

Today I will discuss the conference themes, which are among the most critical issues facing evaluation. One word of caution: There is an old European saying, a very old saying: "Beware of Greeks bearing gifts." I would update that to "Beware of Americans offering advice," including myself, especially these days.

Neo-fundamentalist policies

A long shadow has fallen across the United States. The American tradition of open government is being seriously eroded. We have a regime in Washington that strives to control information beyond anything in my experience. Of course, all governments hide their mistakes and misdeeds and spin news to their advantage. Of course. But I am talking about control of information that threatens evaluation - and perhaps democracy.

In the wake of the September 11 terrorist attacks, George W. Bush instigated neo-fundamentalist policies, a blend of religious fundamentalism and neo-conservatism (House, 2003). Bush is a born-again Christian, and, in his view, his religious conversion saved him from a life of ruin brought on by drink and drugs. After September 11, he assumed the formidable powers of a wartime president with religious intensity. Since then he has come to believe that he was chosen by God to save the United States and the free world from evil forces (Woodward, 2004).

Here are some characteristics of fundamentalism:

- First, there is one source of truth, be it the Bible, the Koran, the Talmud, whatever.
- Second, this source of authority is located in the past. Believers hark back to that time.
- Third, true believers have access to this fundamental truth but others do not, and applying this truth leads to a radical transformation of the world for the better. Fundamentalists have a prophetic vision of the future, revelatory insight.
- Fourth, having access to the one source of truth means believers are certain they are correct. They have moral certitude, a defining attribute.
- Fifth, fundamentalists are not open to counter arguments, indeed, not open to other ideas generally. They do not assimilate evidence that contradicts their views; they dismiss contrary information or ignore it.
- Sixth, they are persuaded by arguments consistent with their beliefs even when others find those arguments incomplete, illogical, or even bizarre.
- Seventh, people who do not agree with them do not have this moral insight, and fundamentalists do not need to listen to them. In fact, sometimes it is all right to muscle non-believers aside since they impede progress.
- Eighth, believers associate with other true believers and avoid non-believers, thus closing the circle of belief and increasing certainty.
- Ninth, they find ways of promulgating their beliefs by means other than rational persuasion, by decree, laws, or coercion, through forcing others to conform rather than persuading them.
- Tenth, fundamentalists curtail the propagation of other viewpoints by restricting the flow of contrary ideas and those who espouse them.

The Bush administration has exercised this new fundamentalism in foreign and domestic policy, even in evaluation. These policies have been evident in the invasion of Iraq. If the Iraqis denied they had weapons of mass destruction, they were hiding them. If they admitted having such weapons, they were violating the UN mandate. If the war might be disastrous for the region, if most nations in the world were opposed, if world opinion was overwhelmingly against, no matter. Others did not understand. They were "old Europe," unwilling to take risks. The Bush team was closed to counter evidence. They presented arguments seen by others as inconclusive, and at times strange. They concocted a revelatory vision of democratic transformation for Iraq that seemed incredible to Middle East experts. Coercion was used against enemies and allies alike. In short, the fundamentalism of the Muslim terrorists was countered with the fundamentalism of the American President.

Methodological fundamentalism

This authoritarian attitude has spread to other levels of government. In evaluation the new policy I call “methodological fundamentalism.” Some government agencies demand that all evaluations must be randomized experiments. Other ways of producing evidence are not scientific and not acceptable. There is one method for discovering the truth and one method only--the randomized experiment. If we employ randomized experiments, they will lead us to a Golden Age in education and social services. Such is the revelatory vision. There are sacred figures from the past and sacred texts. Core believers are absolutely certain they are right, that randomized experiments reveal the truth. They do not listen to critics, and they do not comprehend the arguments against the method. A mandate for randomized trials has been written into legislation without discussion with professional communities. Avoiding contrary ideas is part of the orientation, and, of course, the policy is enforced by government edict.

