20Num%203_files/Webpages%20JMDE%20003/JMDE_003_Part_I.htm#_Toc116196689

Elmore, R. F. (2002). An unwarranted intrusion. *Education Next*, *2*(1). Available at http://www.hoover.org/publications/ednext/3398996.html

Greene, J. C. (1996). Qualitative evaluation and scientific citizenship: Reflections and refractions. *Evaluation*, *2*(3), 277–289.

Guba, E. G., & Lincoln, Y. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage Publications.

Gutmann, A. (1999). *Democratic education*. Princeton, NJ: Princeton University Press.

Harvey, D. (2005). *A brief history of neoliberalism*. New York: Oxford University Press.

House, E., & Howe, K. R. (1999). *Values in evaluation and social research*. Thousand Oaks, CA: Sage.

House, E., & Howe, K. R. (2000). *Deliberative democratic evaluation checklist*. Retrieved April 25, 2007, from http://www.wmich.edu/evalctr/checklists/dd_checklist.htm

House, E. R., & Mathison, S. (1984). Methodology and values in evaluation. *Educational Theory*, *17*(2), 151–152.

Jennings, J., & Hamilton, M. (2004). What's good about public schools? *Our Children*, *29*(6), 4–6.

Kober, N. (2007). *Why we still need public schools: Public education for the common good*. Washington, DC: Center on Education Policy. Retrieved March 3, 2007, from http://www.cep-dc.org/PublicSchoolFacts/why/

MacNeil, C. (2002). Evaluator as steward of citizen deliberation. *American Journal of Evaluation*, *23*(1), 45–54.

MassCARE. (n.d.) *CARE Plan*. Retrieved on March 12, 2007. Available from http://www.parentscare.org/AuthenticAccount/Authentic_Home.htm

Mathison, S. (2000). Promoting democracy through evaluation. In D. W. Hursh & E. W. Ross (Eds.), *Democratic social education* (pp. 229–241). New York: Falmer Press.

Mathison, S. (2008). A short history of assessment and the standards based movement. In S. Mathison & E. W. Ross (Eds.), *The nature and limits of assessment and standards based educational reform* (pp. 3–14). New York: Teachers College Record.

Mathison, S. (2008). Public good and private interest: A history of educational evaluation. In W. C. Ayers (Ed.), *The handbook of social justice in education*. London: Taylor & Francis.

Mathison, S., & Freeman, M. (2003, September 24). Constraining elementary teachers' work: Dilemmas and paradoxes created by state mandated testing. *Education Policy Analysis Archives*, *11*(34). Retrieved February 1, 2006, from http://epaa.asu.edu/epaa/v11n34/.

Mathison, S., & Muñoz, M. (2008). Evaluation of schools and education: Bad practice, limited knowledge. In S. Mathison & E. W. Ross (Eds.), *The nature and limits of assessment and standards based educational reform* (pp. 71–81). New York: Praeger.

Mathison, S., & Ross, E. W. (Eds.). (2008). *Battleground schools: An encyclopedia of conflict and controversy*. New York: Greenwood Publishers.

Mathison, S., Ross, E. W., & Vinson, K. D. (2006). Defining the social studies curriculum: Influence of and resistance to curriculum standards and testing in social studies. In E. W. Ross (Ed.), *The social studies curriculum: Purposes, problems, and possibilities* (pp. 99-115). Albany, NY: SUNY Press.

McNeil, L. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.

Metcalf, S. (2002, January 28). Reading between the lines. *The Nation*, pp. 12–15.

Nye, J. (2002). Globalism versus globalization. *The Globalist*. Retrieved March 1, 2008, from http://www.theglobalist.com/StoryId.aspx?StoryId=2392.

Olssen, M. (1996). In defense of the welfare state and of publicly provided education. *Journal of Education Policy*, *11*, 337–362.

