

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

Part II

Global Review: Regions	45
Evaluation Activities in Africa, <i>Lamphear, P. A</i>	46
Evaluation Activities in Australasia, <i>Risley, J. A</i>	50
The State of Evaluation in Canada, <i>Coryn, C. L. S</i>	54
Evaluation in Europe: An Overview, <i>Schröter, D. C</i>	68
Evaluation Activities in the United Kingdom, <i>Risley, J. A</i>	77
Evaluation in Eastern Europe and the Middle East, <i>Gugiu, P. C</i>	81
Evaluation in Latin America and the Caribbean: An Overview of Recent Developments, <i>Chianca, T. & Youker, B</i>	89

Part III

Global Review: Publications	102
What's Happening in AJE (2003-2004), <i>Wingate, L</i>	103
Evaluation: The International Journal of Theory, Research and Practice (2003-2004), <i>Schröter, D. C</i>	114
The Japanese Journal of Evaluation Studies, <i>Risley, J. A</i>	121
Journal of Evaluation and Program Planning, <i>Switalski Schinker, R</i>	123
New Directions for Evaluation (Vol. 102), <i>Keating, M</i>	129
Education and Personnel Evaluation: A Review of Leading Journals for the Period 2003-2004, <i>Persaud, N</i>	135

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,

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

epistemological) as well as to dimensions of process (i.e., control of technical decision making, stakeholder selection, depth of participation).

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.

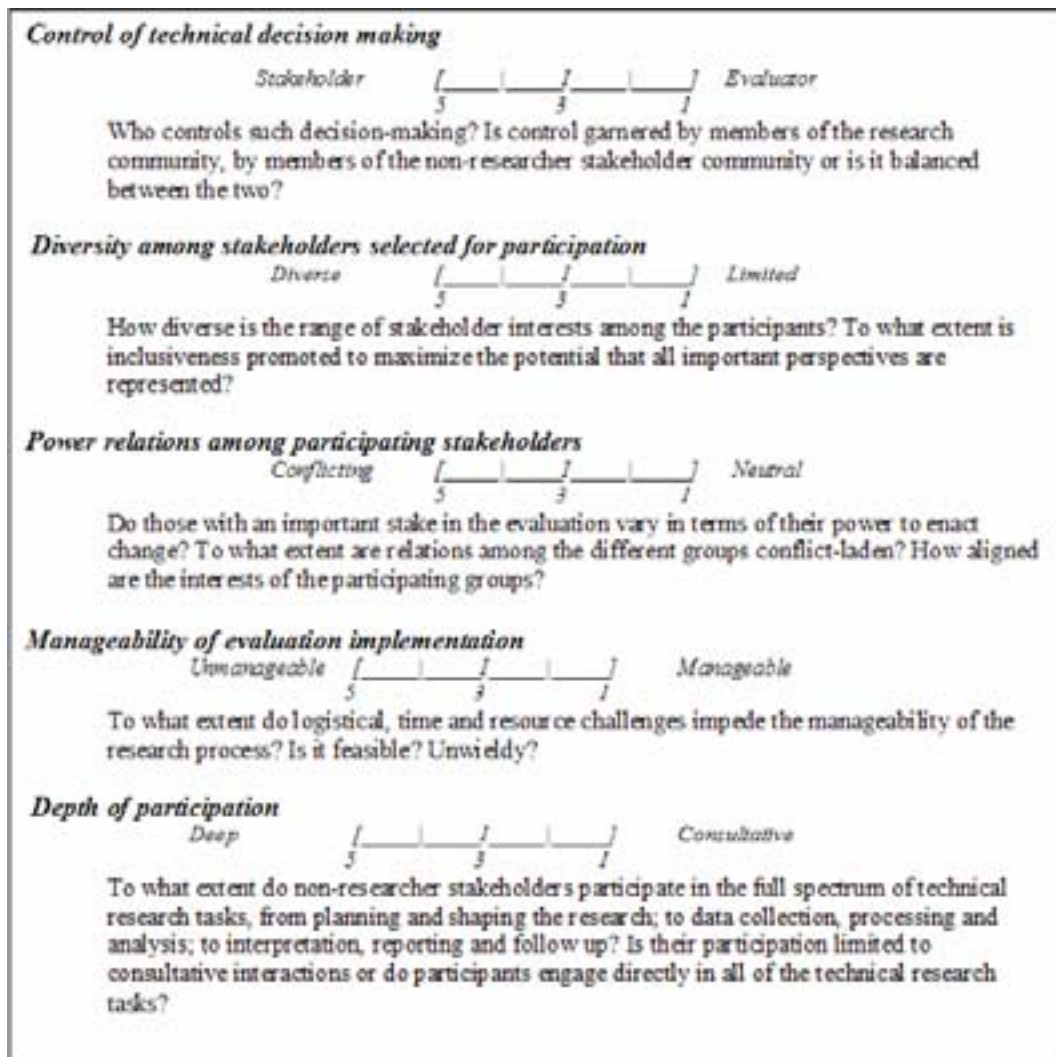
A second justification is *political* and is ideologically rooted in normative conceptions of social justice and the democratic process. The primary interest of collaborative inquiry that subscribes to such political aims is to promote fairness through the involvement of individuals associated with all 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.

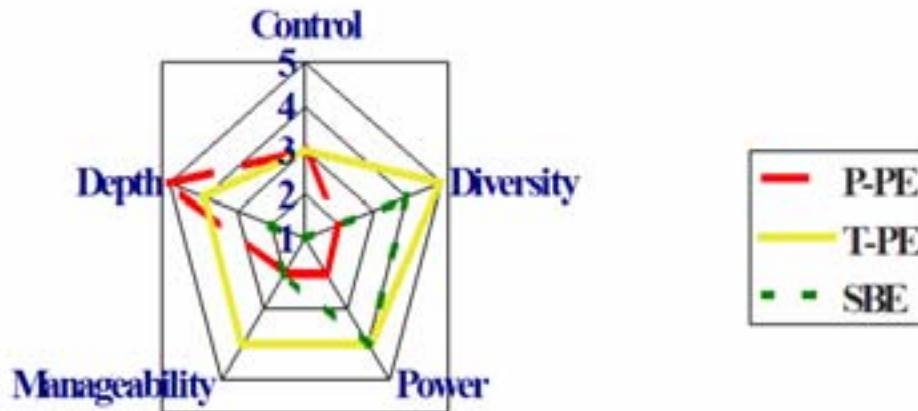


Figure 1: Five dimensions of form in collaborative inquiry

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.

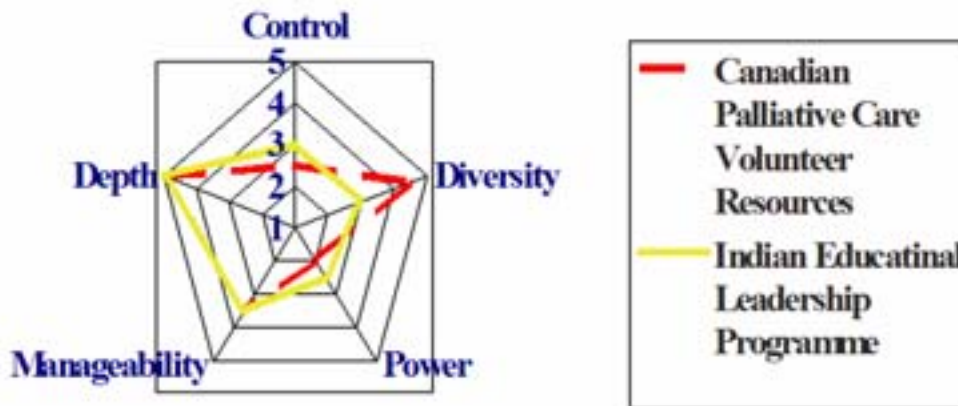


Figure 2: Comparison of Canadian and Indian P-PE cases

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

Lynda 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 floundering 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!

Global Review: Regions

Evaluation Activities in Africa

Paul A. Lamphear

Africa has seen a significant growth in Evaluation networking in the past five years with the founding of the [African Evaluation Association \(AfrEA\)](#) in 1999. As of this writing, AfrEA has 16 national associations under its umbrella, each with evaluator networks supporting their respective countries. The associations are in a wide disparity of development maturity, but are all aimed at supporting socio-economic development programs in their countries, as in this description of the Uganda Evaluation Association as posted in www.kabissa.org.

The major purpose of the association is to build individual and institutional capacity in policy, program and project evaluation in Uganda through local and global networking, training, skills development and other avenues for professional development in evaluation practice⁸.

The first conference of the African Evaluation Association, held in Nairobi in connection with the creation of AfrEA, was attended by more than 300 evaluators from 35 countries. Michael Quinn Patton was invited as keynote speaker, promoting a focus on ethics and guiding values for professional evaluation associations. He also emphasized the need for a culturally defined set of standards for evaluation and the necessity for utilization-focused evaluation.⁹ Patton referred to the historically poor “expatriate” evaluations with poorly briefed evaluators, performed too quickly, with inadequate reports.

⁸ <http://www.kabissa.org/contact/index.php?action=view&id=519>

⁹ Full transcript of Patton’s lectures are available at <http://www.afrea.org/>.

It appears that these lectures, the training seminars, and 80 papers that were submitted on evaluation, helped to ‘jumpstart’ the development of the national associations. Although there does not appear to be any evaluation journals specific to or published in Africa, AfrEA has worked on several projects to increase evaluation capacity and foster a consistent professional approach for evaluators. In 2002, the association completed the "African Evaluation Guidelines", a cultural adaptation of the US “Program Evaluation Standards”, published in both English and French.

At the 2nd AfrEA conference in 2002, a variety of international evaluators were invited, including keynote speakers Prof. Anna Madison of Cornell University, USA, Ada Ocampo, Leader of the Latin American Evaluation Network, and Penny Hawkins, President of the Australasian Evaluation Association. After 5 days of trainings and paper submissions, the association recommended that the African Evaluation Guidelines (AEG) be adopted by all the National Networks, by Government and Public bodies, and by UN Agencies and other Multinational Organizations performing evaluation in Africa. The AEG provides a checklist of 30 items essential for quality control in evaluation and it has now been used by several national governments in major development program evaluations.

The [Niger Network of Monitoring and Evaluation \(ReNSE\)](#) has over 140 documents (in English or French) and in 2004 has published their first newsletter (in French), available from their website, to promote expertise in Nigerian evaluators.

AfrEA supports dissemination of monitoring and evaluation resource materials focused on Africa in the areas of Agriculture, Conservation, Gender, HIV/AIDS,

and Poverty. Additionally, AfrEA encourages members to participate in the Xceval listserve, a discussion forum for evaluators from developing countries.¹⁰

AfrEA has significant connections with international organizations and has activities currently sponsored by:

African Development Bank (AfDB)

Agence Intergouvernementale de la Francophonie

Catholic Relief Services

Canadian Institutes for Health

Canadian International Development Agency (CIDA)

CARE International

Danish Agency for Development Assistance (DANIDA)

Family Health International (FHI)

International Development Research Centre (IDRC)

World Conservation Union (IUCN)

Norwegian Ministry of Foreign Affairs

UNAIDS

UNCHS

UNDP

UNICEF

UNIFEM

World Bank

¹⁰ Xceval-subscribe@yahoogroups.com

The third national [Conference](#) of the African Evaluation Association will be held in Cape Town, South Africa, from December 1st through December 4th, 2004, and is looking for speakers and attendees.

Evaluation Activities in Australasia

John S. Risley

General Summary of Activities

The Australasian Evaluation Society (AES) produces, and posts on their website (www.aes.asn.au), an e-newsletter approximately twice per year. The AES also holds an annual conference, usually in September or October. The 2004 conference is in October near Adelaide, Australia and will focus on “Diverse Voices in Evaluation.” Last year’s conference emphasized evaluation and indigenous peoples. Many pre-conference workshops are offered. AES has regional representatives from throughout Australia and New Zealand. There is a New Zealand Listserv—Evaluation Aotearoa—that discusses “evaluation research.” It only has a few posts per month, mostly dealing with Auckland Evaluation Group activities.

From reading the editorials and other non-refereed articles in the *Evaluation Journal of Australasia (EJA)* it appears that the evaluation profession in Australasia differs from the profession in the United States in two main ways. First, evaluators come from more diverse academic and professional backgrounds in Australasia than in the United States. Second, Australasian evaluators are much less likely to be associated with a university and much more likely to be employed by a government agency than are American evaluators.

Evaluation Journal of Australasia

A recent editorial in *Evaluation Journal of Australasia (EJA)* noted the history of AES publications. The society launched the *EJA* in 1989. Then from 1993 through 2000 AES published both *EJA* and *Evaluation News & Comment*. In 2001 these publications merged to form the new series of *EJA*. The journal is published by the AES bi-annually (though recently there have been delays in publishing new editions). AES posts the two copies preceding their most recent issue on their web site. The journal includes refereed and non-refereed articles, editorials, interviews with evaluators from both within and without the region, book reviews, research reports, and information about the annual AES conference.

Issues addressed in *EJA* included much information concerning cultural appropriateness, indigenous peoples, and diversity in evaluation. This may be a reflection of the recent AES conference themes. There is some material drawing distinctions about evaluation aspects specific to Australasia, but many articles are written by authors outside the region about subjects not specific to the region.

Subjects of refereed articles in recent issues of *EJA* include: evaluation of options for changing port ownership in Belfast, an evaluation of a respite care program in Christchurch, evaluating the cultural appropriateness of human service delivery programs in Australasia, and the TRIAGE (Technique for Research of Information by Animation of a Group of Experts) technique. A few refereed articles were short (3 pages and 5 pages) compared to articles in the *American Journal of Evaluation*, for example.

Some of the refereed articles had very little to say about evaluation. For example, one of these articles (Burton & Rajan, 2002) concerned a case study evaluation of 15 people seriously injured in workplace accidents. The authors described the

project's goal as exploring the social and economic consequences to society from these workplace injuries. The article discussed the methodology of the study, the experiences of the researchers, and the lessons learned from their research experiences. The methodology was basically a semi-structured interview of injured workers, their family members, employers, etc. The lessons learned by the researchers were: 1) interviewing can be exhausting, 2) diversity of the project team was essential, and 3) it was difficult to remain objective after seeing the suffering of the injured workers.

One interesting article (Sigsgaard, 2002) addressed an unusual methodology (in evaluation research), the Most Significant Change (MSC) methodology. The author, Peter Sigsgaard, works at a Danish NGO called "MS" on measurement and evaluation issues. He gave examples of his experience using MSC in evaluating partnership-based economic development programs in Africa, Asia and Central America. Using MSC you ask people to identify positive or negative changes they have observed within a given "domain of interest." People are then asked which change, positive or negative, they think is most significant and why. More important or very large changes that are reported are verified by further investigation.

Sigsgaard (2002) contrasts this approach with one previously used by MS in evaluating these programs, in which they would conceive of indicators to measure and then cast about looking for these indicators. This led to lots of time spent looking for, and not finding, specific data.

