

JMDE

Journal of MultiDisciplinary Evaluation

Number 1, October 2004

Editors

E. Jane Davidson & Michael Scriven

Associate Editors

Chris L. S. Coryn & Daniela C. Schröter

Assistant Editors

Thomaz Chianca

P. Cristian Gugiu

Paul A. Lamphear

Mary Keating

Nadini Persaud

John S. Risley

Regina Switalski Schinker

Lori Wingate

Brandon Youker

Webmaster

Dale Farland

Mission

—The news and thinking
of the profession and discipline of evaluation
in the world, for the world—

A peer-reviewed journal published in association with

The Interdisciplinary Doctoral Program in Evaluation
The Evaluation Center, Western Michigan University

Editorial Board

Katrina Bledsoe
Robert Brinkerhoff
Tina Christie
J. Bradley Cousins
Lois-Ellen Datta
Stewart Donaldson
Gene Glass
Richard Hake
John Hattie
Ana Carolina Letichevsky
Mel Mark
Michael Quinn Patton
Nick Smith
Robert Stake
James Stronge
Dan Stufflebeam
Helen Timperley
Bob Williams

Introduction

Welcome to the first issue (October, 2004) of the Journal of MultiDisciplinary Evaluation! As we 'go to press' there are 629 people signed up for notification of its appearance, from about 50 countries. Please pass the internet address along to your friends and colleagues, and tell them that all issues will continue to be available by a single click directly from our home page.

This issue is close to 150 pages, but it's split into three parts for easier downloading. And it's designed to facilitate selective reading: find your way around by looking at the Table of Contents, below, and clicking on a section or subsection title to go directly there. Be sure to check out the Essay Competition, which is buried in a short piece called "Zen and the Art of Everyday Evaluation"—and consider entering an essay (it only needs to be 500 words or so). Also think about us for an article (or a letter or a memo)—see the Mission Statement for details on submissions. And get a sense of what's happening in evaluation around the world through the 90 pages of our Global Review—of regions (Part II) and of journals (Part III). Can you enrich this with more about evaluation in your part of the world or your publication? Join our emerging group of onsite correspondents by bringing us all up to date—follow the model of our coverage of Latin America. In later issues we'll add coverage of new regions and publications, and update the coverage here with supplements, to provide a historical record of the global development of evaluation. (Next issue; April/May 2005, or before.) If a later amendment is made, we'll put a link on the earlier article taking you directly to it.

In the next issue, we'll have: (i) some serious coverage of the arguments about methods of demonstrating causation in evaluation; (ii) discussion of valid and invalid efforts at controlling cultural bias in evaluation; (iii) the beginnings of an item pool for testing competence and proficiency in evaluation. And more!

Table of Contents

Part I

Mission for the Journal of MultiDisciplinary Evaluation.....	1
Editorial: The Fiefdom Problem, <i>Scriven, M</i>	11
Unpacking the Participatory Process, <i>Weaver, L. & Cousins, J. B</i>	19
Zen and the Art of Everyday Evaluation, <i>Scriven, M</i>	41

Mission for the Journal of MultiDisciplinary Evaluation

Michael Scriven

A. Why a new journal?

1. We have excellent journals in evaluation, and it would be hard to argue for simply adding one more of their kind to their numbers. But if professional evaluation is going to help improve the world, as many of us strongly believe it can, it must take seriously the task of communicating current developments and skills to the evaluators, evaluation users, and would-be evaluators amongst those people in the world who can't afford to subscribe to the traditional journals or attend the traditional workshops and courses of study. Those people include impecunious students in the industrialized nations, as well as impecunious teachers and community members there, and most people in the primarily rural/agricultural nations. So this journal is different in that it's free. It won't reach everyone who could use it, because not everyone can get to and use a computer terminal with online capability, and read English, but it will be available to several million people that can, and that number is increasing fast.
2. As some of you know, the great war between the commercial publishers that control most of the scholarly journals, and the great libraries that have been making those publishers rich via the massive increases in library subscriptions has at last resulted in a battle won for scholarship. After an abortive effort at negotiation by, amongst others, the State University of New York libraries, the University of California recently simply refused to

pay the latest increase, and the publishers backed down, cutting about \$1 million dollars (U.S.) off the annual bill. Harvard and Cornell are simply canceling 300 journal subscriptions between them; the Research Triangle Libraries (Duke, UNC, NCSU) are doing the same. It's hard to say how that war will turn out, but scholarly interests are obviously served by facilitating the option of online publication, and the Senates at Cal, Stanford, SUNY, and Connecticut have moved to encourage scholars to use, and create, less commercial publishing outlets including online ones. As a leading advocate of online publishing recently put the situation on that front, "there are currently a thousand peer reviewed journals that appear only online. Among the "open access" ones (free to read) are the British Medical Journal, BioMed Central (a collection of 50 open access journals), Educational Researcher, First Monday, and College English." (Of course, there are many other non-profit ones charging a small subscription to cover expenses.) We aim to develop some experience in the online approach, which we will make available freely to any other evaluation journals that feel they need to facilitate less expensive access to their contents. It's worth noting that Gene Glass' ground-breaking free access journal, the Educational Policy Analysis Archives, has more readers downloading articles than there are readers for all the main paper-based educational research journals put together.

3. There are many other niches in the journal world that need to be filled besides radically reducing the cost of access, given that we start with the belief that the existing evaluation journals are extremely good, and that direct competition with them would be counter-productive. One of these niches, in our opinion, is the need to move towards some coverage of

significant evaluation happenings in countries outside North America. We will gradually develop this, as we extend our network of correspondents overseas and from overseas, and we will try to provide some periodic overviews of major meetings, movements, and publications that occur in languages other than English. As we develop increasing numbers of readers in regions such as South America, we will move towards publishing articles and some summaries, in (for that case) Portuguese and/or Spanish. Sign up with your e-mail address in the space provided on our site so as to register interest from your area, and rest assured that your address will not be released to anyone else. (If you are using a school or library or internet café computer, and don't have an e-mail address, send us an e-mail from it to tell us where you are.) And if you attend an interesting meeting outside the Anglophone area, or for that matter inside it, or read something that you think is important and that you think will not be covered, send in your report. Send in a couple of these from Ulaanbaatar and you are likely to be approached with an offer of correspondent status for Mongolia!

4. Another niche. We want to publish good ideas, and we don't care whether they are embedded in a typical journal article, although those are the vehicles that get the peer review treatment. If you can express your idea in a clearly written paragraph or two, or in a memo, or in a letter, and it looks to the editors like something worthwhile, we'll publish it. Your thoughts might be reactions to your own experiences, to the experiences of others, or to previously published material, which could include a well-known book or article, not necessarily one reviewed here. No, that last e-mail you sent to EVALTALK probably isn't going to qualify. But it might dress up well, with some serious further thought—and with some attention to reactions

from others on ETALK—if it's not too esoteric. Remember, our readership won't consist of PhDs in philosophy or psychology!

5. And another niche. We'll review some books, sometimes books that have been out for quite a while but that have been gradually gathering importance or a following. But we often won't review them in the usual way: we might use two or three reviewers, who might include an ally, a critic, and a bystander. That's often more interesting and useful to the reader than a single review. And we'll also encourage the authors to reply to the reviews, in the same or the next issue. Later, the reviewers can reply to the author's comments. In other words, we want the serious discussion of major emerging movements or themes in evaluation to be strongly supported in this journal. In the same spirit, we'll hope to get submissions of dialectic pieces—double articles, with one responding to the other.
6. And.... Authors can add postscripts to their articles, a year after they are published.... Or several years later. They can't alter the original text, and the postscript will be date-stamped, but it can set the record straight when they want to do this, or strengthen the arguments if they want to do this. All articles will be archived and available to the searcher in the usual way.
7. Moreover.... This isn't just a research journal. It's a journal aimed at communicating about evaluation to a very diverse readership. That may mean that it should be partly instructional, too. The model of hybrid journal/magazine publications such as *Scientific American* is worth taking seriously. Along with new research results, they often publish overviews of material that the expert knows well, but the outsider or student in that particular field knows little about. In that spirit, too, we'll do some

reportage on what other journals are covering, for those who can get them through a library. Another common feature of publications like *Scientific American* is an inquiries column where an expert responds to questions from the field. To the extent that our resources permit, we'll explore the inclusion of that kind of material. And that means you can submit that kind of material. Instructors might submit what seems to them a neater treatment of logic models than is found in the standard texts; or their responses to the most common misconceptions about evaluation from students in their mid-career extension course for ward nurses, and someone else may respond to their articles. Could we have an Ethics column? Perhaps, if good questions and good answerers can be found.

