

THE ROLE OF THE EVALUATOR IN A POLITICAL WORLD

Ernest R. House
University of Colorado
Boulder, Colorado

FOREWORD

Alan G. Ryan
University of Saskatchewan

Throughout his long and distinguished career, Ernest House has continuously stressed the moral responsibility of evaluators. His social activist perspective has time and again alerted us to the dangers of being seduced by the agendas of those in power. (It was this stance that made him a particularly appropriate keynote speaker for Saskatchewan's first CES annual conference; he is well versed in Saskatchewan's history of co-operatives and social initiatives.) In his keynote address, he points out that the current political climate in the United States presents a threat to the independence and utility of evaluation, that is, the threat of becoming a servant of the power elite. Using Janice Gross Stein's analysis of the cult of efficiency, he shows how political fundamentalism and methodological fundamentalism are intimately linked. As he wrote over 25 years ago in his monograph *The Logic of Evaluative Argument* (1977): "There are those who try to force simplicity atop the complexities of life and thereby eradicate ambiguity ... Often in positions of power, they impose arbitrary definitions of reality for the sake of action" (p. 47).

But House has never been one to wring his hands at the hopelessness of a situation. His prescription for evaluators in the face of such forces is for us to confront the political exigencies head on. We need to challenge simplistic assumptions that impose artificial limits on evaluation questions and methodological approaches. In the extended example from his own work, he takes us through an alternative role for the evaluator, a sort of Transactional Evaluation approach (Rippey, 1973), but one that is updated through a sophisticated per-

Corresponding author: Ernest R. House, Campus Box 249, University of Colorado, Boulder, CO, USA 80309; <Ernie.House@colorado.edu>

ception of the political forces at work in the study being undertaken. In so doing, he makes clear that the responsibilities of the profession of evaluation go far beyond any accountability model.

REFERENCES

- House, E.R. (1977). *The logic of evaluative argument*. Los Angeles: Center for the Study of Evaluation.
- Rippey, R.M. (1973) *Studies in transactional evaluation*. Berkeley, CA: McCutchan.

AVANT-PROPOS

Alan G. Ryan
University of Saskatchewan

Pendant sa carrière longue et distinguée, Ernest House a constamment souligné l'importance de la responsabilité morale des évaluateurs. Son activisme social nous a maintes fois mis en garde contre l'attrait que peuvent susciter les agendas des personnes en position de pouvoir. (C'est cette position qui faisait de lui un conférencier particulièrement pertinent pour le premier congrès annuel de la SCÉ en Saskatchewan; il connaît bien l'histoire de cette province en ce qui a trait aux coopératives et aux initiatives sociales.) Dans son allocution, il mentionne que le climat politique actuel aux États-Unis constitue une menace pour l'autonomie et l'utilité de l'évaluation, c'est-à-dire, celle de devenir un serviteur de l'élite au pouvoir. En se basant sur l'analyse du culte de l'efficacité de Janice Gross Stein, il montre comment le fondamentalisme politique et méthodologique sont intimement liés. Comme il l'a écrit il y a plus de 25 ans dans son ouvrage intitulé *The logic of evaluative argument* (1977): « [traduction] Il y en a qui essaient d'appliquer de force une simplicité aux complexités de la vie et qui éliminent par le fait-même l'ambiguïté... Étant souvent dans des positions de pouvoir, ils imposent des définitions arbitraires de la réalité pour justifier leurs gestes. » (p. 47).

Mais House n'a jamais été de ceux qui se laissent aller au désespoir, et ce, quelle que soit la situation. Il recommande aux évaluateurs qui font face à de telles forces de confronter directement les instances politiques. Nous devons remettre en cause les hypothèses simplistes qui imposent des limites artificielles aux questions d'évaluation et

aux approches méthodologiques. Dans l'exemple détaillé tiré de ses propres travaux, il nous présente un rôle nouveau pour l'évaluateur, une sorte d'approche transactionnelle à l'évaluation (Rippey, 1973) mais qui est mise à jour grâce à une compréhension raffinée des forces politiques en jeu dans l'étude entreprise. Ce faisant, il souligne clairement que les responsabilités de la profession d'évaluation vont bien au-delà de tout modèle d'imputabilité.