It makes intuitive sense to ask program consumers what changes are occurring due to the program. It does highlight the need to be careful how one measures program changes.

References

- Burton, J., & Rajan, R. (2002). Revealing the hidden costs: research experiences from a case study evaluation project. *Evaluation Journal of Australasia (new series)*, 2(2), 69-73.
- Sigsgaard, P. (2002). Monitoring without indicators: an ongoing testing of the MSC approach. *Evaluation Journal of Australasia (new series)*, 2(1), 8-15.

The State of Evaluation in Canada

Chris L. S. Coryn¹¹

Background and General Context of Organized Evaluation in Canada

The [Canadian Evaluation Society](#) (CES)—Canada's official professional organization for evaluation—serves as the country's core for evaluation related activities. CES is similar to many other evaluation associations around the globe, but differs from the American Evaluation Association in that the majority of its members are from the government sector and evaluators practicing in NGOs, para-government, and the public and private sectors. Presently, the CES has 12 regional chapters, including:

- ✓ [Newfoundland & Labrador](#)
- ✓ [Prince Edward Island](#)
- ✓ [Nova Scotia](#)
- ✓ [New Brunswick](#)
- ✓ [Quebec](#)
- ✓ [National Chapter](#)
- ✓ [Ontario](#)
- ✓ [Manitoba](#)
- ✓ [Saskatchewan](#)
- ✓ [Alberta](#)
- ✓ [British Columbia](#)
- ✓ [Northwest Territories](#)

Since 1991 the CES membership has grown to over 1,750 individual Canadian and student members, as well as over 100 international members (CES, 2004).

¹¹ The author would like to thank J. Bradley Cousins (editor of the *Canadian Journal of Program Evaluation*) for his insight and assistance.

The CES offers a wide range of resources and services for practicing evaluators and students of the discipline including: a comprehensive Web site (available in English and French); an evaluation report bank (academic, government, and private sector reports); a fully-searchable database—the [Grey Literature Bank](#) (unpublished documents of interest to evaluators); a professional development series of workshops; an [annual conference](#) (including the upcoming 2005 joint conference with the American Evaluation Association); and the *Canadian Journal of Program Evaluation*. The CES efforts are strongly supported by the Government of Canada, which has its own specialized evaluation unit; [Evaluation and Data Development](#) (EDD). EDD is one of the largest evaluation shops in the Federal Government of Canada, and focuses primarily on governmental initiatives including analysis of government policy and evaluation of government programs, for example, Human Resources Development Canada (HRDC) programs. Other government contingencies which influence the Canadian evaluation field include the [National Science and Engineering Research Council](#), the [Social Science Research Council](#), [Transport Canada](#), [Industry Canada](#), [Health Canada](#), the [Treasury Board Secretariat](#), and the [Canadian International Development Agency](#); each of which are also sponsors of the CES. Informed decision making is further facilitated by [Statistics Canada](#) a provider—federally legislated—of statistical data for the whole of Canada and each of its provinces that is intended to inform Canadian citizens and other key stakeholders regarding Canada's population, resources, economy, culture, and society.

In the summer of 2001 the CES announced their new vision, mission, and goals for the future (Canadian Evaluation Society Newsletter, Summer 2001):

Vision: The Canadian Evaluation Society will be the leader for evaluation in Canada and a major contributor in the global evaluation community

Mission: The society is a Canada-wide non-profit bilingual association dedicated to the advancement of evaluation theory and practice.

Goals:

1. **Leadership**—To provide leadership to individuals and organizations in support of evaluation theory and practice in Canada and the global community.
2. **Knowledge**—To improve the state of evaluation theory and practice.
3. **Advocacy**—To promote the importance of an evaluation culture.
4. **Professional Development**—To promote and facilitate the enhancement of evaluation capacity for members and non-members.

The CES also supports various student initiatives including the [CES Student Case Competition](#) and [student paper contest](#) (for undergraduate and graduate students in the field of evaluation). The CES Student Case Competition (initiated in 1996), is an annual event in which teams of three to five students from Canadian colleges and academic institutions compete in the analysis of an evaluation case file. In a preliminary competition, all teams receive on the same day the key to an evaluation case file that has been hidden on the Web. They have five hours to prepare an analysis and then submit it by e-mail for judging by an expert panel. The three best teams are invited to participate in a final round, held at CES's annual conference, in which they must analyze a new case and present findings and recommendations before a live audience. The team that makes the best presentation takes possession of the Case Competition Trophy for a year, receives prizes, and is given visibility in various publications.

Evaluation Education Programs in Canada

As of 2000 (CES), over 25 Canadian institutions/colleges/universities offered more than 100 evaluation-related courses across a wide array of academic disciplines (e.g., psychology, political science, public administration, economics)—a complete institution, department, and course list is available at <http://www.evaluationcanada.ca/txt/outline200106.pdf>.

Professional Development of Canadian Evaluators

The CES plans to focus on two key areas in the upcoming years: (1) professional development of its members, and (2) advocacy on behalf of the evaluation function. The articulation of a Core Body of Knowledge (CBK) will guide the Society's professional development and advocacy activities (Canadian Evaluation Society, 2004). The CBK comprises theories, skills, and best practices that people must possess to plan, carry, out, and report on valid and reliable evaluations of programs or policies in governments, not-for-profit organizations, and businesses.

Essential Skills. Much of the emphasis on professional development is funneled through the CES Essential Skills Series. Regional chapters offer this series as well as any other form of training they consider adequate for their members. These essential skills include:

1. Understanding Program Evaluation
 - Key terms and concepts
 - Benefits of program evaluation
 - Basic steps in the evaluation process
 - Major approaches to program evaluation

- Formatting evaluation questions
- Designing an evaluation
- Evaluating with limited resources
- Analyzing and reporting evaluation results
- Reducing resistance to evaluation
- Involving staff and clients in the evaluation process
- Increasing evaluation utilization
- Making evaluations ethical and fair

2. Building an Evaluation Framework

- Identifying who the client is and what the client needs
- Basic concepts of needs assessment
- Major approaches to assessing client needs
- Evaluation methods for "getting close to the client"
- Building an evaluation framework through logic models
- Involving managers and staff in building an evaluation framework
- Relating program design to client needs
- Defining program components
- Formulating indicators for program success
- Using the evaluation framework for linking program performance to client needs

3. Improving Program Performance

- Using evaluation as a management tool for improving program performance and enhancing internal accountability
- Basic concepts of monitoring and process evaluation
- Monitoring program performance with existing administrative data and information systems
- Developing ongoing data collection instruments and procedures
- Linking process evaluation to program decision-making
- Assessing client satisfaction
- Understanding continuous quality improvement
- Using program evaluation for building a "learning organization"

4. Evaluating for Results

- Defining program results
- Major approaches to evaluating results
- Developing results measures
- Designing outcome evaluations
- Validity and reliability
- Appropriate use of quantitative and qualitative techniques
- Relating program results to program costs
- Understanding program benefits
- Measuring program equity and responsiveness to community needs

- Communicating evaluation findings
- Using evaluations to improve program effectiveness and accountability

(Canadian Evaluation Society, 2004)

Certification of Evaluators in Canada. As a body representing program evaluators across Canada and promoting the program evaluation function in Canadian institutions, the CES is concerned with the sustainability, growth and strengthening of the profession. In recent years, this concern has led the Society to consider issues related to increasing professionalization, through means such as professional development programs, development and adoption of practice standards and [ethical guidelines](#), and certification of members. This issue remains unresolved, but is becoming increasingly acute in the wake of recent developments in the federal government sector that have raised the profile of auditing (Cousins, 2004).

This latter issue — developing a form of certification for members — would be a major step for the CES. Therefore, it was the subject of an [in-depth study of the experience of several other organizations with certification](#) (Long & Kishchuk, 1997). A second study, carried out in 1999, reports on a [pilot survey of clients and employers](#) (Stierhoff, 1999) on their views regarding certification of evaluators.

Canadian Journal of Program Evaluation

The *Canadian Journal of Program Evaluation (CJPE)* was launched in 1986 and is published twice a year (available at www.cjpe.ca). *CJPE* is sponsored by the CES and the University of Calgary. Individual issues and articles can be downloaded by non-members for a nominal cost. *CJPE* seeks to promote the theory and practice of program evaluation in Canada by publishing:

- Articles on all aspects of the theory and practice of evaluation, including methodology, evaluation standards, implementation of evaluations, reporting and use of studies, and the audit or meta-evaluation of evaluation.
- Research and Practice Notes that provide practical examples of the applications of particular methodologies or procedures within the context of a particular study or group of studies.
- Book Reviews of relevance to the practice in Canada.

(Canadian Journal of Program Evaluation, 2004)

Review of the past eight issues (from Spring 2001 to Spring 2004) of *CJPE* revealed a number of insights into the journal's thematic trends. The journal does, in fact, promote and publish articles on theory, practice, implementation, and standards, for example. Notable examples include Christie & Rose (2003)—*The language of evaluation theory: Insights gained from an empirical study of theory and practice*, Levin-Rozalis (2003)—*Evaluation and research: Differences and similarities*, Morris (2002)—*The inclusion of stakeholders in evaluation: Benefits and drawbacks*.

In 2001, the *CJPE* devoted a special issue to provincial evaluation policy and practice in Canada. Accounts of provincial evaluation activity were provided for British Columbia (McDavid, 2001), Alberta (Bradley, 2001), Manitoba (Warrack, 2001), Ontario (Segsworth, 2001), Quebec (Cabatoff, 2001), Prince Edward Island (Mowry, Clough, MacDonald, Pranger, & Griner, 2001), Newfoundland (Ross, 2001) and the Northwest Territories (Hicks, 2001). "Being the first ever account of evaluation activity at the provincial level, this collection of articles represented a very important contribution to the knowledge of evaluation practice in Canada" (Gauthier, Barrington, Bozzo, Chaytor, Cullen, Lahey, Malatest, Mason, Mayne,

Myers, Porteous, & Roy, 2004). A number of general and specific conclusions were drawn about the state of affairs in Canadian evaluation as a result of this special issue and were summarized in *The lay of the land: Evaluation practice in Canada today* (Gauthier et. al., 2004). The authors conclude that program evaluation in Canada:

- Has not acquired an identity of its own
- Tends to neglect key issues
- Loses emphasis on rigor
- Is dominated by program monitoring
- Is insufficiently connected with management needs

Regional Perspectives

Perspectives across Canada's various regions are briefly summarized below. This summary includes: (1) strengths, (2) weaknesses, (3) threats, and (4) opportunities of and for evaluation in western Canada, Alberta, and Ontario, as well as potentials for evaluation teaching and learning.

The Western Canadian Perspective. (Malatest, 2004)

Strength: Development of evaluation methodologies—in recent years the provincial and federal agencies have recognized the requirement of good evaluation.

Weakness: Inadequate planning of program evaluations—awareness and use of evaluation tools are often an afterthought.

Threat: Reduced program evaluation capacity—the ability to design and manage complex evaluation activities has been compromised (e.g., lack of resources).

Opportunity: Managing for outcomes—activities in British Columbia and Alberta have been strengthened by strong government-wide commitment to measure and report on the key outcomes for almost all ministries and/or departments.

Program Evaluation in Alberta. (Barrington, 2004)

Strength: Growing sophistication—evaluators are more skilled and better qualified.

Weakness: Dependence on performance measurement—to the exclusion of more relevant, complex outcomes.

Threat: Devaluation—avoidance of serious evaluation (e.g., focus on accountability rather than improvement).

Opportunity: Linking accountability and evaluation—evaluators believe that they can make evaluation more rigorous and more useful.

Program Evaluation in Ontario. (Mason, 2004)

Strength: Commitment—Ontario government is being steered toward evaluation by political interest.

Weakness: The paradigm—the current approach is to assist the government in determining redirection of funding.

Threat: Capacity—Public and non-profit organizations need to demonstrate effectiveness, yet they are limited in their capacity to meet this demand.

Opportunity: Collaboration and partnerships—potential to pool evaluation resources across different funders and different funding interest.

Teaching and Learning Evaluation in Canada. (Chaytor, 2004)

Strength: Self-definitional capacity—the time for evaluation to define itself and establish itself as a distinct discipline is "now."

Weakness: Lack of disciplinary focus—disciplines view evaluation differently rather than having a common ground.

Threat: Disconnection—evaluation as part of management is under threat (e.g., lack of common ground).

Opportunity: Demand for skills—recognition of the value of evaluation and the demanding skills required.

This paper is an outsider's perspective of evaluation in Canada. Any errors or omissions are entirely unintentional. Comments, questions, and criticism are certainly welcomed. As new and additional information becomes available it will be made available in upcoming issues of the Journal of MultiDisciplinary Evaluation. If you would like to provide additional information or insight regarding the state of evaluation in Canada please contact the the author via e-mail at: christian.coryn@wmich.edu

References

- Bradley, S. E. (2001). Evaluation in the government of Alberta: Adapting to the "new way." *The Canadian Journal of Program Evaluation*, 16(special issue), 29-44.
- Cabatoff, K. (2001). The long march from evaluation to accreditation: Québec's new government management framework. *The Canadian Journal of Program Evaluation*, 16(special issue), 73-88.
- Canadian Evaluation Society (2004). *Canadian evaluation society website*. Available at http://www.evaluationcanada.ca/site.cgi?s=1&ss=0&_lang=an
- Canadian Evaluation Society (2001-2004). *Canadian evaluation society quarterly newsletter*. Available at <http://www.evaluationcanada.ca/site.cgi?s=4&ss=3&lang=an>
- Canadian Evaluation Society (2004). *CES guidelines for ethical conduct*. Available at http://www.evaluationcanada.ca/site.cgi?section=5&ssection=4&_lang=an
- Canadian International Development Agency (2004). *Canadian International Development Agency website*. Available at <http://www.acdi-cida.gc.ca/index.htm>
- Canadian Journal of Program Evaluation (2004). *Canadian journal of program evaluation website*. Available at http://www.evaluationcanada.ca/site.cgi?s=4&ss=2&_lang=an
- Cousins, J. B. (2004). Personal communication.

Evaluation Data and Development (2004). *Evaluation Data and Development website*. Available at <http://www11.hrdc-drhc.gc.ca/pls/edd/hrdc.main>

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. *The Canadian Journal of Program Evaluation*, 19(1), 143-178.

Health Canada (2004). *Health Canada website*. Available at <http://www.hc-sc.gc.ca/>

Industry Canada (2004). *Industry Canada website*. Available at <http://www.ic.gc.ca/>

Hicks, K. (2001). Program evaluation in the government of the Northwest Territory, 1967-2000. *The Canadian Journal of Program Evaluation*, 16(special issue), 107-114.

Long, B. & Kishchuk, N., (1997). *Professional certification: A report to the national council of the Canadian evaluation society on the experience of other organizations*. Canadian Evaluation Society.

McDavid, J. C. (2001). Program evaluation in British Columbia in a time of transition. *The Canadian Journal of Program Evaluation*, 16(special issue), 3-28.