8. Furthermore.... In the 0th issue of JMDE, which was to be just an introductory flourish to show we're here and working, there's a not-too-serious piece called 'Zen and the Art of Everyday Evaluation'. Zen masters are famous for their use of puzzles, known as koans, which illustrate some deep point in Zen thought. There's an evaluation koan in this article, and it's the first of what we hope will be a series of problems or puzzles that we'll publish from time to time. And of course there will be some prizes for the best answers, usually an interesting book. If you come across or think up an interesting puzzle about evaluation, send it in! We will probably dig up a prize for the year's best entry. This article and many handwritten pages were on a clipboard stolen from Michael Scriven in Canada this summer. It has been replaced in Issue 1.
9. Besides which.... What else could we do that would be interesting and useful? We welcome your suggestions. (10)You're already thinking about the use of photos? Right, so have we, though the technical problems are not

trivial for the software we're using. (11) You thought of color, too, perhaps for concept maps and logic diagrams? You can bet we'll be working on that, it's a potentially substantial advantage of the online medium. (12) How about cartoons? Send them in; become the first famous evaluation cartoonist! (13) What about material from the dozen other fields of evaluation that have attained professional status, such as policy studies and personnel evaluation and product evaluation? That's one of the reasons for the title; we want to encourage border crossing, and there's perhaps room for more of it than finds its way into the existing journals. (14) And how about exploiting the greatest strength of online publication: the response speed? We will put out special issues when it seems urgent to do so: for example, it might have been helpful to do one on the 'Causal Wars' that split the evaluation community last year, with of course both sides well represented. This is not a vehicle for a partisan approach to evaluation: to the extent that we can provide diversity and civility, which will be our aim.

10. We have some other ideas, but perhaps 14 suggestions will be enough to indicate that *JMDE* ("Jim Dee") has a place on the team bench. With your help, we can fill that place and expand it too.

B. Why this title?

We considered many titles. Googling them revealed that almost all had been taken or virtually taken. But we rather like this one, because it suggests something that's important to us, the notion that the essence of evaluation, not just historically but in practice today, is its multiple lineage. We'll try to illustrate that in the pages we publish, and hope that authors will be attracted by it. And there's nothing esoteric about the title: the phrase "multidisciplinary evaluation" generated 318,000 hits on

Google recently, so the term is one in common use, notably in the medical and psychiatric fields where it refers to the efforts at diagnosis that require specialists from very different fields to collaborate. In program evaluation, this most obviously connotes the collaboration between the subject matter expert and the evaluation expert. But that's just an epidermal analysis. The fact is that there's often a need for an expert cost-analyst, an expert focus group or survey specialist, an expert on text analysis or case study, maybe an attorney or an organizational development or a community development specialist, or an expert on another culture or from a distinctive community. Many of us become pretty good at several of these specialties, but the big shops often have them on staff or standby.

Moreover, there is often a multiple disciplinary interaction at the subject-matter level, not just in applied psychology and medicine; for example, an authority on eLearning prefaced an online discussion a couple of weeks ago by saying "e-Learning involves multiple disciplines e.g., philosophy, psychology, pedagogy, anthropology, artificial intelligence (e.g., Artificial Intelligence in Education (AIED)), and human computer interaction." Evaluation of e-learning courses or programs, and many other kinds of evaluand, is often, perhaps typically, like this; and it may be good to pay more attention to this feature of it than we have done in the past. Hence the title. (And why *JMDE*, not *JME*? Out of respect for the *Journal of Moral Education* and the *Journal of Management Education*!)

C. Who is producing it?

The co-editors will be Jane Davidson from New Zealand and Michael Scriven from Michigan, aided by a distinguished and diverse international Advisory and Review Board to which we will continue to add people for some time, as the new network develops. Assistant editors will be a group of the doctoral students at the

Evaluation Center at Western Michigan, headed by associate editors Chris Coryn and Daniela Schröter. We aim to make this the equivalent of the Law Review experience for them. The list of correspondents, like the Advisory Board, will be posted on our website as it develops. Western Michigan is kindly helping with the website, courtesy of Arlen Gullickson, Director of the Evaluation Center, and Dale Farland, our Webmeister. The initial website is evaluation.wmich.edu/jmde, though we're applying for jmde.com.

Special thanks, too, to the Canadian Government, for funding the development and free distribution of the software we are using, designed precisely for the management of online, free access, journals; and to Professor Willinsky, of the University of British Columbia, the expert on electronic publication quoted earlier, who has helped us with access to that software. And thanks to Gene Glass, the founder and editor of the highly successful EPAA, his online refereed journal that invented a number of the ingenious procedures we'll be using; we're especially glad to have him on our Advisory Board.

D. How Can Others Help With It?

(i) Please help to spread the word that a new journal is available, with a broad vision and interests. And, (ii) since its value will depend on what it publishes, make sure to keep *JMDE* in mind for things you'd like to have published. We will make that as easy to do as we can, including eventually an effort to publish material in your native language. Remember that you should be able to reach a whole new audience through us, a very important part of the world's population. And remember that online refereed journals are now widely endorsed as respectable entries in your cv. (iii) If you have special interests or skills that you'd like to be sure are represented in *JMDE*, sent us a note and a sample or

two of your work. (iv) Everyone, please think about other things we can do that aren't already well done; and (v) suggest the most interesting puzzles about evaluation you have or you encounter—they can form the basis for a cutting edge discussion here. Other ways to help are mentioned throughout the earlier sections.

Practical postscripts: (a) In the interests of quality peer-reviewing, articles submitted to *JMDE* should be written without detectable authorship in the manuscript itself, only in the covering letter—which won't go out to the referees. If you can, please use Microsoft Word with 1" margins all round, 1.5 line spacing, and Times 14 point font; e-mail if possible. We don't insist on APA style or any other; just intelligibility and consistency. Please don't submit an article that is under consideration elsewhere, it wastes referee and editorial time. In return, we'll get you a decision very quickly, within three weeks from receipt.

(b) The *JMDE* effort is a kind of safety-net counterpart—in the field of publishing brief scholarly materials—to the AEA Monograph Series. The latter provides direct cost-competition to the publishers of hardcopy books, by publishing books at \$15. That market is one in which one can't compete without some cash flow to cover author's time and printing costs, so free online access is not feasible, and paid online access is still not secure. The big commercial publishers in both domains—books and journals—are substantially similar, led by Elsevier and Kluwer, so the aim is to shake their increasingly life-threatening grip on the distribution of scholarly knowledge, at least in the field of evaluation.

(c) When writing to us, to ensure attention, add “*JMDE*” to whatever else you put in the subject line. These virus-ridden days, no one should open attachments that cannot be identified prior to opening.

Editorial: The Fiefdom Problem

Michael Scriven

NOTE: Editorials in JMDE represent the personal views of the editor who signs them, not of the journal's editors or staff as a group. They are somewhat uncommon in scholarly journals, but JMDE is a somewhat uncommon journal. Correspondingly, you will not be surprised to hear that they are published with the thought of stimulating a discussion, or at least reactions, so please send in your considered reflections on them!

The emergence of dominant countries in world politics is marked by a history of the amalgamation of fiefdoms—mini-empires usually ruled despotically by a baron, prince, king, or maharajah. Usually the fiefdoms were too small to defend against some of their neighbors, and they were often too small for major economies of scale in production. Hence they formed alliances through marriage, trade, or mere covenants. Of course, these are fragile links, compared to complete unification, so the path to better defense, industrialization, and further expansion—as well as riches for the conqueror—lay along the latter path, which often was unilateral and of course it also resulted in an entity powerful enough to invade or dominate still larger but reluctant fiefdoms and eventually countries. The great empires, from West to East, developed in this way, and it is often said that this is the way that the present leadership in the USA is trying to go, under the smokescreen of (selectively applied) slogans such as democratization, the Monroe Doctrine, 'death to tyrants,' or 'protection of vital interests.' Whether or not that rather cynical view is correct is not the issue. The evaluation of that policy is closer to our business, and it's clear that its merits are now considerably compromised by

two new considerations: (i) the proliferation of extremely powerful, portable, and cheap weapons; and (ii) the exemplar of successful guerilla resistance to mighty armed forces. It thus seems possible that the best view of the present situation is that the way the US won the Cold War (or the USSR lost it) may be the only cost-feasible path for world leadership, as the violent alternatives simply continue to falter or fail. It might be called, “takeover by exemplifying a better way”.