Now the problem here is not randomized experimentation as such. It's the belief that there is one source of truth and one only. The error would be equally egregious if the government endorsed qualitative studies as the only source of truth. Wise people know there is no single source of truth, but, alas, wise people are in short supply in the Bush regime. Certainly not everyone who advocates randomized experiments is a methodological fundamentalist. I believe a reasonable case can be made that we should use them more in evaluation.

Relying on randomized field trials is appropriate when you can specify treatments precisely, as in pharmaceutical trials. However, in education and social programs, many uncontrolled factors strongly influence the results. For example, in the Follow Through experiment, the largest evaluation conducted in the US, the same early childhood program placed in Hawaii produced very different results than in New York or Omaha. The reason was that different teachers, different students, different parents, different ethnic groups, and twenty other factors produced varied outcomes when they interacted. In many drug studies these extraneous factors can be controlled, but not in studies where treatments cannot be properly isolated.

In general, the Bush regime manipulates information to conform to policy. Recently, sixty prominent scientists, including twenty Nobel winners, declared that Bush officials deliberately and systematically distort scientific facts to support policy goals (Glanz, 2004). Government officials have changed conclusions of scientific reports and omitted data that did not conform to the new conclusions (Pear, 2004). They have expunged information on their websites that does not reflect policy. Even if governments have the right to protect themselves, surely, they do not have the right to distort scientific findings. In a world in which the government controls and distorts information, what is the role for evaluators? As propagandists? Such control of information threatens not only evaluation but democracy itself.

Politics and Evaluators

Let me generalize. I am not arguing that evaluation should be insulated from politics. Quite the contrary. I am arguing that evaluation is political and that we would do better to face up to this reality. According to Karlsson (2003), following Furubu, Rist, and Sandahl (2002), evaluation in Europe has come in two waves, both driven by government politics. In the 1970s evaluation was used for program development, and in the 1980s governments cut costs, resulting in the New Public Management. EU officials now believe they must demonstrate that their programs produce results to legitimate EU government.

In the US evaluation has followed a similar course. In the 1960s and 1970s the government funded large social programs, and evaluators were asked if those programs worked. The methodology was the large field experiment, like Follow Through. For the most part, those evaluations did not work very well. The findings were too equivocal: a program might work well in one place, be mediocre in another, and fail altogether in a third. Evaluators shifted to evaluating smaller programs using many research methods.

When Ronald Reagan took office, his agenda was to curtail government, and evaluators were asked to find inefficiencies. Performance indicators became common. Later, Clinton believed that if government could not produce goods and services, it could manage them. He called his approach "reinventing government," known elsewhere as the new public management. Again, evaluation was politically driven.

In general, evaluation work is heavily influenced by politics and values outside the professional community. Evaluators are inextricably imbued with value commitments, constrained by the context of the study, the politics of the setting, and the politics of the government. Evaluators are fully "situated" in the deepest sense: valued-imbued, value-laden, and value-based.

The surprising thing is that most evaluators still hold the view that they are insulated from political pressures. They harbor the image of the lone scientist laboring away in the laboratory to produce discoveries that will benefit the world, insulated from the world's forces by research methodology. In fact, evaluating programs is nothing like this. Indeed, the work of scientists is nothing like this either. The image of the lone scientist has more to do with Hollywood than with the communal endeavor that is modern science. Evaluators ask, when will the politics go away so I can attend solely to my evaluation? The answer is, never. Although evaluators try to ignore the politics and values permeating their work, I think this is the wrong approach. We should face up to the value-based nature of our work directly.

The concept of values that has misled us is called the fact/value dichotomy (House and Howe, 1999). It says, facts are one thing and values are something else. The two don't fit together. Facts and values inhabit separate realms. As evaluators we can discover facts, but values are beyond rational investigation, something people have deep inside them. Values might be feelings or emotions. Whatever they are, they are not subject to scientific analysis. People simply hold certain values or do not.

By contrast, I contend that we can deal with both fact and value claims rationally. Indeed, facts and values are not separate kinds of entities. Facts and values blend together in our evaluations. We can better conceive facts and values as being on a continuum like this:

What we call facts and values are fact and value claims, beliefs about the world. Sometimes these beliefs look as if they are strictly factual without any value component, such as, "Diamonds are harder than steel." This statement may be true or false, and it fits at the left end of the continuum. There is little individual preference, taste, or value evident.