Mowry, S., Clough, K., MacDonald, B., Pranger, T., & Griner, D. (2001). Evaluation policy and practice in provincial governments province of Prince Edwards Island. *The Canadian Journal of Program Evaluation*, 16(special issue), 89-100.

- Ross, A. (2001). Evaluation in Newfoundland: Then is then now is now. *The Canadian Journal of Program Evaluation*, 16(special issue), 101-106.
- Segsworth, B. (2001). Evaluation policy and practice in Ontario. *The Canadian Journal of Program Evaluation*, 16(special issue), 59-72.
- Social Sciences Research Council (2004). *Social Science Research Council website*. Available at <http://www.sshrc.ca/>
- Stierhoff, K. A. (1999). The certification of program evaluators: A pilot survey of clients and employers. Canadian Evaluation Society.
- Statistics Canada (2004). *Statistics Canada website*. Available at <http://www.statcan.ca/start.html>
- Transport Canada (2004). *Transport Canada website*. Available at <http://www.tc.gc.ca/>
- Treasury Board of Canada Secretariat (2004). *Treasury Board of Canada Secretariat website*. Available at <http://www.tbs-sct.gc.ca/>
- Warrack, B. (2001). Program evaluation in the Manitoba government: past, present, and future. *The Canadian Journal of Program Evaluation*, 16(special issue), 45-58.

Evaluation in Europe: An Overview

Daniela Schröter¹²

The Landscape of European Evaluation

The umbrella organization of evaluation in Europe is the [European Evaluation Society](#) (EES). Founded in 1994 in Hague (Netherlands), the EES elects presidents for two-year mandates and provides a central secretariat for two to three years in different locations. The EES welcomes all individuals interested in evaluation, professionally or academically. Members of the EES receive a Newsletter,¹³ a one year subscription to *Evaluation: the International Journal of Theory, Research and Practice*, and reductions on EES conference fees and other activities.

The web site of the EES provides a good overview about the evaluation community including lists of European and international evaluation associations and networks, evaluation journals, events, and other online resources. Currently, the EES provides links to 13 national or multinational European organizations as well as 5 regional networks within the United Kingdom (see Figure 1). A Portuguese evaluation society will be established (News from the Community, 2004) and listserv discussions of the German Evaluation Association ([DeGEval](#)) indicate that an Austrian Evaluation Society has been formed.

¹² This is a first draft of an illustration of evaluation in Europe. The author anticipates receiving feedback from those evaluators who are experts on issues of European evaluation. Please send any comments, recommendations, feedback, and additional information to the state of art in Europe or individual European countries to: daniela.schroeter@wmich.edu. I appreciate your feedback and will gladly revise this paper accordingly.

¹³ At this point in time there are three newsletters available online at <http://www.european-evaluation.org/news/newsletters.html>. All of these newsletters are specifically addressing conference related issues in preparation for the biennial events.



[European Evaluation Society](#), [Danish Evaluation Society](#), [Finnish Evaluation Society](#), [French Evaluation Society](#), [German Evaluation Society](#), [International Program Evaluation Network](#), [Irish Evaluation Network](#), [Italian Evaluation Association](#), [Polish Evaluation Society](#), [Spanish Public Policy Evaluation Society](#), [Swedish Evaluation Society](#), [Swiss Evaluation Society](#), [UK Evaluation Society](#) (the following are regional UK networks: [Cymru Evaluation Network](#), [Scottish Evaluation Network](#), [London Evaluation Network](#), [Midlands Evaluation Network](#), [North West Evaluation Network](#)), [Walloon Evaluation Society](#)

Figure 1. *National and Multinational Evaluation Societies in Europe*

Evaluation in Europe appears to be highly influenced by the political environment. One of the most constraining elements to effective communication across the European evaluation community is the diversity of language. In addition, Elliot Stern (2004, p. 9-10) referred attendees at the 5th conference of the European Evaluation Society 2003, to four contextual dimensions that challenge and shape

evaluation in contemporary Europe specifically in public policy and civil society. These are (1) national specificity or convergence (identity), (2) cultural diversity and its limit (solidarity), (3) decentralization or supranational solutions (legitimacy), and (4) the strong state with the weak means (complexity). The task of the EES is to help minimize and overcome any barriers the European evaluation community faces.

The EES holds conferences biennially. From September 30 to October 2, 2004, the sixth conference took place in Berlin, Germany with the title *Governance, Democracy and Evaluation*. There were about 423 evaluators (as indicated on listserv discussions of the German Evaluation Society); 334 of these stem from 36 countries and 5 organizations and presented on issues related to *Governance, Democracy and Evaluation* (see [List of Presenters](#)). More specifically the [call for papers](#) welcomed contributions related to program complexity, accountability, standards and guidelines, policy implementation, knowledge management, and education in evaluation that address needs of national and international level governments. The [program](#) of the conference reflects these issues, and the proceedings may shed light on the specifics.

The Development of Evaluation in Europe

Leeuw (2004) asked if European evaluation is still an “infant industry” and illuminates the European type of “evaluation industry”. His book chapter will serve as the foundation for the following sections. Rist, Furubo, and Sandahl’s (2002)¹⁴ assessed countries worldwide on eight dimensions to determine levels of development in evaluation. The dimensions included:

¹⁴ The author did not have the study at hand and utilized Leeuw (2004) summary and comments of that study only.

- Evaluation activity
- Supply of evaluators
- Training capacity
- National discourse
- Organized evaluation meetings
- Evaluation infra-structure within the public sector
- Evaluation infra-structure within parliament
- Evaluations carried out by Supreme Audit Offices (see Leeuw 2004, 63).

While not all European countries were assessed within this study results indicated most intense evaluation efforts in North and West European countries. However, data was either insufficient or indicated only moderate training capacity for evaluators in Europe, which as Leeuw argues is plausible in view of the fact that evaluation has not been established well at the university level in form of evaluation studies. On the other hand, national discourse and organized meetings were available and as indicated by other contributions in this issue of *JMDE* not only stimulate debate and discussions, but also provide platforms for trainings. Additionally, Rist et al. found that evaluation in the public sector was more widely available than evaluation within parliament. Last but not least, evaluations carried out by Supreme Audit Offices were most developed in Sweden, the Netherlands, and the United Kingdom.

Historically, Sweden, Germany, and the United Kingdom are considered first and second wave evaluation countries where evaluation developed in the 70s and 80s. Since then, many other European countries have been established as frontrunners in evaluation, especially the Scandinavian countries (Denmark, Norway, Finland),

other West European countries (Ireland, the Netherlands, and Switzerland) as well as South European countries (France, Spain, and Italy), and the numbers are growing.

The European Evaluation Market

Based on a study conducted in 1999, Leeuw describes the European evaluation market as a growing market. While the response rate in the study was rather limited, findings indicated that the evaluation market was growing faster on the European and national levels than in regions. Most evaluations conducted were related to policy and respondents indicated that methods utilized usually derived from the evaluators' specific subject areas. Moreover, the regional evaluation market was perceived as rather fragmented and it was thought that international competition on the European evaluation market would be constrained due to cultural factors. For instance, Leeuw pointed out that one respondent said that it was even hard to hire a British evaluator for an Irish setting. This is due to language constraints and an understanding of the different organizational cultures. On the European level, this leads to evaluations which are conducted by teams of evaluators from multiple nations. Leeuw refers to such arrangements as “(quasi)professions” (p.68). Moreover, Leeuw argues that top-down processes thwart good evaluation practice. While evaluation in North America is outcome and impact oriented, European evaluation focuses on resources and administrative processes. Leeuw points out that there may be a slight drift into auditing, focusing on form rather than substance.

Evaluation on the European Union Level¹⁵

On the European level, initial forms of program evaluations began in the 80s, were focused on research and technology development programs, and were based on practices prevalent in first wave evaluations. A shift occurred in 1995, when a new evaluation scheme was introduced that demanded evaluation of research and framework programs in form of annual monitoring and five-year periodic assessments. Leeuw states:

The assessments can be understood as a combination of an ex post evaluation of the previous program, an intermediate evaluation of the current program and an ex ante appraisal of future activities (2004, 69).

However, while evaluation on the Union level always focused on regulatory policy, formal evaluation systems or databases for the Directorates General are insufficient and “the Council and Parliament pass[ed] a small number of ‘sunset’ regulations which include a formal evaluation clause given a deadline (especially in the field of Competition Policy)” (Leeuw, 2004, 69). The results of reporting, however, are neither called nor could be classified as evaluation. Other foci, especially cost-effectiveness and cost-benefit evaluations were yielded by management reforms in the 90’s and are “supervised by the Directorate General for Budgets and Financial Control” (Leeuw, 2004, 71).

In 1996, steps for more systematic evaluations of policies were undertaken and a “decentralized model in which the operational Directorates General are responsible for establishing systematic evaluation procedures for the programs they are executing” was developed to improve evaluation practice (Leeuw, 2004, 71). As a

¹⁵ Based on Leeuw (2004).

result, each Directorate had to designate one evaluation official who is responsible for establishing an annual evaluation plan and for determining program to be evaluated. The Directorates' evaluation plans are assembled into the "Commission's Annual Evaluation Program". The [Directorate General for Budget](#) "coordinates evaluation activities and maintains an overview of the evaluation findings across the Commission services. It also provides [methodological guidance](#) and support, helps with [procurement of evaluation expertise](#) and maintains [evaluation networks](#) within and outside the Commission (see [website](#)). Unique features of the Evaluation Commission include a broad definition of the concept of evaluation and its direct link to budget:

Not only does it [evaluation] encompass ex post and midterm evaluation, but it also cover ex ante exercises... evaluation projects are to be framed so that they correspond to identifiable entities in the Community budget and to be timed so that results are available when they are relevant for budgetary decisions (Leeuw, 2004, 72).

Current Issues in European Evaluation¹⁶

Leeuw refers to different elements of current developments in Europe, including an increasing importance of civil societies, strengthened public management, and "polity". The most interesting aspect here is polity, especially because distinct political traditions in European nations need to be considered. Moreover, Leeuw refers to the valuing component of evaluation, which is especially inherent in political processes in which decisions are being made, values are chosen, and priorities set. However, the traditional practice of social research is challenged in their value-free doctrine.

¹⁶ Based on Leeuw (2004)

Most central topics for evaluation within Europe as stated by Leeuw are:

- The increasing importance of evaluation for civil society
- Evaluation for Parliaments (Do parliament decisions have effects?)
- Evaluation for public policy partnerships
- Decentralization of evaluation
- Potentials for evaluation of social programs from a non-managerial standpoint
- Evaluation of information and communication technology products, processes, and outcomes (web-based communication, training, the internet as knowledgebase)
- Auditing versus evaluation
- Evidenced-based evaluation
- Learning from evaluation
- Effective implementation and utilization of performance management systems in public management.

Overall, evaluation appears to be a vast growing market in Europe. However, as a discipline evaluation is still an “infant,” not only on the European level but internationally.

References

Leeuw, F.L. (2004). *Evaluation in Europe*. In: Stockmann, R. (Ed.). *Evaluationsforschung: Grundlagen und ausgewaehlte Forschungsfelder*. 2nd edition. Opladen: Leske + Budrich.

News from the community (2004). In: *Evaluation: The International Journal of Evaluation Theory, Research and Practice*, 10(3): 380-381.

Stern, Elliot (Ed.). *Evaluation: The International Journal of Evaluation Theory, Research and Practice*, 9(4)-10(3).

Stern, Elliot (2004). What shapes European evaluation: A personal reflection. In: *Evaluation: The International Journal of Evaluation Theory, Research and Practice*, 10(1): 7-15.

The European Evaluation Society (2004). *The European Evaluation Society website*. Available at: <http://www.europeanevaluation.org/>

Evaluation Activities in the United Kingdom

John S. Risley

General Summary of Activities

The UK Evaluation Society (UKES; www.evaluation.org.uk) was founded in 1994 and is composed of over 150 individual and corporate members. Most of these are individual members. UKES hosts an annual conference each year in December and jointly conducts seminars and conferences with other professional organizations. The society also sponsors an e-mail discussion list, Eval Chat, publishes a thrice yearly newsletter, *The Evaluator*, and produces *Evaluation: The International Journal of Theory, Research and Practice*.

UKES has five regional networks. Three of these networks, the Scottish Evaluation Network, the London Evaluation Network, and the North West Evaluation Network are established. The other two, the Cymru Evaluation Network (Wales) and the Midlands Evaluation Network are just forming.

The UKES website offers a host of information and links on evaluation topics, including:

- evaluation guidelines for good practice from different national evaluation associations,
- a list of postgraduate courses on evaluation taught throughout the U.K.,
- links to 21 national/regional evaluation society websites,
- an evaluation glossary (including an entry on “chatty bias”)

- a short but wide-ranging bibliography of evaluation books

Evaluation: The International Journal of Theory, Research and Practice

The journal *Evaluation* is published quarterly by Sage. Through the end of October it is available free online at evi.sagepub.com. I reviewed the last two years of *Evaluation* (the January 2003 issue through the July 2004 issue) and categorized each article according to Lori Wingate's adaptation of Michael Scriven's analogy for understanding disciplines. Wingate identified four categories of focus for journal articles—practice, methods, theory, metatheory—that I used below and one category—history—that I eliminated because no articles fit the description.

Practice issues dominated the 37 articles from the last two years (48.6 percent). The practice articles mainly dealt with the related issues of evaluation use and stakeholder participation. An article by Taut & Brauns (2003) examines social and psychological explanations for resistance to evaluation and offers strategies for overcoming evaluation resistance.

Many articles I categorized in the practice area concerned evaluation in different fields—healthcare, bidding for public services, welfare policy. These articles did not discuss different evaluation approaches or models, so I did not categorize them under theory.

Over one-fifth of the articles (21.6 percent) concerned theory. Three of these eight articles concerned theory-based evaluations—with two generally favorable and one generally unfavorable toward the approach—while other evaluation approaches addressed included qualitative, desk screening and implementation evaluation. Hearn, Lawler and Dowswell (2003) addressed the dominance of the positivist approach to most healthcare evaluation and argued that an inclusion of

“nonpositivist, qualitative, and process-oriented evaluation” would improve our understanding of health programs and policies.

I categorized six articles (16.2 percent) as methods articles. Interestingly, all of these articles focused on quantitative methods of data collection and analysis. Sverdrup (2003) discussed the use of time-series databases of complaints data to evaluate laws and regulations.

The metatheory category included five articles (13.5 percent) across 2003-2004. Virtanen and Uusikylä (2004) address the “paradigm crisis” in evaluation that stems from evaluators’ different assumptions about causality. These authors describe four alternative models (which they term ideal models) for evaluation considering: 1) how explicitly causality has been taken into account, and 2) how well the model enhances public-sector accountability.