These thoughts about fiefdoms and their fate are occasioned by two recent events, and one persistent problem in the evaluation world. The first of the recent events is the Causal Wars that began last year, which remind us that the world of ideas is not immune to the bare-faced use of political power, misrepresentation, and ad hominem argumentation in the struggle for ideological and economic control. The other is a request to all presenters at a major series of educational workshops and seminars this past summer—not the Evaluators' Institute, by the way—that they should adhere to the definitions and structuring of evaluation provided in some online resources provided by the sponsors. This seems harmless enough—and was, I am sure, merely an effort to avoid confusion amongst the attendees—until one studies these definitions and structure. Then one discovers something that, one recalls unhappily, has now become too frequent an occurrence: a multiple and major failure to grasp the essential elements of many of the basic concepts of our field. The definitions provided for terms shared with statistics, social science methodology, or common English are quite adequate: but definitions of terms unique to evaluation reflect a severe lack of clarity about these concepts. And now one recalls that there are other foundations, organizations, and educational institutions that are prominent in the evaluation business, and deserve much credit for their support and work in that field, where the same tendency to standardize on confused interpretations of these concepts has become part of the—conscious or

unconscious—efforts at ‘branding’, that is, the effort to leave a distinctive mark on some part of the field that will demonstrate one’s own contribution.

The result of each fiefdom standardizing on their own (significantly different) usage is of course just the kind of confusion at the macro level that the standardizers are trying to avoid in their own bailiwick: a person learning or using one set of definitions will have trouble understanding and communicating with those trained to another version. We've already seen this happening quite often on Evaltalk. If combined with the kind of economic and political enforcement that has occurred in the Causal Wars takeover of most of the federal funding for educational research, where some \$500 million per annum is now (de facto) reserved for those with the 'right views' on the highly controversial issue of establishing causation, we will seriously undercut the possibility of progress towards an understanding of the nature of our field, and of our discoveries in it, whether it's conceived as a discipline, a profession, or a set of practices. In other words, the political cycle from fiefdom to empire is playing out again in our domain, and we should be concerned that evaluation funding restrictions, for philanthropies, will follow the federal precedent in being totally restricted to those willing to share particular variants of standard conceptual frameworks that lack adequate justification for the variation.

This is a good moment to remind ourselves of the classic disaster of this type, the stupid blunders of the statisticians who casually redefined perfectly good words in the English language in such a way as to confuse millions of students and citizens for most of a century. To redefine ‘reliability’ so as to exclude its common meaning which includes validity, instead of using ‘consistency,’ was the first of a series of analogous mistakes, where ‘significance’ was next to suffer, and then

‘explanation’ as abused by factor analysts¹. The current attempt to redefine ‘evidence-based practice’ in medicine, public health, social services, education, etc., is at least one where more sophisticated arguments are being used.

Back to the fiefdom problem. The third trigger for this concern with the Balkanization of evaluation—that is, unnecessary fragmentation, confusion, and attendant hostility, with the shadow of dictatorship in the background—is of much greater importance to the world at large. In the field of international development, it has become increasingly clear that the situation with the evaluation of interventions is far from satisfactory. This areas has long been one of concern to thoughtful evaluators, because of the combination of limited external oversight with the usual strong (though tragically short-sighted) double-barreled motivation for doing superficial or zero evaluation—namely, that serious evaluation *might* make you look bad, and it uses valuable resources. This appeal to both risk-management and fiscal conservatism is always hard to beat². More detailed analysis, especially by Paul Clements, one of the faculty for our doctoral program in evaluation here, makes clear by on the ground meta-evaluation studies in Africa of the World Bank, CARE International and USAID program evaluations, that these concerns are all too appropriate³. Each maintains a fiefdom of its own

¹ Perhaps the ultimate example was to elevate a contradiction to the level of a recommended procedure, viz., ‘grading on the curve.’

² What beats it here, and in most other situations, are these four cards from the same pack: (i) evaluation is the best way to uncover early warnings that allow you to avoid the worst possible outcomes; (ii) it is a cost-effective way to get that insurance; (iii) it will sometimes produce highly positive conclusions which reward good work and can generate further rewards, including further opportunities for good work; (iv) it will often benefit you far more than it costs, in ways other than those referred to in (i) through (iii), e.g., by showing the way to plain quality improvements. Evaluations that do these things need to be designed to do them, and many are not; so these are not blanket endorsements of evaluation, but only of good evaluation practice.

³ For further information, Google will locate sites that have his doctoral dissertation (from the Woodrow Wilson School at Princeton): search on the first five words of “Development As If

operations, including their evaluations, which has its own rules and indeed culture. Despite some improvements, and—please note—*some* very good evaluations, gross errors persist. The editors hope, and intend, that this journal will provide one source of encouragement for improvement in this area, and hope to include an article by Dr. Clements in the next issue, as well as comments from country evaluators where the big development agencies operate.

Related to this example is the recurrent tendency for agencies to issue RFPs for ‘external evaluations,’ in which they overspecify the design all the way down to overspecifying the requirements for bidders⁴. Doing this of course undercuts externality to the point where it loses most of its contribution to credibility and seriously attacks validity. A tempting way to extend the fiefdom, of course, and nearly as bad as sole-sourcing the contract to a friendly consultant. In other words, how to make an external evaluation into an internal one.

What else can be done to avoid both the linguistic confusion and the Balkanization of research—and the funding of research—on evaluation? We might be able to learn something from what happens in philosophy, the field where nothing is taken for granted, all concepts are up for reformulation, and very different interpretations of the key ones are taught at different colleges, depending on which school of thought is dominant amongst the resident faculty. Doesn’t this just show that one can’t hope to prevent multiple interpretations of key concepts? I believe the main lesson to be learnt is more fundamental: one must treat the definitions of key existing concepts as an extremely serious matter, not a matter of casual linguistic

Impact Matters: a comparative organizational analysis of USAID, The World Bank, and CARE based on case studies of projects in Africa”

⁴ A particularly egregious example is the latest UNICEF RFP to evaluate their “programme of cooperation” to strengthen humanitarian responses. Here the requirements on the bidders amount to requiring membership in the inner circle of ‘acceptable’ players, i.e., the fiefdom country club.

convenience (which is true only with neologisms). Conceptual schemes, and the definitions that go with them, are powerful instruments of analysis and hence persuasive support for particular interpretations, not minor precursors to it (a point well made in *Zen and the Art of Motorcycle Maintenance*, by the way).

Constructively speaking, I will also take two steps myself: first, I will propose to a few leading organizations engaged in teaching, supporting, and propagating evaluation, that we need to hold a small conference of interested parties on a double topic, which we might call “Finding Common Ground”. The agenda would cover: (i) *standardizing terminology* where possible, the reasons for doing this, *and* the limits of such attempts; and (ii) finding *compromise positions* on major conceptual issues, such as the one about causation. This is a natural marriage of goals, since the difference between common definitions and common analyses is only a gradual one.

Second, I will take care, in the doctoral program that I run, to stress the existence of, the case for, and the need to tolerate, alternative conceptual schemes and definitions besides the ones for which I argue—although not to treat this as a matter for arbitrary decision, but rather as something that requires serious justification. That’s a tough distinction to make. I hope others will join in this conscious effort, or write to *JMDE* explaining why they think this is an undesirable strategy—or one in need of major extensions.

ENDNOTES 1. The most important potential relevance of this editorial is to the problem of evaluation in Europe today; and probably in Africa tomorrow. We'll try to carry some news about the conflict between the urge to brand, a.k.a nationalism, and the urge to communicate.