On the other hand, a statement like "Cabernet is better than chardonnay" fits at the right end of the continuum. It is suffused with personal taste. But just because we can distinguish between the fact and value ends of this continuum does not mean facts and values are totally different kinds of entities. Consider a statement like, "Follow Through is a good educational program." This statement encompasses both fact and value. The claim is based on criteria from which the conclusion is drawn and supporting facts. It lies in the middle of the continuum, a blend of fact and value. Indeed, if you examine evaluation reports closely, you will find that facts and values are entangled so intricately it is difficult to pull them apart.

I believe evaluative claims are subject to rational analysis in the way we ordinarily understand rational thought. First, the claims can be true or false. Follow Through may or may not be a good program. Second, we can collect evidence for and against the truth or falsity of that claim. Third, the evidence we use can be biased or unbiased. Fourth, the procedures for determining whether claims are biased or unbiased are decided by the discipline, in forums like this. In other words, evaluators can investigate value claims rationally based on our discipline.

Now this conception of facts and values is different from the old fact-value dichotomy. In the old view, to the extent evaluative conclusions were value-based, they were outside the purview of evaluators. In the new view, value claims are subject to rational analysis by evaluators. Indeed, values that are carefully considered are evaluations.

Value-based Evaluation

So much for theory. But how can we deal with politics and values in actual studies? Here's an example. For years the public schools in Denver, Colorado, have been under federal court order to provide Spanish language instruction for students who do not speak English until those students learn English. This population includes fifteen thousand of Denver's seventy thousand students. These students are new immigrants, mostly from Mexico. In 1999 the legal plaintiffs in the court case--the Congress of Hispanic Educators and the US Justice Department--reached an agreement with the Denver school district as to what a proper bilingual program for these students should be. With the approval of the contending parties, the presiding judge appointed me court monitor, my task being to monitor whether the program was implemented.

The passions on both sides were highly inflamed. Providing Spanish language instruction for immigrants is an explosive issue. For years the school district and the plaintiffs have displayed a complete lack of trust in each other. In conducting a monitoring evaluation, I tried to reduce the distrust by involving the major stakeholders in the evaluation and making what I did transparent. I brought the leaders of the contending groups together face-to-face twice a year to discuss the findings and the conduct of the on-going evaluation. Since many participants were lawyers (adversarial by occupation), these meetings had many conflicts.

For data I constructed a checklist based on the key elements of the program to assess each school. I submitted the checklist to all parties and used their recommendations to revise it. I hired two retired school principals to visit and assess individual schools with the checklist. Since they were former principals, the school district trusted them. Since they were Latinos and supported bilingual instruction, the plaintiffs trusted them. I encouraged the school district staff to challenge the evaluation of each school where they disagreed. We hashed out disagreements face-to-face. (Eventually, the school district developed its own checklist to anticipate which schools might have problems.)

I met with interested groups in the community, including the most militant, those bitterly opposed to bilingual programs and those who wanted total bilingual schools. I listened, responded to their concerns, and included their ideas in my investigations. I followed up on information they provided using traditional research methods. I thought about holding meetings open to the public but decided against such meetings since I was afraid they would degenerate into shouting matches. The emotions were too raw. I developed quantitative performance indicators of program success based on the school district's data management system. I discussed the indicators with all parties until everyone accepted them as indicators of progress.

My written reports went to the presiding judge three times a year. As court documents the reports were public information the local media seized on. I asked the school district officials and the plaintiffs how I should handle these requests. They preferred that I not talk to the media. It would inflame the situation. So I referred all inquiries to the stakeholders and made no public comments beyond my written reports.