The model reflecting both a strong link between causality and the evaluation design and an emphasis on public accountability is termed “transparent democracy”. “Scientific inquiry” signifies a strong link between the evaluation design and causality without an emphasis on accountability. The “explorative inquiry” model is characterized by a high degree of emphasis on accountability and a difficulty in distinguishing causal effects. Finally, an evaluation using the “symbolic evaluation” model serves a symbolic purpose rather than a “true pursuit of learning.” (89)

References

Hearn, J., Lawler, J., & Dowswell, G. (2003) Qualitative evaluations, combined methods and key challenges: General lessons from the qualitative evaluation of community intervention in stroke rehabilitation. *Evaluation*. 9: 30-54.

Sverdrup, S. (2003). Towards an evaluation of the effects of laws: Utilizing time-series data of complaints. *Evaluation*. 9: 325-339.

Taut, S., & Brauns, D. (2003). Resistance to evaluation: A psychological perspective. *Evaluation*. 9: 247-264.

Virtanen, P., & Uusikylä, P. (2004) Exploring the missing links between cause and effect: A conceptual framework for understanding micro–macro conversions in programme evaluation. *Evaluation*. 10: 77-91.

Evaluation in Eastern Europe and the Middle East

P. Cristian Gugiu

The state of evaluation in Europe is ever changing. In 1994, there was only one national evaluation society in Europe. Since then, the evaluation scene in Europe has blossomed to include 10 or 11 national societies—most of which are located in West Europe. Recent efforts, most notably by the European Evaluation Society, have been directed towards integrating all of these organizations under one umbrella. Through training and capacity building, the European Evaluation Society hopes to increase the number of engaged professionals in Europe, develop an academic support base, and strengthen the links to the policy community.

Compared to Western Europe, the state of the evaluation field in the Middle East appears to be less developed. In some way, the state of evaluation in the Middle East resembles that of Eastern Europe. Few of these countries have created national evaluation societies, taught evaluation in schools of higher learning, or published evaluation journals. The present paper intends to examine the current state of evaluation in Europe and the Middle East.

Evaluation Journals and Newsletters

East European Journals

Several representatives of the European Evaluation Society (EES) report that no one knows of any journal or newsletter publications in Eastern Europe.

Middle East Journals

According to Barbara Rosenstein, Ph. D., Chairperson of the Israeli Association for Program Evaluation (IAPE), the IAPE has published, to date, eight newsletters, in both Hebrew and English on evaluation.

Israeli Journal: *Studies in Educational Evaluation*

Studies in Educational Evaluation (SEE) is published in English. The majority of articles were not published by Israelis. Authors were dispersed throughout the world including the United States, United Kingdom, Australia, Spain, the Netherlands, and Germany.

A great many of the articles were purely research articles, a few of them described an evaluation case study, and a fair number of them discussed a specific methodology that could be used in evaluation.

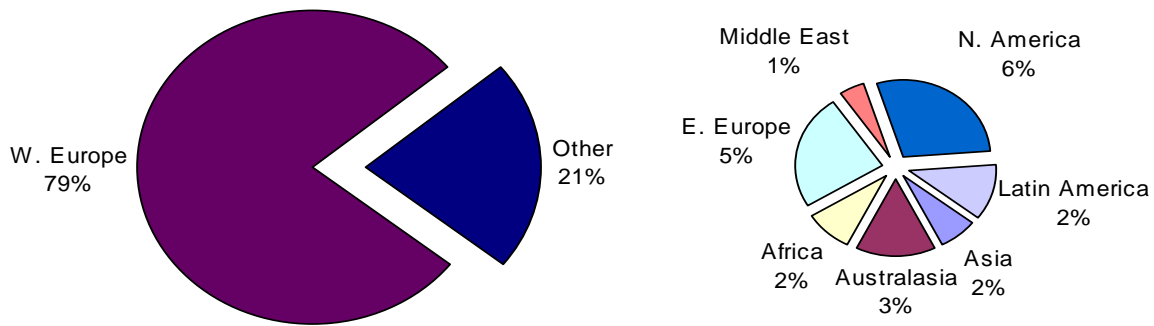
Evaluation Societies

European Evaluation Society (<http://www.europeanevaluation.org/>)

The primary goal of the European Evaluation Society (EES) is to promote theory, practice and utilization of high quality evaluation especially, but not exclusively, within the European countries. This goal is obtained by bringing together academics and practitioners from all over Europe and from any professional sector, thus creating a forum where all participants can benefit from the cooperation and bridge building. The society was founded in Hague in 1994. The first official board was elected in autumn 1995 and started its work in January 1996.

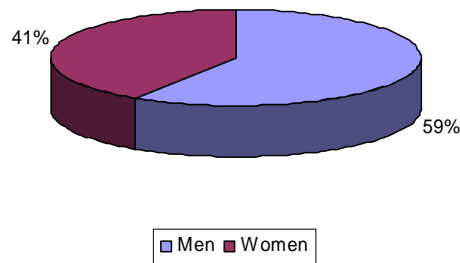
EES held its sixth conference on September 30 to October 2, 2003 in Berlin, Germany. The conference took place at the University of Applied Sciences and featured a total of 334 presenters from 36 countries.

Presenters at 2004 European Evaluation Society Conference by Region



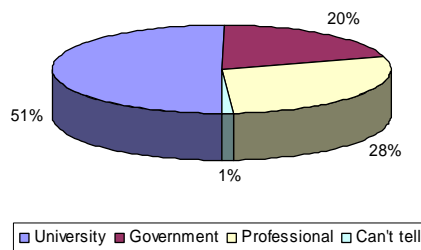
Over three-quarters of the presenters came from West European countries including Belgium (2.4 percent), Denmark (2.7 percent), Finland (4.2 percent), France (4.8 percent), Germany (9.3 percent), Greece (0.3 percent), Iceland (0.3 percent), Ireland (2.1 percent), Italy (15.9 percent), Netherlands (5.4 percent), Norway (1.5 percent), Portugal (2.7 percent), Spain (5.7 percent), Sweden (5.7 percent), Switzerland (4.5 percent), and the United Kingdom (9.0 percent). The remaining presenters included countries from Asia (Japan, 0.6 percent; Korea, 0.9 percent), Australasia (Australia, 2.4 percent; New Zealand, 0.6 percent), Africa (Angola, 0.3 percent; Guinea Bissau, 0.3 percent; Kenya, 0.3 percent; Nigeria, 0.9 percent), East Europe (Austria, 2.4 percent; Bosnia and Herzegovina, 0.3 percent; Czech Republic, 0.3 percent, Poland, 1.2 percent), the Middle East (Egypt, 0.3 percent; Israel, 0.3 percent, Palestine, 0.3 percent), North America (Canada, 0.9 percent; United States, 5.1 percent), and Latin America (Colombia, 0.6 percent; Mexico, 1.8 percent).

Presenters at 2004 European Evaluation Society Conference by Gender



There were slightly more male presenters than female presenters.¹⁷ However, this statistic was primarily influenced by the large number of West European presenters. Five of the seven other regions had an equal or greater number of female presenters than male presenters.

Presenters at 2004 European Evaluation Society Conference by Job Type

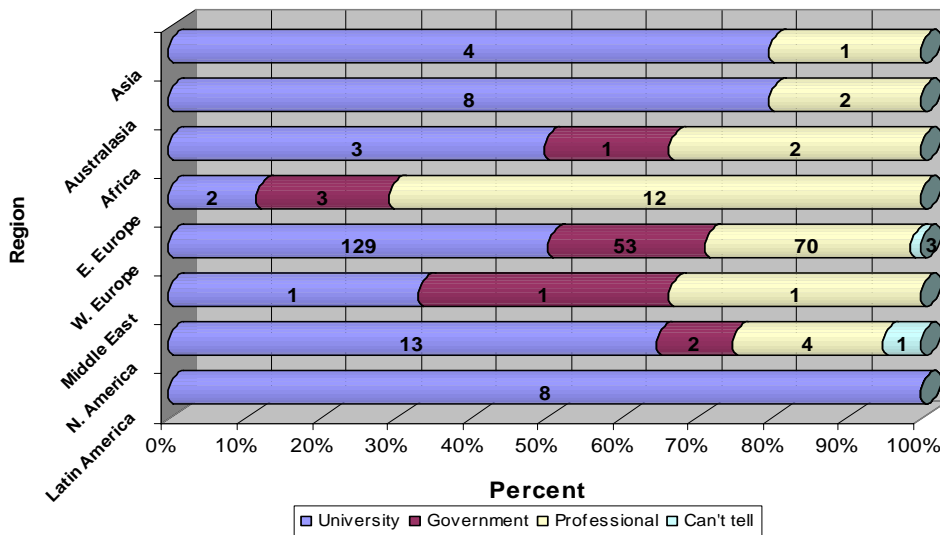


An examination of the type of jobs presenters worked in revealed that the majority of them worked for a university or college in their native country. The two next largest groups included people who worked in private industry or for the

¹⁷ The gender of the presenter was determined by (a) an inspection of their name, (b) an internet search for a website that mentioned the presenter's gender, or (c) an internet search for the gender most common for the presenter's first name. Consequently, there may be a small degree of inaccuracy in the categorization of gender.

government.¹⁸ It was interesting to note the differences in distribution of job type among the eight regions. For seven of the eight regions, presenters typically worked at a university. However, for Eastern Europe, the majority of presenters came from private industry. Possible explanations for this difference include the lack of university programs specializing in evaluation, the low number of professional evaluation associations, and historical factors such as socialism and recent wars.

Presenters at 2004 European Evaluation Society Conference by Region and Job Type



Polish Evaluation Society (<http://www.pte.org.pl/>)

The Polish Evaluation Society (PES) began in 2001 and set out to build an evaluation culture and popularize evaluation as a social and democratic process. To this end, it sought to (a) organize studies, courses and trainings; (b) conduct evaluation research; (c) exchange experiences with other societies, institutions and

¹⁸ The type of job the presenter worked in was determined by the e-mail domain they provided to EES or by an internet search on their name or the name of the organization listed as their employer with EES. For four presenters, however, it was not possible to determine their job type.

organizations; (d) organize meetings, seminars and conferences, (e) publish in the area of evaluation, and (f) provide consulting and advising services.

The Polish Evaluation Society has very strict rules as to the educational qualifications of its members. Most members are still strongly connected with the academic environment, either via didactic activity or scientific research (Warsaw University, Lublin Catholic University, B. Jański School of Administration and Enterprise). Members of PES continuously enrich their knowledge by taking part in trainings, seminars and conferences both in Poland and abroad (UK, The Netherlands, Denmark) and also by co-operation with other similar organizations in the world. (United Kingdom Evaluation Society, European Evaluation Society, IOCE—International Organization for Cooperation in Evaluation, PLS Ramboll Management—Denmark, Eureval-C3E—Centre for European Evaluation Expertise—France).

Members of PES are professional evaluators who also conduct marketing research and other research on social character. They have wide experience in the field of the evaluation which they gained in the process of conducting a variety of research for Polish and international organizations such as Polish Children and Youth Foundation, Public Interest Institute, government organizations such as European Integration Committee, the ministry of Education, service sector companies such as Daewoo, and EU institutions such as European Parliament and European Commission. Members of PES use different paradigms and research perspectives. A Rich variety of the activities and approaches is an advantage of this organization.

Romanian National Assessment and Examination Service

(<http://www.edu.ro/snee.htm>)

The National Assessment and Examination Service (NAES) was established in 1998 by the Romanian Government as the first national, independent body providing professional expertise in educational assessment and examinations in Romania. NAES is responsible for the design and implementation of the new educational evaluation system, namely for: (a) current assessment in pre-university education; (b) school leaving examinations (Capacitate exam and Bacculaureate exam); (c) national assessments at the end of educational cycles (now at the end of 4th grade); and (d) continuous teacher training in the field of assessment and examinations.

NAES is actively involved in national and international projects (e.g. the British Council, QUATRO Fontys—PTH Eindhoven) and maintains professional contacts with universities, research institutes, governmental and nongovernmental institutions and organizations in the field (e.g. CITO—The Netherlands, EDC—USA etc.). Their headquarters in Bucharest provides assessment technologies and facilities for development projects and studies in assessment and examination and their staff offers competence and expertise for cooperation to all those interested, in Romania and abroad.

Israeli Association for Program Evaluation (<http://www.iape.org.il/>)

The Israeli Association for Program Evaluation (IAPE) is a non-profit, professional organization comprised of academics, practitioners and users of program and project evaluation in a variety of fields—psychology, education, social services, health, business, and others. The goals of the organization include (a) increasing the use of program evaluation and its findings, (b) encouraging the development of

the theory of program evaluation, (c) advancing the essential recognition of program evaluation as a means of improving the effectiveness of social and educational interventions, (c) promoting the recognition of program evaluation as a profession, (d) serving the communities and the populations involved in program evaluation, (e) contributing to the influence of program evaluation on decision making, (f) supporting and influencing evaluation practice in Israel, and (g) creating and developing professional ties among evaluators and users of evaluation in Israel. To this end, the IAPE has sought to (a) organize conferences focusing on issues of concern to the evaluation community; (b) create an electronic and regular mail network that provide information about issues concerning evaluation in Israel and abroad; (c) establish connections with evaluation organizations throughout the world; (d) participate in the worldwide forum of evaluation associations, (e) circulate of a list of members to evaluation consumers in Israel; and (f) publish a newsletter containing articles, discussions, and events of interest to the evaluation community in Israel.

Evaluation in Latin America and the Caribbean: An Overview of Recent Developments

Thomaz Chianca and Brandon Youker¹⁹

In the past ten years, evaluation, as a professional field, has undergone significant development in several countries in Latin America and the Caribbean (LAC). Four considerations provide clear evidence of such development: (1) establishment of professional evaluation organizations; (2) intensified dissemination of ideas and use of professional evaluation in the three key societal sectors: government, private, and philanthropic; (3) increased number of evaluation-related publications; and (4) growing establishment of short-term and graduate-level training programs in evaluation.

Evaluation Organizations

The first professional evaluation organization that was formed in the region was the Central American Evaluation Association (ACE) in 1989 that has had its headquarters, since then, in Costa Rica. ACE's main objective is to foster the evaluation of programs and projects to improve efficacy and efficiency of the use of societal resources. The main activities developed by ACE include seminars, workshops, and courses to disseminate evaluation knowledge in Central America.

Only eight years later, in 2002 new evaluation professional organizations were established in LAC. Given their specific contexts, Brazil, Colombia and Peru opted to create networks of evaluators, less formal and more flexible organizations than

¹⁹ The authors want to thank Craig Russon, Marco Segone, Rogério Silva, Xinia Picado, and Emma Rotondo for their great comments and suggestions to improve an early version of this paper.

associations or societies. United Nations agencies such as UNICEF (United Nations Children's Fund), UNESCO (United Nations Educational, Scientific and Cultural Organization), and IFAD (International Fund for Agriculture Development), along with international and indigenous nonprofit organizations (e.g., foundations, institutes, etc.), played a decisive role in the creation of these three evaluation networks. Most membership is comprised of evaluators working in social and educational programs funded either by the government or by foundations.