2. No good evaluator would read the above without noting that it can also be seen as an attempt by someone who invented a fair number of the terms in the evaluation vocabulary to extend his own fiefdom. While I do think that people who invent terms have some obligation to argue against careless shifts from their original meanings, they also have an obligation to be open-minded about *serious arguments* for modification or clarification of the original definitions. I make an effort in the *Evaluation Thesaurus* not to 'brand' the dozen or so terms I have introduced, like meta-evaluation, impactee, and the formative/summative distinction, with any claim to authorship, hoping thereby to free others to suggest modifications to the definitions. And I'm now inclined to think that the arguments, notably by Michael Quinn Patton and Eleanor Chelimsky, for adding a third category to formative and summative have merit, although I originally took those two types to be exhaustive. In an essay in Alkin's *Evaluation Roots* (Sage, 2004) I suggest one might use "ascriptive" to identify certain evaluations—for example, an evaluation done by a military historian of Napoleon's use of cavalry—that are aimed at neither improvement of an evaluand, nor macro-decisions about it⁵, but

⁵ Macro-decisions include decisions to continue or cease funding, purchase, or export. The contrast is with micro-decisions, e.g., whether to improve the programming of the organization's database. The latter but not the former are serviced by formative evaluation.

simply at determining/ascribing merit, worth, or significance ‘for its own sake’.⁶
There, I’m not incorrigible; how about you?”

Example: here’s one of the World Bank’s definitions:

Meta-evaluation—The term is used for evaluations designed to aggregate findings from a series of evaluations. It can also be used to denote the evaluation of an evaluation to judge its quality and/or assess the performance of the evaluators.
Meta évaluation Évaluation conçue comme une synthèse des constatations tirées de plusieurs évaluations. Le terme est également utilisé pour désigner l’évaluation d’une évaluation en vue de juger de sa qualité et/ou d’appré
Metaevaluación Este término se utiliza para evaluaciones cuyo objeto es sintetizar constataciones de un conjunto de evaluaciones. También puede utilizarse para indicar la evaluación de otra evaluación a fin de juzgar su calidad

Comments by MS. The definition treated as primary—the one in the first sentence—is a simple confusion of meta-evaluation with meta-analysis. The second definition is correct and of course quite different. Arguably, the former will not result in an evaluative conclusion, but in an analytic conclusion of the following (non-evaluative) kind: “The evaluations studied lead to the conclusion that on balance, the new meningitis vaccine is not unduly risky for those with compromised immune systems.” A meta-evaluation always leads to an evaluative conclusion, of the form “This evaluation is sound/unsound/clear/unclear/credible/not credible.”

⁶ The example is exotic, because most examples of ascriptive evaluation can be used for formative or summative purposes by users in contexts other than that of the originator, which tends to blur the distinction.

Unpacking the Participatory Process

Lynda Weaver & J. Bradley Cousins⁷

University of Ottawa

Introduction

Interest in collaborative forms of inquiry has increased dramatically in recent years in evaluation and social science research. One consequence of such interest has been the emergence of many different forms or genres of collaborative inquiry, such as stakeholder-based evaluation, deliberative democratic evaluation, practical participatory evaluation, transformative participatory evaluation, empowerment evaluation, and the like. In order to ensure clarity of purpose and application, it is necessary to differentiate among such approaches. One such framework—originally proposed by Cousins, Donohue and Bloom (1996) and later developed by Cousins and Whitmore (1998)—applies not only to collaborative and participatory forms of evaluation but to forms of applied social research in a broader sense. Within the framework consideration is given to both the goals and interests of collaborative inquiry (i.e., pragmatic, political, epistemological) as well as to dimensions of process (i.e., control of technical decision making, stakeholder selection, depth of participation).

⁷ Corresponding author: B. Cousins, Faculty of Education, University of Ottawa, 145 Jean Jacques Lussier, Ottawa, ON, CANADA, e-mail: bcousins@uottawa.ca. **Authors' Note:** The authors are grateful to S. Khana, J. Koudys, B. Li., and A. Tsarouhaus for their valued input. A previous version of this paper was presented at the annual meeting of the American Evaluation Association, St. Louis, November 2001

This paper questions the adequacy of the process dimensions of the earlier version or our framework. Our ongoing analysis of process dimensions reveals that one of the dimensions—stakeholder selection—is problematic and requires reconsideration. In this paper we re-present the framework and describe enhancements to the process dimension component. By way of illustration, we then apply the framework to two separate case examples of practical participatory evaluation. This work is relevant to the study and practice of evaluation because it helps clarify differences among versions of collaborative inquiry and thereby helps reduce confusion that may arise in discussions about, or applications of, such approaches. The enhanced process component of the framework allows interested parties to graphically depict the continua for a given inquiry project in order to portray differences in collaborative evaluation approaches. It also provides the basis for the development of research tools that could be used for empirical inquiry into participatory processes in social inquiry and their effects.

Goals and Interests of Collaborative Inquiry

We identified three primary goals and interests associated with collaborative social inquiry, derived in the first instance, from Levin (1993), but found them to resonate with other conceptions such as Mark and Shotland (1985) and Garaway (1995). Any given collaborative research project, we suggest, would be characterized by a primary emphasis on one or some combination of the three goals and interests. First is the *pragmatic* justification. Collaborative inquiry is purported to lead to instrumental consequences and to increase the usefulness of the knowledge that is created. In this sense, collaborative inquiry takes on a problem-solving orientation. Members of the community of practice engage with researchers or evaluators to produce knowledge that bears upon identifiable practical problems. To the extent that the research is grounded in the context for

use and thereby rendered meaningful to those responsible for problem solving, decision making or policy making, the knowledge produced will be of greater use.

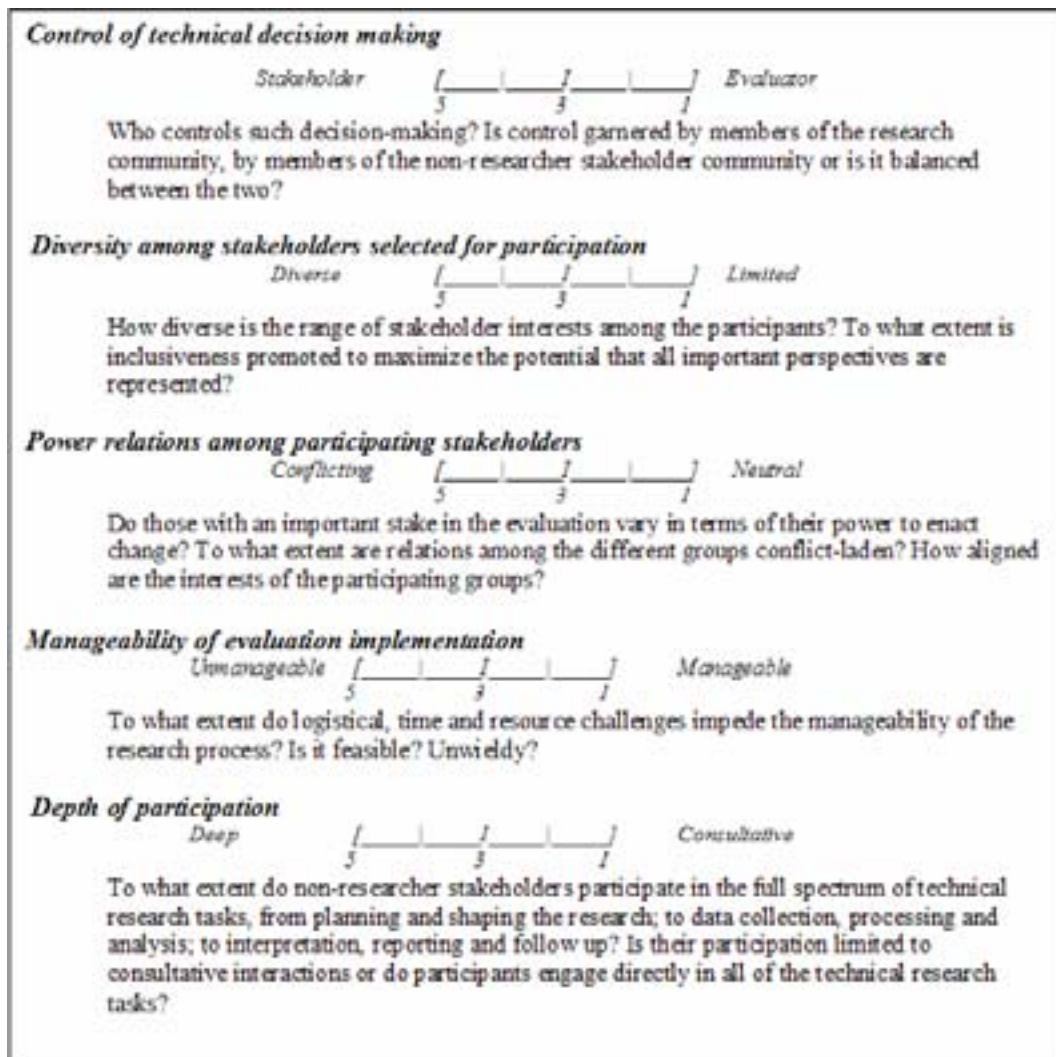
A second justification is *political* and is ideologically rooted in normative conceptions of social justice and the democratic process. The primary interest of collaborative inquiry that subscribes to such political aims is to promote fairness through the involvement of individuals associated with all groups with a stake in the research (e.g., applied study, evaluation) or the focus for research (e.g., programme, policy). Through direct involvement and participation in the research process, persons from oppressed groups or marginalized sectors that do not normally have a voice in policy or programme decision making are now provided with such opportunities. The focus for politically-oriented collaborative inquiry is very much emancipatory or concerned with the amelioration of social inequities inherent in the societal structures of the status quo.