Many different issues arose over a five-year period. For example, the lawyers representing the Latinos suspected the school district was forcing schools to move students into English classes prematurely. So I paid close attention to the proficiency level of the students when they were transferred to English and to the procedures used to assess them. Lawyers from the US Justice Department were afraid students would be taught with inferior materials. So we assessed the Spanish versus English teaching materials to ensure the quality was similar. Even the lawyers on the same side had different concerns.

The Latino community itself was divided. The group descended from the 17th century Santa Fe culture saw themselves as Spanish-Americans, not Mexican. They held many professional jobs in the schools and spoke English and Spanish. Most students were immigrants from rural Mexico who had no formal education, even in Spanish. The cultural and class differences between these groups affected the program. For example, immigrant parents took their children out of school for weeks during school term to return to Mexico for village fiestas, which infuriated the Spanish-Americans.

The parents themselves disagreed. Some wanted their children immersed in English immediately so the students could get jobs. Most wanted their children in Spanish first, then English. Legally, parents could choose what their children should do. We discovered that many schools did not make these choices clear to parents. So we attended to whether the options were presented to parents at each school in ways they could understand.

The most militant Latino group in the city wanted full cultural maintenance of Spanish rather than transition to English. I met with the leader of this group in the café that served as political headquarters in the Latino part of Denver. I listened to her concerns. There was little I could do about cultural maintenance since I had to work within the court document, which precluded it. However, I could investigate issues that caused her to distrust the schools. For example, some school principals were not identifying Spanish-speaking students because they were afraid their teachers would be replaced by Spanish teachers. We reported this to school district officials, who resolved the problem.

The issues were numerous. What was not an issue was how the program compared to other ways of teaching English in general. A randomized field experiment would have convinced no one. There was one group wanting no bilingual instruction, but they had already made up their minds. (I met with the leader of this group and suggested she read my reports to see how long it took students to enter English classes. She led an unsuccessful state campaign to eliminate bilingual instruction throughout the state. Interestingly, the school district and plaintiffs joined forces against her.)

Now, after five years--preceded by twenty years of militant strife - the program is almost fully implemented. The issue seems to be defused for the school district. The opposing groups can meet in a room without casting insults at each other. I am not saying the groups love each other, but they can manage their business together rationally. The conflict is nothing like when we started.

In summary, the Denver evaluation dealt with the politics of the program. It dealt with specific issues arising from the views, values, and interests of those most concerned. The face-to-face meetings among the key stakeholders proved critical. In addition, the transparency of my actions as evaluator was also critical. Without stakeholders understanding what I was doing, I don't believe trust could have evolved. The evaluation became a mode of communication, negotiation, and common understanding. During the study, I employed the usual research methods we use - checklists, tests, and performance indicators. What was different was how the study was framed.

Deliberative Democratic Evaluation

I believe this evaluation incorporated a democratic process that gave voice to stakeholders. Its legitimacy to participants rested on fair, inclusive, and open procedures for deliberation, where those in discussion were not intimidated or manipulated (Stein, 2001). Of course, those involved still do not agree on all issues, and they never will. Some value disagreements are intractable, but that does not mean we cannot handle them.

I call this approach deliberative democratic evaluation. The three principles are inclusion of all relevant stakeholder views, values, and interests; extensive dialogue between and among evaluators and stakeholders so they understand one another thoroughly; and deliberation with and by all parties to reach conclusions (House and Howe, 1999). The conclusions might be jointly constructed rather than made entirely by the evaluator. (A checklist for deliberative democratic evaluation is on the website of the Evaluation Center at Western Michigan University: www.wmich.edu/evalctr/checklists).

By enlisting stakeholders at many points, the evaluators' role is extended beyond the traditional. Since a range of views, values, and interests are considered, the hope is that the conclusions will be better, that participants will accept and use the findings more, and that an evaluation becomes a democratic practice that faces up to the political, value-imbued situation evaluators often find themselves in.

The important thing is not the particular approach but the general direction. Jennifer Greene (2003) has suggested there are three ways of addressing value issues. First, some evaluators see themselves as politically neutral through employing research methods. Second, some accept the value-laden character of their work and try to express multiple values in their studies, but still strive to maintain non-partisan credibility. A third way is to engage with politics and values directly to address particular values and interests. Democratic evaluators are in this last category.