[PREVAL](#) (Program for Strengthening the Regional Capacity for Evaluation of Rural Poverty Alleviation Projects in Latin America and the Caribbean)—a joint effort between IFAD and, from 1995-2000, the Inter-American Institute for Cooperation on Agriculture (IICA), and, from 2000-2007, with the Centro de Estudios para la Promoción del Desarrollo (Center of Studies for Development Promotion)—has played an strategic role in the region since 1995, contributing directly to the creation of the national evaluation networks in Peru and Colombia. In its first two phases (1995-2000 and 2000-2004), PREVAL focused its work on strengthening the evaluation capacity of IFAD projects to reduce rural poverty in the region. In its third phase (2004-2007), PREVAL will broaden its objectives to work more closely with governments, organizations offering technical assistance in monitoring and evaluation, as well as national evaluation and monitoring networks/associations in the region. PREVAL has established an important network of evaluators working with projects aimed at alleviating rural poverty, and has produced an important body of knowledge in this area published in Spanish. It is also important to recognize the key role played by the International Organization for Cooperation in Evaluation ([IOCE](#))—comprising all national and regional evaluation organizations around the globe—in fostering the establishment of

evaluation organizations in the region. IOCE held an important planning meeting in Barbados with all major international leaders and its inaugural assembly in Lima, Peru in 2003²⁰. The creation of both the Peruvian and the Colombian evaluation networks were strongly supported by IOCE.

In September 2003, representatives from the four existing evaluation organizations in the region got together in São Paulo, Brazil, to create the Latin American and Caribbean Evaluation Network ([RELAC](#)). The city of Lima, Peru, will host RELAC's first evaluation conference in October 20-23, 2004. The conference's theme ("Evaluation, Democracy, and Governability: Challenges for Latin America") and main sessions (e.g., democratic evaluation, methods for evaluating human rights programs, evaluation capacity building in social initiatives, monitoring, evaluation and systematizing as a political and social process to strengthen democracy in LAC, etc.) reflect RELAC's strong focus on promoting a social agenda for the region and having evaluation as a major tool. The conference will gather more than 100 evaluators from all over the region including representatives from eight countries that are trying to create their own national evaluation organizations: Argentina, Bolivia, Chile, Ecuador, Honduras, Nicaragua, Uruguay, and Venezuela.

There are at least four electronic discussion lists on evaluation in the LAC region: RELAC, PREVAL, the [Brazilian Evaluation Network](#), and the ILPES/CEPAL.

It is not over-optimistic to assume that very soon we will witness a significant increase in the number of evaluation professional organizations in LAC.

²⁰ Russon, C. & Love, A. (2003) The Inaugural Assembly of the International Organization for Cooperation in Evaluation: The Realization of a Utopian Dream. Occasional Paper #20. The Evaluation Center, Western Michigan University: Kalamazoo, Michigan.

Use of Professional Evaluation in Key Societal Sectors in LAC

There has been significant growth in the use of professional evaluations by the government, the nonprofit sector, and at least in the field of personnel evaluation, in large private businesses. In the government arena, initiatives related to national educational evaluation/assessment systems, innovations in government administration systems, and social development programs supported by international cooperation agencies are major factors influencing such growth.

In education, the establishment of evaluation mechanisms has been extensive from basic (K-12) to higher education in many countries within the region. In Brazil, for instance, the ministry of education has in place at least four major evaluation initiatives, applied countrywide, to assess the quality of education. In higher education there are two initiatives: (1) institutional evaluation (assessment of the general conditions of all higher education institutions in the country), (2) evaluation of undergraduate education (includes a national exam for senior students in each professional area and an accreditation strategy to renew or provide new licenses for universities and colleges). The other two are related to the basic education system: (3) national evaluation system for basic education (bi-annual assessment of the quality of K-12 schools, based on a national random sample), and (4) national exam for high school students (senior high school students have the option to take the exam that will serve as one of the criteria for acceptance in a university—similar to the ACT and SAT tests in the U.S.). Several countries also have official connections with regional as well as worldwide education assessment initiatives such as the Latin American Laboratory for Evaluating the Quality of Education, the International Program for Student Evaluation, and the World Education Indicators Program.

The idea of reducing the size of the state and making it more effective and efficient (state reform) has strongly influenced virtually all countries in the region. Such an idea brings along a strong push for the establishment of control systems on expenditures as well as for implementation of planned activities that usually involve monitoring and, to some extent, evaluation. Several countries have created structures, usually subordinated to the ministry of planning, that are in charge of dealing with monitoring and internal evaluation of governmental efforts. Examples of such structures are the System of Information, Evaluation, and Monitoring of Social Programs (Argentina); Secretary of Strategic Investments (Brazil); Intersectorial Committee for Modernizing Public Administration (Chile); National System for Management and Results Evaluation (Colombia); and the National Evaluation System (Costa Rica).

In the area of social development, virtually all programs supported by international cooperation agencies such as the World Bank, Inter-American Development Bank, World Health Organization, and United States Agency for International Development (USAID), are required to be evaluated both internally as well as by using external evaluators. These organizations have played a major role in introducing innovations in evaluation as well as advocating for the use of quality professional evaluations within government funded initiatives. Several examples of such evaluations are already publicly available from the agencies' websites (e.g., [USAID](#) and the World Bank-Operations Evaluation Department ([OED](#))). The Latin American Institute for Social and Economic Planning ([ILPES](#)), subordinated to the United Nations Economic Commission for Latin America and the Caribbean ([CEPAL](#)), has been an important reference in providing evaluation support to country-level government evaluators by offering supporting materials (publications); evaluation training; and networking opportunities for professionals

working in evaluations of governmental social-development programs in the region.

Initially influenced by international foundations investing in the region the fast—growing nonprofit sector in Latin America has increasingly used and advocated for the use of evaluation as a way of assuring quality of the programs they fund, and also to inform their decisions regarding strategic funding.

The [W.K. Kellogg Foundation](#) is one of the international foundations that have significantly invested in the development of evaluation in LAC. In 1995 and 1997, the foundation sponsored two groups of LAC evaluators (a total of approximately 40 professionals) in in-depth training programs in evaluation at [The Evaluation Center—Western Michigan University](#). Some of the participants of such training opportunities are assuming leadership roles in evaluation in the creation of evaluation organizations in their home countries.

Several foundations and institutes are commissioning and/or developing evaluations throughout the region. In Brazil, for instance, some of the nonprofit organizations that are very active in evaluation include: [Fundação Carlos Chagas](#), [Fundação Cesgranrio](#), [Instituto Ayrton Senna](#), [Fundação ABRINQ Foundation](#), [Instituto FONTE](#), [Fundação Roberto Marinho](#), and [Fundação IOCHPE](#).

Another interesting movement influencing the growth of evaluations in the third sector is the increasing number of private businesses investing in social initiatives, based on the idea of social responsibility. Such organizations have a different culture (focus on control and efficiency) from the nonprofit organizations investing in the sector, and have made an important push to support the establishment of monitoring and evaluation systems in the initiatives sponsored by them. In Brazil, the *Grupo de Institutos, Fundações e Empresas*—GIFE (Group of Institutes,

Foundations, and Enterprises) and the Ethos Institute are examples of two large organizations created to mobilize funds and to support companies and private foundations who invest in social, cultural and environmental projects of public interest, making such organizations partners in building a sustainable and fair society.

The extent of evaluation use in the private sector is not very public. It is evident that several corporations and other private business have made serious efforts to evaluate their products, projects and personnel. Reports on such efforts, however are not easily accessible and the evaluators working in this area have almost no contact with other evaluators working in the public and nonprofit sectors. No doubt more extensive exchange of experiences between these professionals has great potential to be beneficial to all, but some important barriers such as prejudices from both sides (e.g., ‘private sector only look at profits;’ ‘public and nonprofits are always inefficient’) need to be overcome before such approximation has any chance of succeeding.

Body of Original Publications in Evaluation

Though there are virtually no evaluation specific journals in LAC, there are several journals related to education, health, and social sciences with strong evaluation content. Some examples include:

- [*La Revista de Ciencias Sociales*](#) (Journal of Social Sciences—Costa Rica)
- [*Revista Ensaio – Avaliação e Políticas Públicas em Educação*](#) (Evaluation and Public Policy in Education—Brazil)
- [*Estudos em Avaliação Educacional*](#) (Educational Evaluation Studies—Brazil)
- [*Cadernos de Saúde Pública*](#) (Journal of Public Health—Brazil)

- [*Revista Avaliação Psicológica*](#) (Journal of Psychological Evaluation—Brazil)
- [*Revista da Rede de Avaliação Institucional*](#) (Journal of the Institutional Evaluation of Higher Education Network—Brazil)
- [*Cuadernos de Investigación de la Escuela de Gerencia Social*](#) (Journal of Inquiry of the School of Social Management—Venezuela)
- [*Revista del Instituto de Investigaciones en Ciencias de la Educación*](#) (Journal of the Education Science Investigation Institute—Argentina)
- [*Acción y Reflexión Educativa*](#) (Educative Action and Reflection—Panama)
- [*Planejamento e Políticas Públicas*](#) (Planning and Public Policy—Brazil)
- [*Revista de Administração Pública*](#) (Journal of Public Administration—Brazil)

The footnoted social science journals²¹ have regularly published the intellectual products of LAC evaluators.

It is critical to acknowledge the substantial collection of accessible evaluation publications such as books, manuals, newsletters, technical reports, etc. that are available in most Latin American countries. There are several websites such as the Latin American Institute for Social and Economic Planning (ILPES), and the Programme for Strengthening the Regional Capacity for Evaluation of Rural Poverty Alleviation Projects in Latin America and the Caribbean (PREVAL) that provide an extensive collection of evaluation publications in the field of evaluation throughout the region.

²¹ A very useful source to find out about Latin American journals in education and related areas is the *Centro de Recursos Documentales Informático—CREDI* (Center for Digital Documental Resources), sponsored by the Organization of Ibero-American States.

There are two excellent annotated bibliographies that provide published reference materials that address several aspects of evaluation in LAC. The first publication, *The Annotated Bibliography of International Programme Evaluation*, edited by Russon & Russon²² has a chapter by Antoinette B. Brown and Ada Ocampo, on Latin America and the Caribbean. The reviewed documents (books, manuals, journal articles, etc.) are grouped into three sections: (1) manuals and guides, (2) participation, and (3) case studies and evaluation reports. The second publication, *Annotated Bibliography on Project Evaluation* by Viñas²³ which includes a broad range of documents in the review, is divided into fourteen categories: (1) definition of basic concepts; (2) design, implementation and evaluation; (3) environment impact; (4) evaluation approaches; (5) evaluation design; (6) gender; (7) indicators; (8) methods and instruments for information collection and analysis; (9) monitor and evaluation systems; (10) organizational development; (11) participation by beneficiaries; (12) producing reports and presenting conclusions; (13) types of evaluation; and (14) use of evaluations.

Capacity Building in Evaluation

There are a few LAC universities and training institutions that offer masters level programs, specifically in evaluation. At the masters' level, there are at least five universities offering such program:

- [Professional Masters in Evaluation of Social Programs and Projects.](#)
Universidad de Costa Rica. San José, Costa Rica.

²² Russon C. & K. Russon. Eds. (2000) *The Annotated Bibliography of International Programme Evaluation*. Norwel: Kluwer, Academic Publishers

²³ Viñas, V.E. (1997). *Annotated Bibliography on Project Evaluation*. San José, Costa Rica : International Fund for Agricultural Development (IFAD), Inter-American Institute for Cooperation on Agriculture (IICA), Programme for Strengthening the Regional Capacity for Evaluation of Rural Poverty Alleviation Projects in Latin America and the Caribbean (PREVAL).

- [Masters in Socio-Economic Evaluation of Investment Projects](#). Universidad Panamericana. Mexico City, Mexico.
- [Masters of Science in Project Management and Evaluation](#). University of the West Indies—Cave Hill Campus. Bridgetown, Barbados.
- [Masters in Project Evaluation](#). Universidad del CEMA, Buenos Aires, Argentina.
- [Masters in Social Projects Evaluation](#). Universidad Autónoma de Guadalajara. Guadalajara, Jalisco, México.

At the certification level there are quite a few programs offered in different countries including:

- [Course on Evaluation of Social Programs and Projects](#). Centro de Empreendedorismo Social e Administração em Terceiro Setor—CEATS (Center of Social Entrepreneurship and Administration for the Third Sector—Brazil) and the FONTE Institute.
- [Diploma in Evaluation of Projects](#). Universidad de Concepción. Concepción, Chile.
- [Diploma in Evaluation of Social Projects](#). Pontificia Universidad Católica de Chile. Santiago, Chile.
- [Diploma in Planning and Evaluation of Projects](#). Universidad de Chile. Santiago, Chile.
- [Diploma in Planning and Evaluation of Socioeconomic Projects](#). Centro de Análisis y Evaluación de Política Pública—CAEP—Monterrey, Mexico. (Center of Analysis and Evaluation of Public Policy)

- [International Certificate of Project Planning, Evaluation, and Management—Inter-American Development Bank](#). Centro de Investigaciones Territoriales del Ecuador—CITE. (Territorial Investigation Center of Ecuador).
- [Post Graduate in Formulation and Evaluation of Projects](#). Universidad Americana. Managua, Nicaragua.

There are also several short-term evaluation training courses facilitated by different organizations within the region. Some of the best sources to identify such training opportunities include: (a) Nota Informativa del ILPES sobre Evaluación de Proyectos y Programas ([ILPES Informative Note on Program and Project Evaluation](#)); (b) [PREVAL](#); and [FONTE Institute](#). The following is a sample of the recently offered short-term courses in some LAC countries:

- X International Course on [Planning and Evaluating Public Investment Projects](#). Offered by [CEPAL/ILPES](#). Sept 27 to Oct 22, Santiago, Chile.
- [Internet-based course on Planning and Evaluation of Agricultural and Agri-Industrial Projects](#). Offered by REDCAPA and [Austral University of Chile](#). Sep 1 to Nov 30 2004.
- [International Course on Logic Model, monitoring and Evaluation](#). Offered by ILPES/CEPAL and the Spanish Cooperation Agency (AECI). Jun 21 to Jul 2, 2004. Cartagena de Indias, Colombia.
- [International Course on Use of Socio-Economic Indicators for the Evaluation of Impact of Poverty Reduction Programs](#). Offered by ILPES/CEPAL and the Spanish Cooperation Agency (AECI). May 3—14. Santa Cruz de la Sierra, Bolivia.

- [Utilization-Focused Evaluation by Michael Quinn Patton](#). Sponsored by the Brazilian Evaluation Network, UNICEF-Brazil, and FONTE Institute. Salvador, Bahia—Brazil. March 2004.
- [Collaborative Evaluation by Rita O’Sullivan](#). Sponsored by the Brazilian Evaluation Network, UNICEF-Brazil, and FONTE Institute. São Paulo, SP—Brazil. September 2003.

Final Comments

The report makes no claims to be comprehensive and does lack significant information, mainly about the state-of-art of the evaluation field in the Caribbean countries. We hope to fulfill the existing information gaps in the next revisions of and addenda to this paper.