The third and final justification for collaborative inquiry is *epistemological*, the primary aim being the production of valid knowledge or representations of underlying social phenomena. Recent challenges to the dominant paradigm for research in the social sciences—logical empiricism—have been many and varied and stem from fundamental distinctions made in conceptions of reality and of knowledge. In his comprehensive review and integration of constructivist conceptions of research in the social sciences Schwandt (1997) epitomizes the concept of the ‘localness’ of knowledge and the importance of context as the essence of constructivism. While constructivist conceptions of research are undeniably rooted in relativist epistemologies, others have argued from different footing and similarly placed a premium on context. Huberman (1994), for example, proposes a perspective regarding knowledge production, utilization and dissemination that might be termed ‘revisionist-traditionalist.’ He argues that

knowledge can indeed be transported from one context or setting to another but that its reception, interpretation and integration into the local context determines its impact and sustainability. His construct ‘sustained interactivity’ suggests that reciprocal effects on knowledge user and producer communities will arise from enhanced contacts between the two. The argument is aligned with a justification for collaborative inquiry that aims to enhance the validity of the produced knowledge.

Process Dimensions of Collaborative Inquiry

Quite apart from considerations of the aims of collaborative inquiry, we identified dimensions of form as being important and suggested them to be fundamental in characterizing various collaborative approaches to systematic inquiry (Cousins & Whitmore, 1998). Each may be thought of as a Likert-type rating scale along which any given application of collaborative inquiry may be described. Initially, we identified three such dimensions—control of technical decision making, stakeholder selection, depth of participation—but through ongoing analysis came to the view that one of these dimensions was confounded and therefore conceptually inadequate. We ultimately teased apart the dimension ‘stakeholder selection’ into three distinct dimensions of form or inquiry process. The resulting framework consists of five dimensions of form.



Taken together, a given collaborative inquiry might be represented diagrammatically in the form of a ‘radargram,’ shown in Figure 1. In the figure we represent hypothetical examples of three distinct forms of collaborative evaluation. We now turn to a discussion of each in terms of its justification and depiction according to our process dimensions.

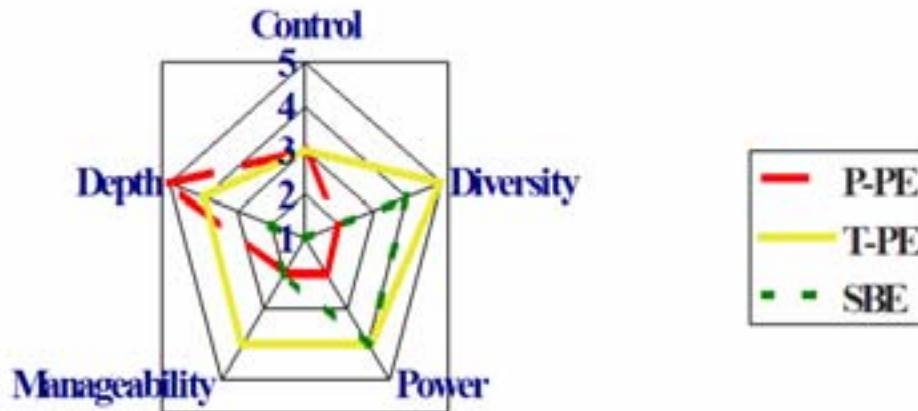


Figure 1: Five dimensions of form in collaborative inquiry

Practical-participatory evaluation (P-PE): Our prior work (Cousins & Whitmore, 1998) differentiated between two streams of participatory evaluation on the basis of the primary aims of the inquiry. The first we called Practical Participatory Evaluation (P-PE) an approach that is very much concerned with practical problem solving and providing support for ongoing programme and/or organizational decision making (see, e.g., Cousins & Earl, 1995). In P-PE, members of the evaluation community work in partnership with members of the programme community to implement evaluations typically seeking to inform programme improvement initiatives. Instrumental (support for discrete decisions) and conceptual (educative function) uses of evaluation findings and process use, are likely to be observed as a benefit of P-PE. Figure 1 shows that technical decision making in P-PE is typically shared between the evaluator and non-evaluator stakeholders. Diversity in participation is likely to be limited as non-evaluator stakeholders are typically primary users, those with vested interest in the programme who are in a position to enact change. Power relations among non-evaluator stakeholders are likely to be neutral since the interests of programme

managers and implementers are usually those most often represented. This, however, is not necessarily the case. Since only a limited number of non-evaluator stakeholders participate in the inquiry, the process would be logistically manageable and feasible. Finally, in P-PE participants are normally involved extensively in a wide variety of the inquiry tasks, including data analysis and reporting.

Transformative Participatory Evaluation (T-PE): Brunner and Guzman (1989) describe an approach to participatory evaluation that has been implemented in evaluations of programmes in developing countries for some considerable time. The approach has decided links with other forms of collaborative inquiry such as participatory action research (PAR) and participatory rural appraisal (PRA) which are normative in intent and seek to ameliorate identified social inequities. Through participation, non-evaluator stakeholders develop their capacity for self-determination and develop rich understandings of the often oppressive forces operating in the local context. This stream of inquiry, which is ideologically grounded and political in intent, we labelled transformative participatory evaluation (T-PE) (Cousins & Whitmore, 1998). In T-PE control of technical decision making is also likely to be balanced between trained evaluators and non-evaluator stakeholders. While evaluators wish to adopt the role of facilitator, there is a need for them to teach participants inquiry methods and the logic of evaluation. Participants would include programme practitioners but in most cases would also involve intended programme beneficiaries as members of the evaluation team. Other interested parties including government officials, NGO personnel, and representatives of donor agencies are equally likely to be involved. Participation, then, would be highly diverse, and given the range of value perspectives having legitimate input a degree of conflict in interests is to be expected. The diverse

nature of participation would naturally lead to logistical challenges and raise into question the feasibility of the inquiry. Finally, as was the case with P-PE, non-evaluator stakeholders would be involved in a wide range of technical inquiry tasks and activities; this being an important element of the capacity building and empowering force of T-PE.

Stakeholder-based evaluation (SBE): Many years ago the concept of stakeholder-based evaluation was introduced through a collection of papers by such renowned contributors as Weiss, Stake and Murray (Bryk, 1983). It was portrayed as being a recommended evaluation strategy when values conflict among stakeholder groups regarding programme purpose or goals was evident. Evaluators would seek to understand evaluation issues from multiple perspectives and the evaluation would be responsive to the exigencies of the local context. In SBE, the evaluator would remain firmly in control of the evaluation and its implementation. Normally a range of stakeholder perspectives would be systematically taken into account and therefore a significant degree of diversity in perspective was to be expected. Best suited to circumstances where programme goals and means are contentious, SBE processes are normally witness to significant differentials in power relations and conflicts of interest. However, with the evaluator firmly in control of the evaluation implementation, the project could be expected to be manageable. Finally, evaluators would most often involve non-evaluator stakeholders in deliberations about the evaluation issues to be addressed and then later, in helping to interpret evaluation findings. Therefore depth of participation would be limited to a consultative role on behalf of non-evaluator stakeholders.