RÉFÉRENCES

- House, E.R. (1977). *The logic of evaluative argument*. Los Angeles: Center for the Study of Evaluation.
- Rippey, R.M. (1973) *Studies in transactional evaluation*. Berkeley, CA: McCutchan.

■ In her Massey lectures in 2001, Janice Gross Stein recalls the time her 85-year-old mother shattered her hip and required surgery. Stein began a frantic search for someplace her mother could live when she left the hospital. On the seventh day after surgery, a hospital administrator stopped Stein in the hallway.

“When are you removing your mother from the hospital?”

“Ah, we need a little more time to find someplace she can live,” Stein said.

“Your mother is now a negative statistic for this unit. Every additional day she remains in hospital, she drives our efficiency ratings down.”

Stein, a political scientist at the University of Toronto, addresses this “cult of efficiency” in her lectures. She says, “Efficiency turned inward and became silent about values, neutral about goals, but vocal about means ... efficiency becomes a code word for an attack on government as a provider of public goods” (Stein, 2001, p. 28, 35).

Well, what's the problem here? How is it that performance indicators, an important tool of evaluators, should play an ugly role in people's lives? Are evaluators complicit in practices that few of us would condone? Or is it that politicians and administrators have

misused tools that we've given them? There is some merit in the misuse idea, but I think that we evaluators also have to take responsibility. It's not just them. It's us, too.

HOW WE GOT HERE

How did we get to this place? First, let me provide a thumbnail sketch of evaluation policy in the U.S.A. (I am acutely aware that Canada is not the U.S.A.—lucky you—but if I attempt such an account for Canada, I am bound to make mistakes.) In the U.S., big-time evaluation was first mandated in the Great Society legislation of Lyndon Johnson. The idea was that the federal government could fund large social programs, and evaluators could determine whether those programs worked. The method of choice was the large field experiment utilizing control groups, with the Follow Through evaluation a paradigm case. For the most part, these studies did not work very well. The findings were too equivocal: a program might work splendidly in one place, be mediocre in another, and fail altogether in a third. Evaluators shifted to evaluating smaller programs using a variety of methods.

With variations, this phase lasted until Ronald Reagan. His agenda was to drastically curtail the size of the government, a policy legitimated by Chicago economics. The economic theory was that governments spend money poorly and that pushing money into the private sector is more efficient. This was a direct attack on the welfare state. During the Reagan years, the role for evaluators was to find inefficiencies in government to reduce expenditures, a search that included everything from spotting weak programs to kicking welfare mothers off the dole. Performance measures, such as high stakes testing in education to spot suspect schools and underperforming students, became commonplace.

Clinton had a more favourable view of the welfare state, but not enough political support to reverse the anti-government trend. He attempted a compromise — “reinventing” government. If government could not *produce* goods and services, it could *manage* them. The paradigm is the “M-structure” of the corporation. Central management becomes a small group of administrators and evaluators who control and monitor the activities of several profit centres that operate autonomously. For example, if education is privatized, the government does not produce educational services, but manages them through inspections and indicators. Evaluators play a moni-

toring function. Hence, Stein's encounter with the hospital coordinator in Toronto.

In the U.S. there is one more (sad) chapter — George W. Bush. Bush has instigated neo-fundamentalist policies, a melding of religious fundamentalism and neo-conservatism (House, 2003). Here are 10 characteristics. First, there is one source of truth, be it the Bible, the Koran, the Talmud, or 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 source of truth means believers are certain that 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 outsiders find those arguments incomplete, illogical, or even bizarre. Seventh, people who do not agree with them do not have this insight, and fundamentalists do not have to listen to them. In fact, sometimes it is all right to muscle non-believers aside since they don't understand and they only 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, policy, or laws, through forcing others to conform rather than persuading them — in short, through coercion. Tenth, fundamentalists try to 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 policies. In foreign policy it has been evident in the invasion of Iraq. If the war might be disastrous for the region, if most nations in the world were opposed, if world public opinion was overwhelmingly against, no matter. Others didn't 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. The more criticism from outside, the more they banded together. Coercion was the tool of choice for compliance, used against enemies and allies alike. The fundamentalism of the Muslim terrorists was countered with the new fundamentalism of the American president.