It does, however provide unquestionable evidence of the impressive advances the whole region has made in the evaluation field in the recent past. Evidently, even though not to the same degree in each country, it is reasonable to say that basic conditions have been established to make such advances even more comprehensive and effective in the future.

The current efforts to establish national and regional evaluation organizations, the growing number of quality publications in both Spanish and Portuguese on evaluation, the increasing use of professional evaluation by different organizations in all societal sectors and the broad recognition of evaluation as important for improving society are some of the factors influencing such advances. One major challenge still to be faced in order to have evaluation in a better position as a recognized professional field is the creation of more formal graduate-level training for evaluators in a wider range of countries.

This paper is a work in progress that will be modified and/or improved as we gain new information. If you would like to provide additional information or point out any errors or misunderstanding in the text, please do not hesitate to contact Thomaz Chianca (thomaz.chianca@wmich.edu) or Brandon Youker (brandon.w.youker@wmich.edu).

Global Review: Publications

What's Happening in AJE (2003-2004)

Lori A. Wingate



“The *American Journal of Evaluation (AJE)* publishes original papers about the methods, theory, practice, and findings of evaluation. The general goal of *AJE* is to present the best work in and about evaluation, in order to improve the knowledge base and practice of its readers. Because the field of evaluation is diverse, with different intellectual traditions, approaches to practice, and domains of application, the papers published in *AJE* will reflect this diversity. Nevertheless, preference is given to papers that are likely to be of interest to a wide range of evaluators and that are written to be accessible to most readers.”

AJE Web site: <http://www.sciencedirect.com/science/journal/10982140>

The *American Journal of Evaluation (AJE)* is the flagship publication of the American Evaluation Association, the world’s largest organization for professional evaluators. As such, *AJE* plays an important role in defining the relatively young discipline of evaluation and influencing the work and thought of many practicing evaluators, many of whom have never had any formal training in evaluation.

In the *Evaluation Thesaurus* Scriven (1991) provides an analogy for understanding how various disciplines, and the levels of activities within those disciplines, relate to one another. In this analogy, he suggests we think of disciplines as estates in the “country of the mind.” He explains, “The houses on an estate have a ground floor

representing applied work; a floor above that which is devoted to developing instruments, methods, and techniques, and a top floor where the theoretical work is done. Up in the attic, out of sight for most of the time, is the den of metatheory” (pp. 13-14).

I used this framework to analyze the contents of *AJE* articles (from Spring 2003 through the present issue, which is Autumn 2004). I categorized the articles (65 in all) according to whether they focused on practice, methods, theory, or metatheory, and one additional category—history. The breakdown is shown Figure 1. Below I describe these categories and summarize the articles associated with those categories, highlighting what I believe to be the most important articles.

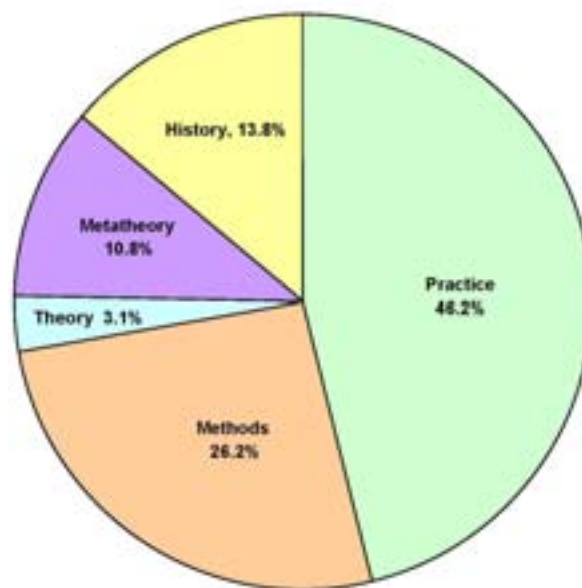


Figure 1. Focus of 2003-2004 *AJE* articles

Practice

“Practice” articles deal with ways of working with stakeholders and clients, ethical challenges, evaluation contexts, managerial aspects of evaluation, and evaluation

use. Almost half (46.2 percent) of the articles published in *AJE* since 2003 focus primarily on such practical aspects of the evaluation profession.

Eight of the 30 articles in the Practice category are part of *AJE*'s "Ethical Challenges" series, in which the section editor, Michael Morris, presents a brief scenario in which an evaluator faces an ethical challenge. In response, two commentators, in two separate articles, analyze the nature of the ethical problem and describe what they believe to be the appropriate response by the evaluator in the scenario, especially in light of the American Evaluation Association's Guiding Principles for Evaluators and *The Program Evaluation Standards* by the Joint Committee on Standards for Educational Evaluation (1994).

Seven articles in the Practice category focus on evaluation use, with five of these appearing as a series in a single issue. These use-oriented articles explore the many facets of evaluation utilization. They provide exemplars of useful evaluation, identify factors that promote and impede evaluation use, and weigh the sometimes conflicting values of evaluation utility and scientific rigor. Evaluation is an inherently applied discipline—intended to be used—but it is something that many people shy away from, or downright fear. Given these conflicting conditions, it is no surprise that many evaluators are interested in improving evaluation utilization. I categorized two other use-oriented articles (by Henry [2003] and Henry and Mark [2003]) in the "Metatheory" category, because they go beyond the practical issues related to use and venture into a theory about evaluation influence, which I discuss in greater detail in that section.

The remaining articles that I included in the Practice category address a variety of issues that have emerged out of the experience of real people engaged in the practice of evaluation—for example, how certain evaluation contexts present

particular challenges or opportunities, the managerial aspects of evaluation (e.g., contracts, resource constraints), and how to communicate effectively with stakeholders. One article that stands as particularly useful is by Bamberger, Rugh, Church, and Fort (2004). They offer several practical solutions for common problems that evaluators face when working under severe constraints. Their recommendations are most relevant for impact evaluations in which the use of control groups, baseline data, and random sampling would be ideal but not feasible due to timing, resources, and/or availability of data.

Articles focusing on practice offer readers insights into the real world of evaluation, where textbook methods and theory meet politics, red tape, ethical dilemmas, and stakeholders and clients who may or may not be interested in participating in evaluation or using its results. These types of articles provide readers with opportunities to learn from others' mistakes and successes in the uncertain world of evaluation practice. They offer students and established evaluators insights into how evaluation happens in the real world—lessons often not provided in textbook expositions on theory and methods.

Methods

“Methods” articles focus on a particular approach to data gathering and/or analysis. Seventeen of the 65 AJE articles (26.2 percent) deal primarily with methods. Such articles typically describe an innovative method or a modification of an existing method. These articles were equally divided between qualitative (8) and quantitative methods (8), with one article featuring a blend of both.

The qualitative methods covered by the articles include concept mapping, site visits, qualitative phone interviewing, the “most significant change” technique, methods for reconstructing and analyzing program theories, the Delphi technique,

methods of values inquiry, and methods for formatively evaluating educational technology.

Four of the seven articles on quantitative methods discussed methods used to overcome problems associated with randomized controlled trials, including the use of longitudinal data on program outcomes to estimate program effects, two different methods for analyzing impacts on beneficiary subgroups, and an approach for blending experimental and quasi-experimental methods. Other articles focused on the development of intervention-specific measures, techniques for assessing the quality of program implementation, and the use of post-plus retrospective pretests for measuring change.

The one article that focused on a method that incorporates the use of both qualitative and quantitative data described the development and use of a rubric for evaluating collaboration.

Methods articles highlight innovative and cutting edge approaches to evaluation data gathering and analysis. Journal articles and professional conferences are probably the most important ways practicing evaluators learn about new and useful methods. The methods are typically described in the context of a particular evaluation, which may help readers to discern the method's applicability to the areas in which they work.

Theory

“Theory” articles center on the use of a particular evaluation approach or model. Evaluation theory was the focus of just two articles (3.1 percent) published in *AJE* since 2003. One provides an in-depth look at an evaluation that blended two approaches to evaluation—theory-driven and utilization-focused. The other

theory-focused article offers an adaptation of Michael Fetterman's empowerment evaluation model (by Carolyn Sullins, a Senior Research Associate at The Evaluation Center). Both deal with practical applications of theory, but the emphasis is on the applied theory, rather than the specific methods or findings. (There are other *AJE* articles that feature the use of a particular theory, but the thrust of these articles is on practice, not theory.)

No articles in the timeframe examined (2003-2004) focused exclusively on an evaluation theory/model/approach in its pure form. As Christie and Alkin (2003) remark in their article about using a theory-driven approach in a user-oriented evaluation, "theories are rarely, if ever, flawlessly translated into practice" (p. 381). Given this, "in order to develop a deeper understanding of how evaluation theories are best applied in practice, it is important to describe cases where evaluation theories have been used in practice" (p. 381). That, indeed, is the nature of these two Theory articles.

It was somewhat surprising to me that only 2 articles out of 65 focused purely on evaluation theory. It is an important area of inquiry would seem to warrant more space in *AJE*.

Metaethy

Scriven (1991) defines metatheory as a "'theory' about the nature of a field of inquiry, engineering, or craft. It deals with matters such as the definition of the field's boundaries, its differences from neighboring fields or disciplines, the reason why certain methods work well for it and others are inappropriate.....it is the self-concept of the discipline" (p. 232). Seven (10.8 percent) articles in *AJE* directly discuss or contribute to the evaluation discipline's self-concept, or metatheory.

Two of the Metaethy articles focus on use. Both articles address the issue of evaluation use not simply as a practical matter, but as a sort of lens through we can view the role of evaluation discipline. Henry and Mark (2003) address the shortcomings in the existing literature on evaluation use, particularly the “inattention to the intrapersonal, interpersonal, and society change processes through which evaluation findings and process may translate into steps toward social betterment” (p. 294). They urge evaluators to look beyond immediate use of findings as the primary utilitarian purpose of evaluation, and instead focus on social betterment as the ultimate desired outcome. They outline a general theory of evaluation influence. Similarly, Henry (2003) offers several examples of evaluations that have been influential and offers a “clearer picture of what evaluation should look like in the future” (p. 515).

Two articles that I placed in the Metatheory category have to do with evaluation education. These articles do not directly contribute to the metatheory of evaluation in terms of content, but the way in which and what students and others learn about evaluation—its practice, methods, and theory, and history—is probably the primary vehicle by which evaluation metatheory develops. One article provides an overview of a one-year evaluation course that employs a mentoring approach. The other, by Christie and Rose (2003), provides an account of an informal discussion group. This group, facilitated by Marvin Alkin at UCLA, includes both students and faculty members who meet every other week to discuss an article in a recent issue of the *American Journal of Evaluation*. In addition to providing a venue in which members can share and test ideas, relate theory to practice, refine thinking, and hypothesize (among other things), the group also promotes socialization into the field. Such groups, write Christie and Rose, “are an alternative mechanism for encouraging the kinds of dynamic dialogue that facilitates the advancement of both

theoretical and practical notions of a field, such as evaluation, that is so dependent up on the interchange of ideas” (p. 238).

In his article on the Joint Committee evaluation standards, Stufflebeam (2004) addresses the applicability of the Program, Personnel, and Student Evaluation Standards to other cultural contexts. These are essentially standards for evaluation practice, but they have played an important role in shaping the field’s self-concept. At issue is whether the Standards can or should be transferred to other cultural contexts, and Stufflebeam argues they should not. The widespread interest in doing so is a testament to the Standard’s relevance to the discipline’s self-concept.

Stake (2004) addresses the role of advocacy in evaluation. He outlines six types of advocacies found to some extent in most evaluations. Roughly, they are advocacy for (1) a program’s success, (2) the evaluation discipline, (3) rationality, (4) evaluation use, (5) the alleviation of underprivilege, and (6) democracy. He argues that these advocacies shape evaluators’ interpretations of findings, which are “are enriched by personal experience” (p. 107). He concludes the article by stating, “Comprehensive, idiosyncratic interpretations are small steps toward saving the world” (p. 107).

The final article dealing with metatheory views evaluation itself as an important object of inquiry and provides a framework for researching the processes, contexts, obstacles, and knowledge claims in public sector evaluations. In this article, Segerholm (2003) reviews existing research on evaluation and concludes that it is “fairly scarce” and tends to focus on particular aspects of the evaluation cycle (i.e., initiation, implementation, results, and utilization) (p. 356). Likewise, she notes, metaevaluations (evaluations of evaluations) usually focus on a single evaluation. Segerholm argues that we need more research on evaluation to “gain knowledge

and a more thorough understanding of the phenomenon and practice of evaluation *in general*” (p. 357).

History

In addition to Scriven’s disciplinary categories of practice, methods, theory, and metatheory, I added History as a fifth category. I found this to be necessary because articles that focus on the development of the evaluation field cut across all the other categories, dealing with evaluation practice, methods, and theory, as well as influential personalities in the field; groundbreaking evaluations; important books; and key agencies, organizations, and educational institutions. These articles also contribute to the development and refinement of evaluation’s metatheory, since they help interpret and shape the field’s self-concept. Nine (13.8 percent) of the *AJE* articles since 2003 delve into the history of evaluation.

Most of the articles included in this category (6 out of 9) are oral history accounts of evaluation leaders collected for The Oral History Project—an effort by Robin Miller, Jean King, Melvin Mark, and Stacey Stockdill to document the “genealogy” of program evaluation. These oral history articles have featured interviews with Lois-ellin Datta and William Shadish, as well as brief articles by Laura Leviton, Roger Straw, Charles Reichardt, and Melvin Mark, who reflect on their experience in the Methodology and Program Evaluation program in the Psychology Department at Northwestern. Additional evaluation leaders will be featured in future issues, leading to the compilation of a rich and detailed history of the development of the evaluation field.

Margaret Mead’s evaluation of the 1947 Salzburg Seminar on American Civilization is the focus of the three other History articles.

References

- Bamberger, M., Rugh, J., Church, M., & Fort, L. (2004). Shoestring evaluation: Designing impact evaluations under budget, time and data constraints. *American Journal of Evaluation*, 25(1), 5-37.
- Christie, C. A., & Alkin, M. C. (2003). The user-oriented evaluator's role in formulating a program theory: Using a theory-driven approach. *American Journal of Evaluation*, 24(3), 373-385.
- Christie, C. A., & Rose, M. (2003). Learning about evaluation through dialogue: Lessons from an informal discussion group. *American Journal of Evaluation*, 24(2), 235-243.
- Henry, G. T. (2003) Influential evaluations. *American Journal of Evaluation*, 24(4), 515-524.
- Henry, G. T., & Mark. M. M. (2003) Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation*, 24(3), 293-314.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards (2nd ed)*. Thousand Oaks, CA: Sage Publications.
- Scriven, M. (1991). *Evaluation thesaurus*. Newbury Park, CA: Sage Publications.
- Segerholm, C. (2003). Researching evaluation in national (state) politics and administration: A critical approach. *American Journal of Evaluation*, 24(3), 353-372.

Stake, B. (2004). How far dare an evaluator go toward saving the world? *American Journal of Evaluation*, 25(1), 203-107.