With these three hypothetical examples we can see that the approaches discussed differed considerably in both goals and interests as well as the operational form taken. The framework described above provides a useful means of capturing such

variation among the different collaborative approaches. We now turn from the hypothetical to the actual case in order to demonstrate the utility of the framework in more concrete terms.

Actual Case Applications

The case examples we selected are independent projects on which we worked separately in the capacity of evaluators. The first case (reported by Weaver) is in the domain of hospice/palliative care in the Canadian context: a P-PE of the Volunteer Resources to determine how to improve the programme and to prepare for downsizing of the palliative care unit. The second case (reported by Cousins) is a cross-cultural P-PE of an educational leadership training programme in India. We independently completed the analyses and reports on each case in order to test out the conceptual framework by applying it to actual evaluation cases and to see how the cases might compare, given that they fall within the P-PE stream.

Evaluation of Canadian Hospice/Palliative Care Unit: The sole chronic care hospital in Ottawa, Canada houses a palliative care inpatient unit with 45 beds—the largest unit in Canada. It is staffed with a comprehensive interdisciplinary team, including nurses, doctors, volunteers and many allied health professionals such as a physiotherapist, occupational therapist, chaplain, pharmacist, recreation therapist, psychologist, and Volunteer Coordinator. The team strives to work in harmony to provide specialized symptom management to maximize the quality of life for terminally ill patients and their families. The Director of Patient Care oversees the nurses and allied health professionals, and a Medical Director oversees the physicians.

The part-time Volunteer Coordinator has the responsibility of training and supervising the compliment of volunteers. At the time of the evaluation, there were

approximately 60 volunteers on the roster. Each volunteer comes to the unit weekly for a four-hour shift anytime from 0700h to 2300h any day of the week. Usually, three volunteers are scheduled at the same time to cover the entire unit.

The need to evaluate the volunteer resources arose from the proposed restructuring of the unit 12 to 18 months in the future. As part of the overall preparations to downsize the number of beds and allocated resources, management made plans to obtain feedback from the team members about the future unit. Attention was focused on the volunteer resources because they had not been evaluated formally for many years, and they are an integral, essential part of patient and family care. A commitment was therefore made by senior and middle management to conduct a formative evaluation for two purposes: (1) to evaluate the current volunteer resources and (2) to plan for the restructured, downsized unit. Senior management made the decision to conduct the evaluation. A working committee, of which Weaver was a member, was created and a work plan was drawn up.

The major reason behind choosing to be participatory in this evaluation was to be pragmatic. The working committee could make decisions quickly if the stakeholders were sitting together at the table, and the content of the questionnaires would be exhaustive with all stakeholder groups' input. The political rationale was an important consideration because if management had not included volunteers and nurses in the process, they would not be as likely to accept the recommendations for change to the volunteers' working conditions and policies. Lastly, the philosophy of collaboration in the evaluation reflected the nature of the interdisciplinary and holistic care rendered on the palliative unit. The collaborative evaluation effort would inform management of volunteers' issues, and the volunteers would feel integral to decision making that affects their working

conditions. In summary, while the primary justification for the evaluation was practical, political concerns most certainly factored in.

Having described the evaluation in terms of its background and motivation, we now turn to an analysis of its implementation in operational terms. Weaver rated the inquiry project in terms of its process dimensions using the five dimensions described above. The results appear in Figure 2. These we describe below.

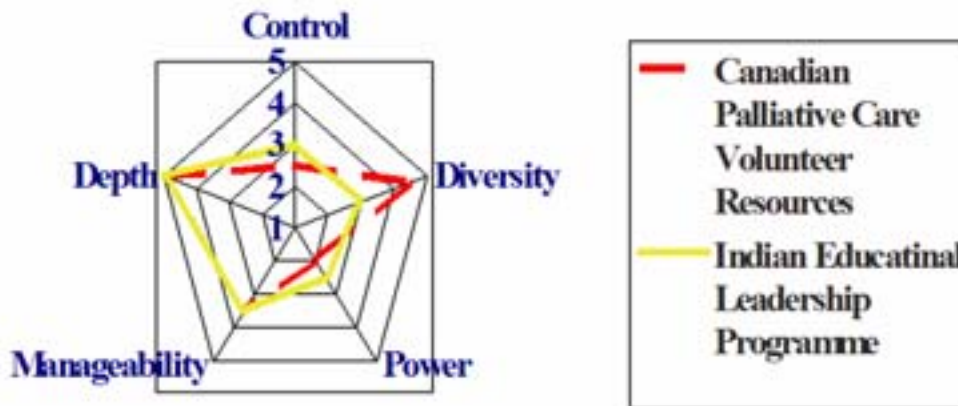


Figure 2: Comparison of Canadian and Indian P-PE cases

1. *Control of technical decision making (2.5)*: Control of technical decisions was shared equally by all committee members. This dimension was actually one that caused strain among group members. At first, questions about technical aspects of the evaluation from the volunteers, Volunteer Manager and nurse were handled quickly by the evaluator and/or the Director of Patient Care (DPC). Resentment was expressed by one of the volunteer committee members. She stated she felt like her purpose was to be a rubber stamp for decisions “already made”. The conflict stemmed from trying to follow evaluation rigour without

enough explanation or without consideration for the non-evaluators' ideas. By consciously realizing the problem associated with this dimension of participatory evaluation, the group overcame the friction.

2. *Diversity among stakeholders selected for participation (4.5)*: Diversity was achieved on the working committee by recruiting representatives from four groups of stakeholders. Management was represented by the DPC, the Palliative Care Volunteer Coordinator (PCVC) and the hospital's Director of Volunteer Resources. Three volunteers were asked to participate, each one with a different length of service on the unit (range from one year to over 10 years). A nurse brought the care team's perspective. Weaver, the evaluation consultant from the Institute of Palliative Care made up the eighth member. Diversity was about as complete as it could have been save for non participation by patients and/or family members whom had not been asked to join the group.
3. *Power relations among participating stakeholders (2)*: The intent of the group was to ensure a balance of power among all committee members. In reality, this balance took time to achieve since it was first necessary to overcome the more customary hierarchy in the work setting where management has power over others. Having three volunteers helped them feel more powerful as a group, then as individuals. The conflict mentioned above concerning 'control of decision making' also skewed the power structure at first. In the end, the group was cohesive and respectful of each other and conflict seemed to dissipate.
4. *Manageability of evaluation implementation (3.5)*: Resources and timing for the evaluation project impacted directly on this dimension. The committee was capped at eight members to balance diversity with functionality. Initially, the data collection was to be limited to a literature search and a mailed volunteer

survey. An outspoken nurse suggested that an evaluation would not be complete without the nurses' opinions since they work so closely with volunteers. A brief survey was, therefore, also administered to the nurses on the unit. Some logistical challenges were experienced, with the amount of data collected in the two surveys being fairly voluminous.

5. *Depth of participation (5)*: Each member of the working committee participated extensively in the evaluation process. As a group they determined the necessary information required to answer the evaluation goals, edited the questionnaires drafted by Weaver, assisted with qualitative data content analysis, and interpreted the findings. As a group, they will put forth recommendations to the Programme Management Committee in terms of how volunteers will function in the restructured unit. The only jobs that were conducted by Weaver alone were the analysis of the quantitative data and the creation of the presentation material. Participation in all aspects of the evaluation was evident.

Evaluation of Indian Educational Leadership Programme: The Educational Leadership Programme (ELP), centred in New Delhi and in existence since 1996, is grounded in an ethos of effective leadership for equity and excellence in education, reflective practice, organizational change and collaboration. Principal foci are the development of personal educational awareness and philosophy, instructional leadership and systemic organizational management. The programme was developed on the basis of mostly American principal training programmes such as Harvard and Danforth.

The impetus for the evaluation came from ELP's creators, developers and implementers, specifically administration and staff of the Centre for Educational Management and Development (CEMD) in New Delhi. The Centre, a non-

governmental organization (NGO) somewhat dependent on external donor funding, has a staff of over 25. The ELP represents an important Centre activity, but one of several. Interest in evaluation stemmed from a desire to understand, through systematic inquiry: (i) the programme's strengths and limitations; (ii) its comparability to other leadership programmes, particularly those in western cultures; and (iii) considerations for ongoing development and improvement. The evaluation was coordinated by Evaluation and Assessment Group at Queen's University (Kingston, Canada) and was contracted by the Aga Khan Foundation, a donor agency providing significant recourses to CEMD.