In the field of evaluation, the Bush policy has been methodological fundamentalism. Some government agencies, like the U.S. Department of Education, have demanded that all evaluations must be based on 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 social services. Such is the revelatory vision. A mandate for randomized trials was written into legislation without discussion with professional communities. Avoiding contrary ideas is part of the orientation. And the policy is enforced by government decree.

Now the problem here is not randomized experimentation as such. The problem is the policy that there is one source of truth and one only. The error would be equally egregious if the government endorsed qualitative studies or performance indicators or any method as the only source of truth. Wise people know there is no single source of truth. Not surprisingly, the American Evaluation Association has strongly criticized this policy of the U.S. government.

As for randomized experiments, a reasonable case can be made that we should use them more often in evaluation. Relying on field trials is appropriate when you can specify and isolate treatments precisely, as in pharmaceutical trials. However, in many education and social programs, many other uncontrolled factors influence the results. For example, in the Follow Through experiment, the same early childhood program placed in Hawaii produced very different outcomes in New York and Omaha. The reason is that different teachers, different students, different parents, different ethnic groups, and 20 other factors can produce varied outcomes when they interact with each other. These factors can be controlled in many drug studies but not in studies that cannot be so isolated. To say all studies should be based on randomized experiments is ridiculous. (For discussions, see Scriven, in press, and Maxwell, 2004.)

In the Bush world, information is restricted and controlled to conform to the official line in general. Recently, 60 prominent scien-

tists, including 20 Nobel laureates, issued a statement saying that the Bush administration deliberately and systematically distorts scientific facts to support policy goals (Glanz, 2004). One must wonder, in a world in which the government controls and distorts information, what is the role for evaluators? Fortunately, Canada does not have to contend with such forces, but you never know what might creep across the border camouflaged as a maple leaf.

THE ROLE OF THE EVALUATOR

I now want to generalize. Evaluation is political. It is deeply affected by external forces. What evaluators do is heavily influenced by political factors outside the professional community. Evaluators are inextricably bound by value commitments and political constraints, only some of which they themselves decide. When they conduct studies, they are constrained by the context of the study, the politics of the setting, and even the politics of the government. Evaluators are fully “situated” in the deepest sense: value-imbued, value-laden, and value-based.

Now this is not news to anyone long in the profession. But, surprisingly, we still hold a view of evaluators as value-neutral and insulated from political pressures. It is the image of the lone scientist labouring away in the laboratory to produce discoveries that benefit the world, isolated from political forces, protected by methodology and theory. In fact, evaluating programs is nothing like this. Actually, 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 endeavour that is contemporary science.

Back in Canada, Stein puts the situation this way: “Judgements about the effectiveness of public goods are inevitably laced with political claims, but without those judgements, no calculation of cost effectiveness can be made. Technical language should not obscure the fundamentally political calculations of efficiency in public goods” (Stein, 2001, p. 71).

What we evaluators have often tried to do about the politics and values permeating our work is to ignore them, to pretend that we are value-neutral and immune to politics, value commitments, and job pressures. We pretend that methods, theories, and dedicating ourselves to discovering the facts protect our neutrality. But I think

this is the wrong approach. I think we should face up to the value-based nature of our work.

The conception of values that has misled us is called the fact/value dichotomy. It says that facts are one thing and values are something else. The two don't come 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, perhaps. Values might be feelings or emotions. Whatever they are, they are not subject to scientific analysis. People simply hold certain values or do not.

That conception is wrong. By contrast, I contend that we can deal with both facts and values rationally. Indeed, facts and values are not separate kinds of entities, though they sometimes appear that way. Facts and values blend together in our evaluation studies. We can better conceive facts and values as being on a continuum like this:

Brute Facts _____ **Bare Values**

What we call facts and values are fact and value claims. They are beliefs about the world. Sometimes these beliefs look as if they are strictly factual without any value component built in, 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 in it.

On the other hand, a statement like “Cabernet is better than chardonnay” fits better at the right end of the continuum. It is suffused with personal taste. What about a statement like “Follow Through is a good educational program”? This statement encompasses facts and values. The evaluative claim is based on criteria from which the conclusion is drawn and on facts to support it. The statement lies toward 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 tightly it is difficult to pull them apart.