Compatible ideas have been advanced by MacDonald (MacDonald and Kushner, 2004), Simons (1987), Kushner (2000), and Saunders in the UK, Barry MacDonald being the first to develop a concept of democratic evaluation back in the 1970s; by Karlsson (1996, 2003), Segerholm (2003), Hanberger (2001), Murray (2002), and Vedung, in Sweden, Krogstrup (2003) in Denmark, and Mosen in Norway, the Scandinavians having carried democratic ideas farther than anyone. In Australia, Elsworth and Rogers have introduced such ideas into their work. In Canada, Cousins and Whitmore (1998) have stressed participatory evaluation, and in the US Greene (2003), Schwandt (2003), King (1998), Ryan (Ryan and DeStefano, 2000), Mark, Henry, and Julnes (2000), and Patton (2002) have addressed similar issues. I have learned something from all these people and others, though perhaps they would say I have not learned enough.

Why Now?

Finally, why should we develop new forms of evaluation now after forty years of practice? Sometimes it helps to think about evaluation as a social institution that develops over long periods of time, what the French Annales historian Fernand Braudel called the "long duration." Braudel's (1981) analysis of capitalist institutions from the 15th to 18th centuries is a classic.

In the long view, evaluation exists as a social practice because it provides legitimacy for government actions, one way in which governments seek legitimacy. It is no accident that program evaluation emerged early in the US, the most capitalist country, since American government is constantly pressured by corporations and the private sector. Nor is it surprising that EU officials believe they must demonstrate their programs are effective to justify the powers they have assumed.

But societies change, and the old relationship between government and evaluation may not suffice, an idea I owe to Mauro Palumbo from Italy (Palumbo, 2002). Societies (and programs) are becoming more complex. Political demands have intensified, and these demands come from more diverse groups. In some ways life is less certain and less predictable. Countries face economic dislocation as jobs disappear, and people migrate in large numbers. With capital free to cross borders, workers follow. Capitalism is a radical social force.

In some countries the economic structures are more difficult to justify. Inequalities in income and wealth have increased dramatically. The United States has the most unequal distribution of income and wealth of any developed country. Other countries have also become more unequal, contrary to political promises and public expectations. Economic well-being is central to government legitimacy.

Perhaps most critical to the legitimacy of regimes is maintaining security. Terrorist threats have given both the US and Russian governments reasons to become more authoritarian. Indeed, fundamentalisms of all types are responses to these same social forces. Meanwhile, information flows at a stunning pace via the internet. The pace itself challenges the way we do things. I believe the move towards randomized experiments and the move towards democratic evaluation are both responses to these social changes. Randomized experiments promise legitimacy by increasing methodological rigor, and democratic evaluations promise legitimacy by increasing transparency and participation. If I were an EU official, I would worry about showing what is happening inside government rather than just results, though both are important. You cannot substitute one for the other.

I began this talk with a warning derived from Virgil's Aeneid to beware of Greeks bearing gifts. One of Virgil's purposes in writing the Aeneid was to legitimate the rule of Augustus. How well invoking the Gods and inventing a myth about the founding of Rome succeeded in legitimating the Julian line of rulers, I don't know. I do know that modes of legitimation can lose their credibility. I end with a warning from a Canadian - political scientist Janice Gross Stein - based on her analysis of performance indicators and the new public management. Stein says, "Political leaders often prefer to put the debates that engage our important and contested values into a supposedly neutral measuring cup. They do so to mask the underlying differences in values and purposes, and to dampen political disagreements. They seek the consensus they need and the political protection they want by transforming conflict over purposes into discussions of measures, and in the process they hide and evade differences about values and goals. But ... numbers cannot bear the political burden they are being asked to carry" (Stein, 2001, p. 198).