For the initial formative phase of the evaluation, we adopted a participatory approach with external evaluation team members from Canada working in partnership with CEMD staff, the programme developers and implementers. Cousins was contracted as the evaluation team leader. Advisory input was provided by a variety of interested stakeholder groups including ELP alumni, educational consultants and university professors, and representatives of funding agencies.

On the first of two planned site visits, we developed collaboratively a set of guiding evaluation questions and a programme logic model and then proceeded to systematically examine programme implementation and effects using a mix of quantitative and qualitative methods. Methods employed were an extensive document review of archival information, a questionnaire survey of ELP alumni and a comparison group of non-alumni counterparts, focus groups of alumni and instructional staff, case studies in schools at which ELP alumni were currently located, a cost-effectiveness analysis of financial records, and a comparative analysis of structure and content of the ELP against five other educational leadership programmes, mostly situated in western cultural jurisdictions.

Once planning was complete, data collection, analysis and reporting responsibilities were assigned, with members of the Canadian and Indian teams both contributing. Reports were sent to Cousins electronically by Indian team members and he subsequently developed a complete draft of the report. This draft served as the basis for the second site visit, where a series of meetings over a four-day period were used to develop the draft report, correct inaccuracies, identify and fill omissions and most importantly, to develop a draft set of recommendations for programme improvement.

Following the site visit, Cousins revised the report and presented a list of 25 recommendations for ongoing development of the ELP. Through distance the list was finalized and the report completed and printed and bound. The plan was for CEMD to work with these recommendations for approximately one year, at which time an external team from Canada would conduct a site visit to examine and report on the extent to which recommendation implementation has been achieved. This final summative component would bring a close to the evaluation. Cousins rated this evaluation process according to the dimensions of the framework. As with the prior case, ratings on each dimension appear in Figure 2 and are described in the text to follow.

1. *Control of technical decision making (3)*: Control was shared and balanced. The evaluation began with a site visit and three days of planning. Cousins acted as facilitator in the analysis of stakeholder groups, their interests, and the implications for evaluation issues and questions to be addressed. He also provided input about the participatory model and expectations for shared decision making. Throughout the project, Indian evaluation team members relied on their knowledge of context and the program itself to inform evaluation

decision making. The resulting evaluation was quite sophisticated involving several sources of data, methods of inquiry and bases for comparison.

2. *Diversity among stakeholders selected for participation (3)*: Non-evaluator stakeholders participating directly in the evaluation were predominantly members of the CEMD staff and included the Director. The organization was very collaborative and the Director supportive of her staff. The five or so staff members participating directly on the evaluation had extensive professional backgrounds and skills in program development and implementation. They had prior training in business, education and other applied social science fields. In addition, several members of the leadership programme alumni were occasional participants in evaluation team meetings. They served in an advisory capacity as did a few other individuals, including a university professor and an American who had participated in the development of the programme in the mid 90's.
3. *Power relations among participating stakeholders (2.5)*: Among the Indian team members, occasional differences of opinion surfaced but the process was, for the most part, conflict-free and highly cooperative. Considerable support was provided to the Canadian members of the team. Indian team members felt comfortable in voicing their opinion and challenging proposals for planned action. They routinely questioned assumptions and raised concerns. One such concern had to do with the overarching goal of comparing the ELP with western educational leadership programmes. The Director of the NGO, and original architect of the ELP, remained intent on her resolve that the evaluation would yield such a comparison but not without extended dialogue about the merits of this strategy. Why, for example, could the programme not be considered more directly in terms of its relevance to education in the South Asian context? Another related conflict emerged over a recommendation concerning expected

contact hours for the ELP participants. The exchange was between Canadian and Indian team members, Cousins successfully arguing from the point of view of western standards, as had been agreed by the entire team.

4. *Manageability of evaluation implementation (3.5)*: The process, by and large, was manageable although complications arose as a function of the scope of the project relative to allocated resources and limits on communication due to geographic separation between Canadian and Indian counterparts. Telephone communications were highly impractical. Initial spotty use of e-mail exchanges became more streamlined and useful as the project unfolded. One Indian team member was identified as the project contact person and all communications went through her. Ultimately, large quantities of data and draft reports were transferred electronically in condensed format, a system that proved to be very reliable and efficient. Other challenges to manageability were grounded in competing demands especially on Cousins, but also on members of the Indian evaluation team. At times, evaluation tasks were difficult to get to in the face of more immediate and pressing demands. The preparation of the final polished and formatted version of the report was delayed for several months, for example.
5. *Depth of participation (5)*: Without question Indian team members participated in all phases of the evaluation process. Planning was done collaboratively during the first site visit. The program practitioners drafted initial versions of questionnaires and interview schedules and reacted to drafts of focus group questions. They implemented the questionnaire survey of alumni and a comparison group of practising principals and helped to interpret statistical summaries provide by Cousins. They carried out several focus groups and case school data collection site visits. Through exchanges with the Canadian

counterparts, they acted on recommendations for data analysis and reporting. Ultimately, the second site visit was a protracted and intensive cross-method interpretation session. Once the final report was compiled as a complete whole by Cousins, the Indian team members provided extensive constructive feedback and suggestions for change.

Discussion

Figure 2 shows the distribution of process dimension ratings for each of the two cases. Empirically, these ratings should be treated with caution since we did not endeavour to establish inter-rater agreement, and therefore inter-subject differences are likely to be inherent in the ratings. The point of the exercise was to test the application of the process framework to concrete collaborative projects.

We were successful in applying the ratings and showing similarities and differences between the two projects. Both projects had similar rationales with the main emphasis being practical. Conceptual, instrumental and symbolic consequences of the project were anticipated. The projects looked quite similar in terms of the five process dimensions that we identified. Control was balanced, a diverse group of participants were involved, and power relations were not a defining issue. The projects tended to be somewhat unwieldy and to involve non-evaluator stakeholders in a full range of evaluation tasks.

If the projects were to be framed as P-PE's it is interesting to note some differences from the hypothetical example in Figure 1. The hypothetical example was developed by Cousins based on his experience over time with P-PE (e.g., Cousins & Earl, 1995). In the present cases more diversity was observed than would be expected. Also, probably for a related reason, the projects were somewhat difficult

to manage. Otherwise, the P-PE experiences were similar to previous reported experiences.

One interesting observation regarding the use of the process framework was that intra-project variability was in evidence. Ratings according to some process dimensions could be observed to shift over time as was the case with the ‘control of technical decision making’ dimension in the palliative care case. Also the nature of conflict among participants was seen to shift to a more neutral posture during the evaluation in over time in that case. In the ELP context, advisory structures were set up and informed the evaluation in various ways. These committees revealed diversity in participation at an aggregate level but such diversity was seen to be more limited at the evaluation team level. These observations may be construed as limitations in the current application because ratings were made on an aggregate or holistic basis. However, they speak to the dynamic nature of the participatory process. The implications of the aforementioned limitations for ongoing research using the framework would be to invoke longitudinal designs that capture varying units of analysis.

Despite the limitations of the present test of the framework, the reconceptualized version of process dimensions for collaborative inquiry shows promise for being a helpful way to think about collaboration. Potentially the framework could be used to guide research on collaborative, participatory and empowerment processes, conditions affecting them and their consequences and effects, preferably using longitudinal, multilevel designs as mentioned above. We have argued elsewhere that such research is badly needed (Cousins, 2003). Despite a good deal of reflective anecdotal reporting of practice (not unlike that reported in the present paper) more intensive empirical efforts such as indepth case study research, longitudinal qualitative and quantitative designs are few and far between. Yet

interest in participatory inquiry is on the rise. Further, some studies have shown that implementation can be extraordinarily challenging and may lead to blatantly unsuccessful outcomes. The present tool will help researchers to clarify important implementation issues perhaps as a way of linking these to antecedent conditions or even consequences, intended and unintended.

The tool can also be of use to evaluation practitioners, donor agencies and others interested in collaborative modes of inquiry. Much is written about such processes but evidence suggests that projects touted to be participatory are anything but. This was the clear conclusion of a recent study of alleged participatory studies in the education sector in sub-Saharan Africa (Meier, 1999). Some writers would argue that so called participatory models and approaches should be ‘problematized’ since they may become effective tools for maintaining the status quo (e.g., Gregory, 1998). Understanding more about participatory processes and how they relate to intended and unintended consequences could be useful for helping practitioners to operationalize participation and collaboration in ways likely to bring about the sorts of benefits anticipated in the first place.