I believe that evaluative claims are subject to rational analysis in the way we ordinarily understand rational analysis. First, the claims can be true or false. Follow Through may or may not be a good educational 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, good or bad. Fourth, the procedures for determining whether

claims are biased or unbiased are decided by the discipline. In short, evaluators can investigate value claims rationally.

This conception of facts and values is quite 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, values are subject to rational analysis by evaluators. Indeed, values are evaluations, and evaluative conclusions are reasoned fact-value claims.

VALUE-BASED EVALUATION

What difference might this conception make for conducting evaluations? To return to Stein's analysis of performance indicators, "The critical question then becomes, who defines the standards? Is accountability to be imposed or negotiated through a deliberative political process that gives voice to all the stakeholders...?" (Stein, 2001, p. 81). She continues,

An open, democratic process to construct the content of accountability—what accountability means, how it is measured, and to whom providers are accountable—is no less than what citizens deserve and expect. It is also the best—indeed the only—hope governments have for active, engaged providers that are committed to the delivery of high-quality public goods. (Stein, 2001, p. 185)

I agree with Stein, and I advocate procedures for collecting and analyzing data that involve those who are most concerned about the program under review. The best way to understand this is through an example.

For years the Denver Public Schools 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 15,000 of Denver's 70,000 students. In 1999 the plaintiffs in the federal court case — the Congress of Hispanic Educators and the U.S. Justice Department — reached a 50-page agreement with the Denver school district as to what the program would be. With the approval of the contending parties, the presiding judge appointed me court monitor. My task was to monitor for the court whether the program was being implemented and to report to the court, the school district, and the plaintiffs.

The passions on all sides were highly inflamed. Providing Spanish-language instruction is an explosive issue in Colorado. For decades the school district and the plaintiffs have had a total lack of trust in each other. In conducting a monitoring evaluation, I decided to try to reduce the distrust by involving the stakeholders in the evaluation and by making everything I did transparent. I brought the leaders of the contending parties face-to-face twice a year to discuss the findings of my ongoing evaluation. Since many participants were lawyers (adversarial by occupation), the meetings had many conflicts.

For data I devised a checklist that included the key elements of the program with which to rate each of the 100 schools. I submitted the checklist to both parties for review, and used their recommendations to revise the form. I hired two retired school principals to visit and rate individual schools. Because they were former principals, the school district trusted them. Because they were Hispanics, spoke fluent Spanish, and supported bilingual instruction, the plaintiffs trusted them. I encouraged the district staff to challenge the ratings for each school when they disagreed. We hashed out differences face-to-face. (I could have obtained better inter-rater reliability by using graduate students, but the students would not have fully understood the issues or known where the bodies were buried. They could have been easily fooled and resented by school principals.) Eventually, the school district developed its own checklist based on mine so they could anticipate which schools would have problems.

I met with the involved groups in the community, including the most militant — both those bitterly opposed to bilingual programs and those who wanted full bilingual schools. I listened, responded to their concerns, and included their ideas in my investigations. I followed up on information they provided about schools. I considered holding open public meetings but decided against it since I was afraid such hearings would degenerate into shouting matches. I developed indicators of program implementation based on the district's own data management system. I discussed these indicators with both parties until everyone accepted them as significant measures of progress.

My periodic written reports went directly to Judge Matsch, the presiding judge (also the judge in the Oklahoma bombing trial). As court documents, the reports were public information that the local media seized on. I asked the district and the plaintiffs how I should handle the media. Both said they preferred that I *not* talk to the media, as it would inflame the situation. So I referred media inquir-

ies to the two parties and made no public comments outside my written reports. The media accepted this arrangement, albeit reluctantly.

Here are a few of the many issues that arose, and how we dealt with them. One set of lawyers represented the Hispanic educators who brought the original lawsuit. They were concerned with making certain the students were taught in their native language until they learned English and that students not be forced into mainstream classes prematurely. The plaintiffs suspected the district was forcing schools to move students to mainstream classes. So we paid close attention to the students' proficiency level in English when they were transferred and the procedures for identifying and exiting them.