Puryear, J. M. (1999). The Americas: Educational reforms, external forces, and internal challenges. In D. Morales-Gomez (Ed.), *Transnational social policies*. Ottawa: The International Development Research Centre. Available at http://www.idrc.ca/en/ev-85585-201-1-DO_TOPIC.html

Ross, E. W., & Gibson, R. (Eds.). (2007). *Neoliberalism and education reform*. Cresskill, NJ: Hampton Press.

Saul, J. R. (1992). *Voltaire's bastards: The dictatorship of reason in the west*. New York: The Free Press.

Sedgewick, R. (2003). The Bologna process as seen from the outside. *World Education News & Review*, *16*(5). Retrieved March 2, 2008, from http://www.wes.org/ewenr/03Sept/Feature.htm

Treanor, P. (n.d.). *Neo-liberalism: Origins, theory, definitions*. Available at http://web.inter.nl.net/users/Paul.Treanor/neoliberalism.html

U.S. Department of Education. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*. Available at http://www.ed.gov/rschstat/research/pubs/rigorousevid/index.html

Policies and Programs (with Gary Henry and George Julnes), the *SAGE Handbook of Evaluation* (with Ian Shaw and Jennifer Greene), *What Counts As Credible Evidence in Applied Research and Evaluation Practice* (with Stewart Donaldson and Tina Christie), *Evaluation in Action*: *Interviews With Expert Evaluators* (with Jody Fitzpatrick and Tina Christie), and the forthcoming *Social Psychology and Evaluation* (with Stewart Donaldson and Bernadette Campbell).

Jacob Marszalek is Assistant Professor of Counseling and Educational Psychology in the School of Education at the University of Missouri-Kansas City. His research interests include applying new quantitative techniques to address research questions in education, program evaluation, testing, and psychology. As a student at the University of Illinois, he assisted with several program evaluations funded at the local, state, and national levels. He consults regularly on the design and implementation of grant evaluations in the Midwest.

Sandra Mathison is Professor of Education at the University of British Columbia. Her research focuses on educational evaluation and especially on the potential and limits of evaluation to support democratic ideals and promote justice. She has conducted national large- and small-scale evaluations of K-12, postsecondary, and informal educational programs and curricula; published articles in the leading evaluation journals; and edited and authored a number of books. She is editor of the *Encyclopedia of Evaluation* and co-editor (with E. Wayne Ross) of *The Nature and Limits of Standards Based Reform and Assessment* and *Battleground Schools*; she is co-author (with Melissa Freeman) of *Researching Children's Experiences*; and she is Editor-in-Chief of the journal *New Directions for Evaluation*.

Gerry McNamara, PhD, is Professor of Education at the School of Education Studies, Dublin City University. His research interests include educational evaluation and practitioner research. He is an active member of both the European Evaluation Society and the Irish Evaluation Network, and he is a member of the Council of the British Educational Leadership, Management and Administration Society. Their (see O'Hara) most recent publications include *Trusting Schools and Teachers: Developing Educational Professionalism Through Self-Evaluation* (Peter Lang, 2008) and "The Importance of the Concept of Self-Evaluation in the Changing Landscape of Educational Policy," *Studies in Educational Evaluation*, *34*, 173–179.

Matthew Militello is Assistant Professor in Educational Leadership and Policy Studies at North Carolina State University and has taught at University of Massachusetts, Amherst. Prior to his academic career, Militello was a middle and high school teacher and administrator. His current research centers on preparing school leaders. Most recently, Militello coauthored *Leading With Inquiry and Action*: *How Principals Improve Teaching and Learning*, a book that provides explicit examples of how principals enact collaborative inquiry-action cycles to increase student achievement. He has published in *Education and Urban Society*, *Harvard Educational Review*, *Journal of School Leadership*, and *Qualitative Inquiry*.

David Nevo, Tel Aviv University, Israel, is Professor Emeritus at the School of Education, Tel Aviv University. His professional interests include evaluation theory,