REFERENCES

- Braudel, F. (1981, 1982, 1984) *Civilization and capitalism: 15th to 18th centuries* 3 vols. New York: Harper Row.
- Cousins, J. B. & Whitmore, E. (1998) Framing participatory evaluation. In E. Whitmore (Ed.), *Understanding and practicing participatory evaluation* San Francisco: Jossey-Bass, 5-23.
- Fetterman, D. (2001) *Foundations of empowerment evaluation* Thousand Oaks, CA: Sage.

- Furrobo, J-E., Rist, R., & Sandahl, R. (Eds.). (2002) *International atlas of evaluation* New Brunswick: Transaction.
- Glanz, J. (2004) Scientists say administration distorts facts *New York Times* (February 9)
- Greene, J. (2003) War and peace...and evaluation. In O. Karlsson (Ed.). *Studies in Educational Policy and Educational Philosophy* (2). Sweden: Uppsala University (www.upi.artisan.se).
- Hanberger, A. (2001) Policy and program evaluation, civil society, and democracy *American Journal of Evaluation* 22 (2): 211-228.
- Mark, M. M., Henry, G. T., Julnes, G. (2000) *Evaluation* San Francisco: Jossey-Bass.
- House, E. R. (2003) Bush's Neo-Fundamentalism and the new politics of evaluation. In O. Karlsson (Ed.) *Studies in Educational Policy and Educational Philosophy*, 2. Sweden: Uppsala University (www.upi.artisan.se)
- House, E. R. & Howe, K. R. (1999) *Values in evaluation and social research* Thousand Oaks, CA: Sage.
- Karlsson, O. (2003) Evaluation politics in Europe: Trends and tendencies. In O. Karlsson (Ed.). *Studies in Educational Policy and Educational Philosophy* 1. Sweden: Uppsala University (www.upi.artisan.se).
- Karlsson, O. (1996) A critical dialogue in evaluation: How can interaction between evaluation and politics be tackled? *Evaluation* (2), 405-416.
- King, J. A. (1998) Making sense of participatory evaluation. In E. Whitmore (Ed.), *Understanding and practicing participatory evaluation* San Francisco: Jossey-Bass, 57-67.
- Krogstrup, H. K. (2003) User participation in evaluation--Top down and bottom-up perspectives. In O. Karlsson (Ed.). *Studies in Educational Policy and Educational Philosophy* 1. Sweden: Uppsala University (www.upi.artisan.se)
- Kushner, S. (2000) *Personalizing evaluation* London: Sage.
- MacDonald, B. and Kushner, S. (2004) Democratic evaluation. In S. Mathison (Ed.). *Encyclopedia of evaluation* Thousand Oaks, CA: Sage.
- Murray, R. (2002) Citizens' control of evaluations. *Evaluation* 8 (1): 81-100.
- Palumbo, M. (2002) *Quality and quantity in evaluation, social research, and democracy*. European Evaluation Society, Seville, Spain.
- Patton, M. Q. (2002) A vision of evaluation that strengthens democracy. *Evaluation* 8:1, 125-139.
- Pear, R. (2004) Taking spin out of report that made bad into good health. *New York Times* (February 22)
- Ryan, K. A. and DeStefano, L. (Eds.). (2003) *Evaluation as a democratic process: Promoting inclusion, dialogue, and deliberation* New Directions for Evaluation, 85. San Francisco: Jossey-Bass.
- Schwandt, T. A. (2003) In O. Karlsson (Ed.). *Studies in Educational Policy and Educational Philosophy* 2. Sweden: Uppsala University (www.upi.artisan.se).
- Segerholm, C. To govern in silence: An essay on the political in national evaluations of the public schools in Sweden. In O. Karlsson (Ed.). *Studies in Educational Policy and Educational Philosophy* 2. Sweden: Uppsala University (www.upi.artisan.se).
- Simons, H. (1987) *Getting to know schools in a democracy* London: Falmer.
- Stein, J. G. (2001) *The cult of efficiency*. Toronto: Anansi.
- Woodward, B. (2004) *Plan of attack* New York: Simon and Schuster.

Paper presented to the European Evaluation Society, Berlin, October 2, 2004.