Stufflebeam, D. L. (2004). A Note on the Purposes, Development, and Applicability of the Joint Committee Evaluation Standards. *American Journal of Evaluation*, 25(1), 99-102.

**Evaluation: The International Journal of Theory, Research and Practice
(2003-2004)**

Daniela C. Schröter



Evaluation is a quarterly, European-based journal that in addition to interdisciplinary and multidisciplinary peer-reviewed articles occasionally provides *Special Issues*, *Visits to the World of Practice*, *News from the Community*, *Book Reviews*, *Speeches and Addresses*, and *Debates*, *Notes*, and *Queries*. The publisher highlights the journal's interdisciplinarity (across fields of research), interconnectivity (across fields of practice), and internationality (across countries and continents).

The journal appears quarterly and can be purchased for £51 per individual (approximately \$92) and £372 for institutions (approximately \$671) per annual subscription.

In the following, the most recent four issues of *Evaluation*—Volumes 9(4), 2003 to 10(3), 2004—are analyzed in terms of theory-specific, research-specific, and practice-related articles published and themes from the *News from the Community* and *Speeches and Addresses*.

Articles reflecting theory-orientation illuminate general or abstract principles of the nature or practice of evaluation. There will not be a distinction here between theory and meta-theory.

Contributions categorized as research are those that discuss mainly methodological issues or evaluation tools and how they are employed in specific contexts. It is believed that this methodological approach best summarizes the core of research, that is, “disciplined investigation” (Scriven 1991, p. 312).

Practice-reports are those that describe a specific evaluation project. While some of those articles may attempt to build evaluation theory or contribute to methodology, specific evaluation projects are the core of discussion.

Overall, there were 22 articles from contributors of 11 countries published within the last year. Of these contributions, 6 are theory-oriented, 8 are research/methodology-oriented, and 8 are accounts of practice (see Figure 1).

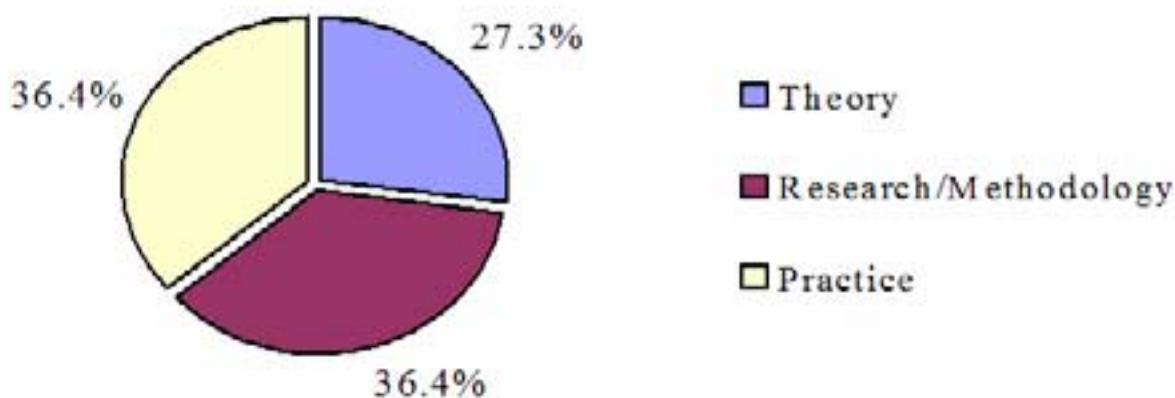


Figure 1: Distribution of articles

Theory

All six theoretical articles were published in a Special Issue of *Evaluation* (Vol. 10(1)), which contains a selection of papers presented at the European Evaluation Society’s 5th biennial conference entitled *Learning, Theory, Evidence: Three Movements in Contemporary Evaluation*. The first of these articles is Stern’s

(2004) personal reflection on “What Shapes European Evaluation?” Stern generalizes features of the European culture and infers tasks for policy and practice of the European community of evaluators. The second contribution, “Theory-Based Evaluation and Learning: Possibilities and Challenges” (Van der Knaap, 2004) attempts to overcome the split between positivist and constructivist paradigms through theory-driven evaluation in public policy which also considers the complexity of society thus providing learning opportunities to governments. Mark and Henry (2004) present a framework of evaluation influence and the underlying mechanisms to be considered. Stame (2004) revises theory-driven evaluation and provides ideas for coping with complexities in hierarchical governmental programs. Virtanen and Uusikylä (2004) try to re-conceptualize the traditional logic of cause and effect under consideration of “ontological premises, conceptions of causality, and relationships to rational theories of action of different program evaluation paradigms (p.77).” The last article in this category is Rieper’s (2004) discussion of Popper’s significance to and influence in evaluation by illustrating his impact on two opposite figures in evaluation, Campbell and Lindblom.

Research/Methodology

There are eight articles within the last year across the four issues that contain research and methodology-oriented contributions to *Evaluation*. Seven articles provide frameworks, approaches, or tools for evaluation in specific contexts such as public policy, social programs, training, and aid (Ling, 2003; Thurston & Potvin, 2003; Brousselle, 2004; Crawford et al, 2004; Bulgarelli & Gori, 2004; Johnson et al, 2004; Saunders & Heflinger, 2004). Davies (2004) provides a methodological approach to representing theories of organizational change to evaluate impacts of international aid agencies. This article is especially interesting

because it provides other approaches to depicting program theory in addition to the traditional logic model. Davies (2004) argues that some theories may have to be represented in their complexity through networks rather than directional simplifications. While his contribution focuses on international aid agencies, it may be useful to other contexts; programs of all types are initiated through organizations, none of which are simple, unidirectional entities. Overall, the research/methodological contributions embrace the interdisciplinary nature of the journal.

Practice

The practice-oriented contributions to the last four issues of *Evaluation* contain specific examples of evaluation in policy, health, environment, welfare, and crime prevention evaluation among others. These are of specific use for those readers who have an interest or are active in the specific areas, but also provide ideas for evaluation work in other contexts. As such these contributions reflect the interconnectivity feature of the journal.

News from the Community

Each issue within the last year included *News from the Community* containing special evaluation-specific developments in individual countries (e.g., developments in France in autumn 2003, in Vol. 9,4), announcements of conferences and awards to be distributed, calls for papers (e.g., to the 6th biennial EES conference, Vol. 10,1), references to specific research (e.g., the IOCE initiative to conduct a survey, Vol. 10, 2), recommendations for and announcements of training, seminars, or workshops; and special honors in evaluation. Overall, *News from the Community* is not limited to the European community but embraces the evaluation community internationally.

Speeches and Addresses

This component of the journal includes addresses and speeches that were presented at evaluation events and are to be disseminated to a wider audience—the readers of *Evaluation* (see [Sage](#)). Within the last year there have been three speeches and addresses. Issue 9(4) includes two contributions: Pawson (2003) and Stek (2003). Pawson’s address is based on a presentation given at the 5th biennial EES conference and discusses tasks for developing theory in evaluation, which include: (1) developing a theory-driven approach to systematic review of existing evaluations, (2) improving the transferability of findings in program evaluation to other programs, and (3) elucidating program complexity. Stek’s (2003) contribution “Evaluation at the World Bank and Implications for Bilateral Donors” was presented at the “25th anniversary of the IOB, the evaluation body of the Netherlands Ministry of Foreign Affairs and the Ministry for Development Cooperation” (p.491).

The next contribution followed in issue 10(3), a “keynote address to the 9th UK Evaluation Society Annual Conference” (p. 366), in which Sanderson (2004) presented “perspectives on rationality” while arguing for “a more ‘practical’ conception of rationality as a basis for an appropriate institutional framework for ‘getting evidence into practice,’ serving the emphasis the UK government puts on “evidence-based policy making (p.366).”

References

Stern, Elliot (Ed.), (2003). *Evaluation: The International Journal of Evaluation Theory, Research and Practice*, 9(4)-10(3).

Stern, Elliot (2004). What shapes European evaluation: A personal reflection. In: Evaluation: The International Journal of Evaluation Theory, Research and Practice, 10(1): 7-15.

Appendix I Summary of classifications

Theory

Vol. 10(1)

Stern, E.: *What shapes European Evaluation? A Personal Reflection*

Van der Knaap, P.: *Theory-Based Evaluation and Learning: Possibilities and Challenges*

Mark, M.M. and Henry, G.T.: *The Mechanisms and Outcomes of Evaluation Influence*

Stame, N.: *Theory-Based Evaluation and Types of Complexity*

Virtanen, P. and Uusikylä, P.: *Exploring the Missing Links between Cause and Effect: A Conceptual Framework for Understanding Micro–Macro Conversions in Programme Evaluation*

Rieper, O.: *Exploring Popper's Relevance for the Evaluation Community*

Research and Methodology

Vol. 9(4)

Ling, T.: *Ex Ante Evaluation and the Changing Public Audit Function: The Scenario Planning Approach*

Thurstun, W.E. and Potvin, L.: *Evaluability Assessment: A Tool for Incorporating Evaluation in Social Change Programmes*

Vol. 10(1)

Davies, R.: *Scale, Complexity and the Representation of Theories of Change*

Vol. 10(2)

Brousselle, A.: *What Counts is not Falling... but Landing: Strategic Analysis: An Adapted Model for Implementation Evaluation*

Crawford, P.; Perryman, J. and Petocz, P.: *Synthetic Indices: A Method for Evaluating Aid Project Effectiveness*

Bulgarelli, A. and Gori, E.: *Information Systems for the Evaluation of the Effectiveness and Efficiency of Vocational Training Programmes*

Vol. 10(3)

Johnsen, A.; Robertsen, K.; and Åsland, D.Y.: *Contracting Out Municipal Auditing: Conceptual Framework and Assessment*

Saunders, R.C. and Heflinger, C.A.: *Integrating Data from Multiple Public Sources: Opportunities and Challenges for Evaluators*

Practice

Vol. 9(4)

Bustelo, M: *Evaluation of Gender Mainstreaming: Ideas from a Meta-Evaluation Study*

Alexander, H.: *Health-Service Evaluations: Should We Expect the Results to Change Practice?*

Mickwitz, P.: *A Framework for Evaluating Environmental Policy Instruments: Context and Key Concepts*

Vol. 10(2)

Skinner, D.: *Primary and Secondary Barriers to the Evaluation of Change: Evidence from Two Public Sector Organizations*

Ashworth, K.; Cebulla, A.; Greenberg, D. and Walker, R.: *Meta-Evaluation: Discovering What Works Best in Welfare Provision*

Vol. 10(3)

Marra, M.: *The Contribution of Evaluation to Socialization and Externalization of Tacit Knowledge: The Case of the World Bank*

Halliday, J.; Asthana, S.M.N. and Richardson, S: *Evaluating Partnership: The Role of Formal Assessment Tools*

Johnson, S.D.; Bowers, K.J.; Jordan, P.; Mallender, J.; Davidson, N. and Hirschfield, A.F.G.: *Evaluating Crime Prevention Scheme Success: Estimating 'Outcomes' or How Many Crimes Were Prevented*

The Japanese Journal of Evaluation Studies

John S. Risley

General Summary of Activities

The Japanese Evaluation Society (JES) —available at www.idcj.or.jp/JES/—began publishing *The Japanese Journal of Evaluation Studies (JJES)* in 2001. The journal is published twice each year. Complete copies of each issue of *JJES* can be found at the JES web site. The complete text of most articles is in Japanese. Some articles (about 15% of all articles since 2001) are printed in English. All other articles are accompanied by an abstract in English. Each issue of *JJES* also contains an English-language index.

JJES is dominated by articles written by authors from Japanese organizations. Nearly 90% of the 64 articles published since the journal's founding are by Japanese authors. The large majority of these articles concern examples of evaluation practice, methodology, and theory as they relate to Japan.

I reviewed all issues of *JJES* and categorized each article according to Lori Wingate's (in this issue) adaptation of Michael Scriven's analogy for understanding disciplines. Wingate identified four categories of focus for journal articles—practice, methods, theory, metatheory—that I used below and one category—history—that I eliminated because no articles fit the description.

Practice issues dominated the 64 articles from the journal's first four years (50%). The practice articles dealt with a range of evaluation issues concerning Japanese government activities. Local and national government attempts to institute systems

for evaluating their policies were a consistent theme. The evaluation of development assistance programs run through the Japanese International Cooperation Agency (JICA) was also a popular topic.

One out of four articles concerned theory. Typically, an author would highlight a model of evaluation (utilization-focused, empowerment, etc.) and discuss how the model was being used in Japan. Another area that was prevalent under this category involved articles concerning gender equity, gender mainstreaming, or gender policy evaluation. These articles emphasized the importance of addressing gender differences when evaluating a policy or program.

I categorized fifteen articles (23.5%) as methods articles. Many of these articles addressed methodological issues around developing indicators for performance measurement. Other articles focused on methods that were helpful in evaluating gender equality in programs and policies.

The metatheory category included only one article. This article²⁴, one of the few English-language articles, concerned the preparation of professional evaluators. The author, James W. Altschud of Ohio State University, argues that evaluation training is more than just learning a set of methodological tools. He reviews some research on the state of evaluation training and raises several issues that need to be addressed when thinking about how we train evaluators.

²⁴ Altschud, J.W. (2002). The preparation of professional evaluators: Past tense and future perfect. *The Japanese Journal of Evaluation Studies*. 2(1): 1-10.

Journal of Evaluation and Program Planning

Regina Switalski Schinker

Journal Description

Below is the explanation of the purpose and intent of the *Journal of Evaluation and Program Planning (JEPP)* and a description of the types of articles published in this journal as presented by the publisher.

Purpose and Intent of the Journal

JEPP is based on the principle that the techniques and methods of evaluation and planning transcend the boundaries of specific fields and that relevant contributions to these areas come from people representing many different positions, intellectual traditions, and interests. In order to further the development of evaluation and planning, *JEPP* publishes articles from the private and public sectors in a wide range of areas: organizational development and behavior, training, planning, human resource development, health and mental, social services, mental retardation, corrections, substance abuse, and education. The primary goals of the journal are to assist evaluators and planners to improve the practice of their professions, to develop their skills and to improve their knowledge base.

Types of Articles Published

JEPP publishes articles, "special issues" (usually a section of an issue), and book reviews. Articles are of two types: 1) reports on specific evaluation or planning efforts, and 2) discussions of issues relevant to the conduct of evaluation and planning.

Reports on individual evaluations include presentations of the evaluation setting, design, analysis and results. Because of *JEPP*'s focus and philosophy, however, they also devote a special section to "lessons learned". This section contains advice to other evaluators about how you would have acted differently if you could do it all over again. The advice may involve methodology, how the evaluation was implemented or conducted, evaluation utilization tactics, or any other wisdom that you think could benefit your colleagues. More general articles provide information relevant to the evaluator/planner's work. This includes theories in evaluation, literature reviews, critiques of instruments, or discussions of fiscal, legislative, legal or ethical affecting evaluation or planning.

Special issues are groups of articles which cover a particular topic in depth. They are organized by "special issue editors" who are willing to conceptualize the topic, find contributors, set up a quality control process, and deliver the material. Often several editors share responsibility for these tasks. Suggestions for special issues are encouraged.