Author Biographies

TLynda Weaver

For more than 2 decades, Lynda Weaver has worked in the area of health care services research, with a focus on program planning and evaluation. She completed a Masters of Health Administration from the University of Ottawa in 1994, and a Masters in Education for Health Professionals in 2001 from the Ontario Institute for Studies in Education/University of Toronto. Lynda has been with SCO Health Service’s Palliative Care Program since 1995 and is currently in the role of Coordinator Palliative Care Education and Quality Management. Her current

interests include the evaluation of palliative care education program planning and evaluation.

J. Bradley Cousins

Brad Cousins, Ph.D. (Toronto), is professor of educational administration at the Faculty of Education, University of Ottawa, Canada. His research interests are located broadly in the domain of evaluation and knowledge utilization with particular focus on collaborative and participatory modes on social inquiry. Professor Cousins is winner of the 1999 'Contributions to Evaluation in Canada Award' sponsored by the Canadian Evaluation Society and is currently editor-in-chief of the Canadian Journal of Program Evaluation.

References

- Brunner, I. & Guzman, A. (1989). Participatory evaluation: A tool to assess projects and empower people. In R. F. Conner & M. Hendricks (Eds.), *International innovations in evaluation* (pp. 9-18).
- Bryk, A. (1983). *Stakeholder-based evaluation. New Directions in Program Evaluation, No. 17*. San Francisco: Josey-Bass.
- Cousins, J. B. (2003). Utilization effects of participatory evaluation. In T. Kellaghan, D. Stufflebeam, with L. Wingate (Eds.). *International Handbook of Educational Evaluation* (pp. 245-266). Boston: Kluwer.
- Cousins, J. B., Donohue, J. J., & Bloom G. A. (1996). Collaborative evaluation in North America: Evaluators' self-reported opinions, practices, and consequences. *Evaluation Practice, 17*(3), 207-226.

- Cousins, J. B., & Earl, L. M. (Eds.). (1995). *Participatory evaluation in education: Studies in evaluation use and organizational learning*. London: Falmer Press.
- Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. In E. Whitmore (Ed.), *Understanding and practicing participatory evaluation: New Directions in Evaluation, No. 80* (pp. 3-23) Vol. 80,. San Francisco: Jossey Bass.
- Garaway, G. B. (1995). Participatory evaluation. *Studies in Educational Evaluation, 21*, 85-102.
- Gregory, A. (2000). Problematizing participation: A critical review of approaches to participation in evaluation theory. *Evaluation, 6*(2), 179-199.
- Huberman, M. (1994). Research utilization: The state of the art. *Knowledge and Policy, 7*(4), 13-33.
- Levin, B. (1993). Collaborative research in and with organizations. *Qualitative Studies in Education, 6*(4), 331-340.
- Mark, M., & Shotland, R. L. (1985). Stakeholder-based evaluation and value judgments. *Evaluation Review, 9*, 605-626.
- Meier, W. (1999). *In search of indigenous participation in education sector studies in Sub-Saharan Africa*. Unpublished Master's thesis, University of Ottawa, Ottawa.
- Schwandt, T. A. (1997). Reading the "problem of evaluation" in social inquiry. *Qualitative Inquiry 3*(1), 4-25.

Zen and the Art of Everyday Evaluation

The First *JMDE* Essay Competition

Michael Scriven

Zen Buddhism was often said, by those subscribing to other varieties of Buddhism, to be “the last guest at the table.” This was a kindly reference to its nouveau status, since it was not part of the centuries-long Indian and then Chinese phases in the history of Buddhism, only emerging in the last historic phase of development, in Japan. Although Buddhism is, in general, a sect not intensely concerned with theology in the Western sense, the Zen version is opposed to the intellectualization of religious ontology and epistemology more strenuously than its predecessors, perhaps because of the direct encounter with the theology the Jesuits brought to Japan. The Zen attitude towards reason was associated with the legacy of the koans, questions and puzzles that were designed to show the limitations of reason, and, by implication, of reason about deep matters concerning life and existence. Of the supposedly 1700 koans, by far the most famous is the question, “What is the sound of one hand clapping?”

Despite this skepticism about rational theology, some highly intelligent, although disparate, efforts have been made by Western intellectual writers such as Aldous Huxley and Arthur Koestler to express the essence of Zen. The disagreement amongst them is considerable: it would not be hard to find sources amongst them that would deny every comment made so far in this note, although these are remarks based on august sources in Western history of philosophy. Robert Persig’s *Zen and the Art of Motorcycle Maintenance* is one of the most interesting of these

high literary efforts and is perhaps unique in a respect that is not commonly remarked: it is about evaluation. This is quite overt, since the subtitle of the book is *An Inquiry into Value*, and the Platonic inscription is:

And what is good, Phaedrus,

And what is not good—

Need we ask anyone to tell us these things?

And here is what Persig says about his own effort:

“I would like not to cut new channels of consciousness but simply dig deeper into old ones that have become silted in with the debris of thoughts grown stale and platitudes too often repeated. “What’s new?” is an interesting and broadening eternal question, but one which, if pursued exclusively, results only in an endless parade of trivia and fashion, the silt of tomorrow. I would like, instead, to be concerned with the question “What is best?” a question that cuts deeply rather than broadly, a question whose answers tend to move the silt downstream. There are eras in human history in which the channels of thought have been too deeply cut, and no change was possible, and nothing new ever happened, and “best” was a matter of dogma, but that is not the situation now.” (p. 16). The quotes are all from the importantly altered 25th anniversary edition (Morrow, 1999).

That can be an inspiring thought for us in evaluation, at the start of a new century. Evaluation is a discipline built by constructing a science out of extensions of everyday evaluation, just as probability theory is an extension into mathematics of everyday reasoning about games of chance. From time to time in these pages we will revisit Persig’s theme and thoughts. On this occasion, we’ll simply pose an anti-koan, a puzzle about what is “best,” a puzzle that is intended to produce, not

laughter at the flounderings of reason but rather reason's best exercise in the pursuit of value. The puzzle is this:

“In the evaluation of revolutions—political or intellectual, in medicine, in warfare, or in education—should one use, as a basis, the values of the victor; or those of the vanquished; or both; or neither?”

We read all the time about revolutionary new technology, revolutionary new ways to prevent or cure diseases, or to deal with crime or terrorism; or revolutions in the government of countries or the treatment of the oppressed. Each of these presents an evaluation problem for the historian or other evaluator and it's a notoriously difficult problem to deal with. Is there any common thread that should run through the best approaches, whatever disciplines are involved? Perhaps a good problem for those interested in multidisciplinary evaluation!

There will be a small prize for the best short essay on this topic; perhaps appropriately, a good book, suitably inscribed. The winning submission, and perhaps others deemed amongst the best entries, will be published here. Five hundred words should suffice; a thousand will be considered excessive in the present context. The prize will be awarded by mid-summer, 2005; but the topic will not be closed to further discussion in these pages thereafter. Suggestions for the Second JMDE Essay Competition will also be welcomed and considered carefully.

Evaluation Humor

Michael Scriven

A section of *JMDE* whose importance is much greater than its size . . . maybe.

Here's an opening entry . . . could this be the most important reference in your professional life?

The *Journal of Nondestructive Evaluation*. Convince your clients that evaluation is a positive force! give them a copy of a good article on Appreciative Inquiry and a copy of the Table of Contents of, for example, the current issue of *JNDE* (vol. 23), which follows:

1. Influence of Wall Thickness on the Ultrasonic Evaluation of Small Closed Surface Cracks and Quantitative NDE
2. Rayleigh Wave Propagation for the Detection of Near Surface Discontinuities: Finite Element Modeling
3. Residual Magnetic Flux Leakage: A Possible Tool for Studying Pipeline Defects
4. Review of Advances in Quantitative Eddy Current Nondestructive Evaluation
5. Using a Single Transducer Ultrasonic Imaging Method to Eliminate the Effect of Thickness Variation in the Images of Ceramic and Composite Plates

A gift subscription is probably not a good idea, especially at \$650!