The co-plaintiff in the case was the U.S. Justice Department. These lawyers were concerned that students receive an equal education while they were in Spanish-language classes. So we rated the difficulty of Spanish versus English materials to ensure the teaching materials in the two languages were similar and assessed the materials available in Spanish. Even the plaintiff lawyers had somewhat different perspectives.

Most immigrant parents wanted their children in Spanish-language classes first, then English. But some wanted their children to go directly into English classes so they could get jobs. Parents had a choice as to what their children should do. However, we discovered that many schools did not make these choices clear to parents. We examined whether the program options were presented to parents in ways they could understand.

The most militant Hispanic group wanted full cultural maintenance of Spanish, in addition to English. I met with the leader of this group in the café that served as political headquarters in the Hispanic part of town, and listened to her concerns. There was little I could do about cultural maintenance since the court agreement, which I had to work within, did not include that. However, I did investigate practices that reinforced her view that the school district was insincere. For example, some schools were deliberately not identifying their Spanish-speaking students because the principals were afraid some teachers would be replaced by Spanish-speaking teachers. We reported this to the district, which took actions to ameliorate the problem.

The issues were numerous. What was *not* an issue was how the program compared to other ways of teaching English, an issue that

might have been addressed by a randomized experiment, an evaluation that would have convinced no one, regardless of the results.

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

I believe this evaluation meets Stein's idea of an open democratic process that defines the content of evaluation standards through a deliberative process that gives voice to all stakeholders — what accountability means, how it's measured, and to whom providers are accountable. Of course, those involved do not agree on all issues. It was never my goal to reach consensus. Stein says:

The conflict among these values is often intractable and incommensurable. It is because these conflicts are intractable that we turn to conversation in public space, and to those we choose to govern, to set legitimate rules for a conversation that is not about interests, but about principles and values. The legitimacy of this conversation rests on recognized, fair, inclusive, and open procedures for deliberation and persuasion, where those who join in reflective discussion are neither intimidated nor manipulated. (Stein, 2001, p. 225)

DELIBERATIVE DEMOCRATIC EVALUATION

I call my particular approach deliberative democratic evaluation, and it rests on three general principles: *inclusion* of all relevant stakeholder views, values, and interests in the evaluation; extensive *dialogue* between evaluators and stakeholders and among stakeholders so they understand one another; and *deliberation* among and by all parties to help reach valid conclusions (House & Howe, 1999). A suggestive checklist for deliberative democratic evaluation is on the website of the Evaluation Center at Western Michigan University <<http://www.wmich.edu/evalctr/checklists>>.

By enlisting the participation of diverse stakeholders at many points in the study, the evaluators extend their role beyond the traditional one. I expect the outcome to be better conclusions, since a range of

views, values, and interests have been considered; more acceptance and use of findings; and an evaluation practice that proceeds democratically, facing up to the political, value-imbued situation that evaluators find themselves in. Notice that in the Denver evaluation I addressed the political and value issues raised by those involved.

Of course, participation takes time and resources. Stein says,

A negotiated process of accountability is time-consuming, expensive, and difficult, but only this process has a chance of providing the quality of public goods that governments and citizens repeatedly say they want, and that providers say they want to deliver. (Stein, 2001, p. 185)

In my opinion, not every study requires such an extended approach: sometimes the conflict that such an approach addresses does not exist. Even so, the Denver study cost less than \$70,000 (U.S.) a year, not terribly expensive.

Deliberative democratic evaluation is only one of several approaches to more democratic, participatory, and collaborative evaluations. Other ideas have been advanced by Cousins and Whitmore (1998) in Canada, who have focused on in-depth participation; MacDonald in Britain (MacDonald & Kushner, in press); Karlsson in Sweden (2003); and Fetterman (2001), Greene (2003), King (1998), and Ryan (Ryan & DeStefano, 2000) in the U.S.

CONCLUSION

In a recent issue of the *Canadian Journal of Program Evaluation*, 12 contributors cited three major problems in Canadian evaluation: lack of identity as evaluators, lack of independence in conducting evaluations, and the dominance of evaluation by performance measurement and program monitoring (Gauthier et al., 2004). In my opinion, the unique contribution of evaluators is their ability to arrive at evaluative conclusions in a disciplined manner. Producing disciplined evaluative conclusions is the defining feature. There are many ways to do this, but to do it properly, evaluators must maintain some independence from managers and others, even when they work closely with stakeholders.

John Mayne has suggested that results-oriented public administration should focus on results that matter to citizens. I interpret that

to mean the results that really do matter, not necessarily the results managers think matter or think should matter. Mayne says, "Measurement in the public sector is less about precision and more about increasing understanding and knowledge" (Mayne, 2001, p. 6). I agree, and I believe that evaluation done properly is compatible with monitoring. After all, the Denver evaluation was a case of monitoring. Of course, if the evaluators' hands are severely tied as to what they can look at, whom they may involve, and what methods they must use, that is another matter. That moves toward government control, which I do not see as being compatible with evaluation or democracy.

Finally, to return to Janice Gross Stein, what was wrong with using performance indicators the way they were used? The problem was that it did not take account of the needs and interests of all the parties involved, particularly the patients and their families. Certainly, it is reasonable to be concerned about costs, but there must be a way of including the views, values, and interests of those affected without sacrificing cost concerns. Evaluation requires balancing different values and interests, not an exclusive focus on one or the other.

Again, Stein says this better than I do:

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)

ACKNOWLEDGEMENTS

Thanks to Alan G. Ryan for helpful comments on earlier drafts. This paper was presented to the Canadian Evaluation Society, Saskatoon, May 19, 2004.

REFERENCES

- Cousins, J.B., & Whitmore, E. (1998). Framing participatory evaluation. In E. Whitmore (Ed.), *Understanding and practicing participatory evaluation* (pp. 5–23). San Francisco: Jossey-Bass.
- Fetterman, D. (2001). *Foundations of empowerment evaluation*. Thousand Oaks, CA: Sage.
- Gauthier, B., Barrington, G., Bozzo, S.L., Chaytor, K., Cullen, J., Lahey, R., Malatest, R., Mason, G., Mayne, J., Myers, A., Porteous, N.L., & Roy, S. (2004). The lay of the land: Evaluation practice in Canada today. *Canadian Journal of Program Evaluation*, 19(1), 143–178.
- Glanz, J. (2004, February 19). Scientists say administration distorts facts. *New York Times*.
- Greene, J. (2003). War and peace ... and evaluation. In O. Karlsson (Ed.), *Studies in Educational Policy and Educational Philosophy, 2*. Sweden: Uppsala University. <<http://www.upi.artisan.se/docs/Doc186.pdf>>.
- 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. <<http://www.upi.artisan.se/docs/Doc187.pdf>>.
- 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. <<http://www.upi.artisan.se/docs/Doc171.pdf>>.
- King, J.A. (1998). Making sense of participatory evaluation. In E. Whitmore (Ed.), *Understanding and practicing participatory evaluation* (pp. 57–67). San Francisco: Jossey-Bass.
- MacDonald, B., & Kushner, S. (in press). Democratic evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation*. Thousand Oaks, CA: Sage.

- Maxwell, J.A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher*, 33(2), 3–11.
- Mayne, J. (2001). Addressing attribution through contribution analysis: Using performance measures sensibly. *Canadian Journal of Program Evaluation*, 16(1), 1–24.
- Ryan, K.A., & DeStefano, L. (Eds.). (2003). *Evaluation as a democratic process: Promoting inclusion, dialogue, and deliberation*. New Directions for Evaluation, 85. San Francisco: Jossey-Bass.
- Scriven, M. (in press). Causation. In S. Mathison (Ed.), *Encyclopedia of evaluation*. Thousand Oaks, CA: Sage.
- Stein, J.G. (2001). *The cult of efficiency*. Toronto: Anansi.

Ernest R. House is professor emeritus at the University of Colorado, a visiting professor at Royal Melbourne Institute of Technology in winter, and a long-time labourer in the evaluation field. Recent books include *Values in Evaluation and Social Research* (with Ken Howe) and an evaluation novel, *Where the Truth Lies*, the unexpurgated adult version of which can be found at <<http://house.ed.asu.edu>>.