Book reviews cover any area of social science or public policy which may interest evaluators and planners. As the special issues, suggestions for books and book reviewers are encouraged.

This journal publishes four times a year.

**‘Somewhere Out There:’ A Survey of the Oft Lone Journey of Evaluators
Working in Business and Industry Settings**

Trude J. Fawson, Vanessa Moss-Summers and Sharon Marie May

[Evaluation and Program Planning](#)

[Volume 27, Issue 3](#) , August 2004, Pages 335-340

Summary: The purpose of this study was to (1) locate evaluators working in the for-profit sector and; (2) to obtain detailed information about what they do, how they do their jobs, and the barriers and supports that they encounter in their work environments.

This article was very helpful for a new professional in evaluation because it gave the “status quo” of the life of a professional evaluator in a for-profit organization. As the article states, there is not a lot of work done on evaluation in the for-profit sector so this article provided useful insight. Particularly since most organizations are for-profit.

This article was very comprehensive. Information was gathered from members of the AEA, the ROI Network of the American Society for Training and Development, the International Society of Performance Improvement and a variety of industries such as airlines, banks, technology companies, insurance, health care/pharmaceutical, manufacturing, retail, and transportation.

First, the article reported on the nature of the job of evaluation. The researchers found that even the largest global corporations “were unlikely to have large, full time staffs of evaluators.” Only one third of the respondents consider evaluation as their primary focus. In addition, the evaluators working in the corporate setting felt “isolated with few peers to rely on for mentoring and support.” These evaluators also felt that they had a ways to go to receive the recognition and respect that their jobs should bring.

Secondly, the article collected information on the types of tasks the evaluators in for-profit organizations perform. They include analyzing data, collecting data, collecting information about participant reactions to new programs/products/interventions, creating measurement instruments, and

recommending solutions/interventions. “The evaluator within the for-profit sector was most likely to design instruments, collect data, analyze data, and recommend solutions.”

Next, the challenges and frustrations that were reported for the professional evaluator were similar to many professionals. With corporate downsizing, there is a concern for lack of time and resources to complete a worthwhile evaluation. There was also the concern that the demands for immediate results from an evaluation were unrealistic. Often times, programs were cancelled before important changes could occur. Management fear of negative reports and pressure from top managers to make the results “look good” were also hurdles for evaluators.

For a successful evaluation, the article found that the client organization must have leadership that is open and willing to participate in honest evaluation. The evaluators must be knowledgeable, ethical, willing to work in a collaborative role, have a high level of skill, be objective, and be able to communicate well with the organization and other team members.

Finally, the most important statement from the article was, “Although inroads are being made, our profession has a tough sales job ahead of it.” In addition, the authors' state:

“Great opportunities exist for evaluators in corporate America in the post Enron era. There is much that the profession can contribute to management accountability and achievement of corporate objectives. However, a big sales effort will be required before evaluators will hold key jobs in all areas of corporation.”

Not Drinking the Poison You Name: Reflections on Teaching Ethics to Evaluators in For-Profit Settings

Michael Morris

[Evaluation and Program Planning](#)

[Volume 27, Issue 3](#) , August 2004, Pages 365-369

The first sentence in the abstract of this article states: “Recent corporate scandals represent a type of ethical challenge frequently encountered by evaluators within and outside of for-profit settings: the pressure to misrepresent findings.” The article refers to the “Enron Era” and the scandalous affairs that seem to happen more frequently in the for-profit world. Yet, it discusses how we must train future evaluators, with the leadership of the AEA’s Guiding Principles for Evaluators, to be professionals of integrity.

The article cited a Morris and Cohn 1993 study that asked a national sample of evaluators the ethical problem that they encountered most frequently. Their response was “being pressured by a key stakeholder (usually the primary client) to alter one’s presentation of the findings was, by far, the challenge most often reported.” While we can certainly hope that this type of ethical problem is a rarity, we, as evaluators must be aware of the type of pressure we may receive from the stakeholders of our evaluation projects.

This article makes suggestions for successful preparation of professional evaluators. Evaluators must received more training in areas of finance and accounting if they intend to practice their skill in the business world. Evaluators must be familiar with multiple approaches for analyzing data and be able to communicate well their findings. Evaluators must have respect for all stakeholders

in an evaluator project, not just the powerful. Evaluators must remember that “core societal values are embedded in the professional principles and standards that guide them as evaluators.” What evaluators do goes beyond the organization they are studying. Evaluators should keep an “ethics journal” to review and share with colleagues.

This article concludes with suggesting that an ethics course should be included in a professional evaluators training. Ultimately, large or small, all evaluators will be faced with an ethics issue at some point. We should be able to assume that for the sake of the professional and the profession, the evaluator will make the correct choice.

New Directions for Evaluation (Vol. 102)

Mary Keating

Views on Cultural Competence

A summary of the Summer 2004 edition of New Directions for Evaluation entitled “In Search of Cultural Competence in Evaluation—Toward Principles and Practices”

“Culture is present in evaluation not only in the contexts in which programs are implemented but also in the designs of these programs and the approach, stance, or methods evaluators choose to use in their work. A common thread between culture and evaluation is the concept of values. Culture shapes values, beliefs, and worldviews. Evaluation is fundamentally an endeavor of determining values, merit, and worth.” (p. 6)

The most recent edition of *New Directions for Evaluation* reviews the impact that cultural influences may have in evaluations. The journal begins with an overview of the historical and current ideas surrounding cultural competence and then looks at the topic from the perspective of different cultures: African American, American Indian and Alaskan tribal communities, and Latinos. The journal then examines issues with program evaluation in multicultural settings and potential next steps for the future.

Many evaluators may attempt to consider cultural influences on a project and think only of the physical characteristics of a community such as food, music,

celebrations, and clothing. However, there is a much deeper layer “...that influences customs, norms, practices, and social institutions, including psychological processes, language, caretaking practices, media, educational systems, and organizations (Fiske, Kitayama, Markus, and Nesbitt 1998, p. 6).” Many of these factors are often either topics themselves of evaluation or relevant values that should be considered in doing an evaluation’s KEC. Evaluators are asked to consider the stakeholders’ and clients’ values when processing an evaluation—but how, if they have no heritage in that culture?

“The APA’s multicultural guidelines (2003) posit that culture is the embodiment of a world view through learned and transmitted beliefs, values, and practices, including religious and spiritual traditions. It also encompasses a way of living informed by the historical, economic, ecological and political forces acting on a group.” (p. 6)

When there are value differences they often lead to value conflicts that can impact an evaluation from problem definition all the way to recommendations. The journal gives several examples of Eastern versus Western cultures or even the impact of Caucasian researchers interpreting the experiences of low income minorities. The editors also debate the approach used for many policy evaluations involving funding that are done on needs based rather than a strengths based approach. Rather than trying to understand and determine the needs, it may be more beneficial to focus on the strengths and how those positive attributes can be leveraged in a program. The evaluators must possess an “active awareness, understanding, and appreciation for the context at hand, and uses responsive and inclusive means to conduct evaluation” (p.11).

In the second chapter of the journal, “A Journey to Understand the Role of Culture in Program Evaluation: Snapshots and Personal Reflectors of one African American Evaluator” by Stafford Hood, the author states: “It is my continuing belief that few evaluative approaches of the past (or for that matter the present) have seriously considered race, culture, poverty, or contextual context as anything more than ‘error variance’” (p. 22). Hood reviews various historical views from leading experts such as Cronbach, Scriven, Stufflebeam, and Tyler. He refers to evaluation standards that recommend evaluators understand the audience and their needs as part of the evaluation process. Hood takes it one step further in identifying stakeholders:

1. Include less powerful groups or individuals as stakeholders, such as racial, cultural, or language minority groups,
2. Determine how they would like to use the results and for whom the information will be particularly useful,
3. Include clients and stakeholders in designing and conducting the evaluation as well as interpreting the results (p. 30)

Evaluators should also have to meet PES criteria in order to perform culturally diverse evaluations. There are guidelines that state that the evaluator must be “knowledgeable of the social and political forces affecting the less powerful stakeholders and use this information in designing and conducting the evaluation.” (p. 30) However, Hood questions whether these criteria are being enforced in the evaluator interview process.

Lastly, the AfrEA has started to amend many U.S. standards to address cultural considerations. For example, written law as the highest U.S. value versus “law is

not always more important than tradition or custom”.(p. 34) Hood believes that there is a great foundation in the evaluation industry, however, and as the applications evolve, the standards should also evolve to be inclusive of all cultures. Hood states: “Properly used with a culturally responsive approach, the good work of Tyler, Stake, and Stufflebeam permits mid-course corrections in evaluation. It permits feedback to the decision makers for change.” (p. 35)

The third module of the latest *New Directions for Evaluation* journal is entitled “Culturally Competent in Indian Country” by Joan LaFrance. Similar to the cultural differences described thus far, this module stresses the uniqueness of tribal sovereignty and the need for cultural integration into the evaluation practice. Throughout the article, La France reinforces the need to seek the knowledge of the community, build relationships and reflect those learnings in the methodological process. This strategy continues the momentum of truly getting to know your audience and the cultural values that influence them.

The fourth article, “Developing & Implementing Culturally Competent Evaluation: A Discussion of Multicultural Validity in 2 HIV Prevention Programs for Latinos,” Ross F Conner examines the importance in distinguishing internal, external and construct validity.

“Cultural issues and differences can be important factors in understanding which variables did or did not cause differences in programs (internal validity), which effects generalize over other settings and times (external validity), and what effects mean for higher order constructs and implications (construct validity). Multicultural validity therefore extends the issues evaluators need to

be attentive to if they are to draw valid conclusions, set out well-grounded implications, and make accurate recommendations.” (p. 52)

The writer goes on to explain key factors in the two programs that were run and once again proves that participant inclusion throughout the evaluation to enable cultural context is critical for success. One additional point discussed in this module is the inclusion of both the literal and figurative language of the participants in the evaluation process. This inclusion of speech allows the cultural influences to permeate and not be hidden during the examination.

The remaining articles continue to support the need for cultural inclusion and immersion in evaluation programs. These articles reinforce the three future needs that are identified by the editors: policies regarding cultural competence and developing practice guidelines, a pool of qualified multicultural evaluators, and more reports and best practice sharing of multicultural evaluation programs (p. 15). Evaluators should be reminded that cultural context has an impact on both international programs and domestic programs. Domestic programs that are in a community unfamiliar to the evaluator may have just as many cultural competency complexities as foreign lands. Evaluators must realize the need to understand cultural values in all scenarios. The definition of cultural competence as articulated below, embodies the current issues surrounding evaluation and the severe impact that cultural context has played in the past and could in the future if not addressed:

“A systematic, responsive inquiry that is actively cognizant, understanding, and appreciative of the cultural context in which the evaluation takes place; that frames and articulates the epistemology of the evaluative endeavor; that employs culturally and contextually appropriate methodology; and that uses stakeholder generated,

interpretive means to arrive at the results and further use of the findings.” (p. 13)

References

Thompson-Robinson, M., Hopson, R., & SenGupta, S. (Eds.). (2004). In search of cultural competence in evaluation: Toward principles and practices. *New Directions for Evaluation*, 102.

Education and Personnel Evaluation: A Review of Leading Journals for the Period 2003-2004

Nadini Persaud

This literature review highlights interesting research during the period 2003-2004, based on the Journal of Personnel Evaluation in Education (last issue produced June 2003), Journal of Human Resources, Journal of Teacher Education, NASSP Bulletin and Education Administration Quarterly.

Psychological and Contextual Influences on the Teacher Selection Interview: A Model for Future Research by D. A. Delli and E. M. Vera

The article “Psychological and Contextual Influences on the Teacher Selection Interview: A Model for Future Research” suggested that psychological influences can bias the interview process. According to the authors, research suggests that variables such as *attractiveness, likeability, expertness, and similarity* can influence interviewers’ perceptions of prospective employees and applicants’ perceptions of the interviewer and the job itself. The article notes that interviews, not paper qualifications, have been rated as the most important criteria in making hiring decisions; therefore education administrators need to be aware of these issues to ensure that the selection process is not biased by psychological influences. The authors suggest that in certain professions such as teaching and nursing where the ratio of candidates to jobs is very low, selection processes may deviate from the norm and be subject to contextual influences (interview structure, interview format and the influence of local and federal policies on interview outcomes) as well as psychological influence. The authors conclude by noting that

awareness of both the psychological and contextual issues that can influence the hiring of teachers can help ensure that the best qualified teachers are recruited.

Summative Evaluation of Student Teachers: An Enduring Problem by J. Raths and F. Lyman

The article “Summative Evaluation of Student Teachers: An Enduring Problem” noted that education faculty are not taking sufficient care to prevent weak and incompetent student teachers from attaining state licenses. Factors contributing to the problem include:

- *Conflation of summative and formative evaluation:* As defined by Scriven (1967), summative evaluation involves decisions such as final grades, promotion to the next class, acceptance to a university, granting a license, giving merit pay increases and other similar performance incentives. Formative evaluation is akin to coaching and is designed to improve student performance or improve the products on which students are working. According to the authors, in the mode of formative evaluation, teachers rarely issue overall judgments about the quality of a performance or a product, but instead point out particular strengths and/or weaknesses, suggest how weakness might be addressed, and encourage reflection. The authors also note that there is an inevitable conflict of interest when coaches make summative evaluations of the performances of their students, because initially they start out acting as a coach but later become the evaluator.
- *Minor roles within the larger teacher education profession:* Teachers may not be full time, may not be faculty and may not have received adequate coaching. They have their own duties and see the supervision process of student teachers as secondary.

- *Uncertainty*: There is no clear distinction between what constitutes competent and incompetent student teachers. This makes it difficult to make a high-stakes judgment about an individual student teacher armed only with vague decision criteria, and administrators want to avoid the threat of lawsuits and grievance petitions.
- *Rhetoric of excellence*: Language on evaluation forms give emphasis to “excellence” and “high quality” in teaching, but provides minimal guidance to enable supervisors to determine what constitutes incompetence and what is really excellence and high quality.

Responding to incompetence clarification, the authors make two distinct contributions: (a) they place incompetence in a continuum of teaching behaviors from criminality and malpractice through best practice and (b) the teaching acts in specific settings are pronounced as indicators of teacher incompetence. The article concludes with suggested caveats that need to be taken into account as the definitions of incompetence are shared or adopted.

International Comparisons of the Relationships among Educational Effectiveness, Evaluation and Improvement Variables: An Overview by C. Teddlie, S. Springfield and J. Burdett

This study focused on teacher effectiveness research, school effectiveness research, teacher evaluation, staff development, teacher improvement and school improvement in five countries: Cyprus, Hong Kong, The Netherlands, UK and the USA. The study concluded that teacher evaluation is used for three major purposes across the countries, accountability, promotion and staff development. It noted that links between teacher evaluation and other constructs, such as staff development and educational improvement, are undeveloped in all five countries and that there

is substantial room for strengthening the links among teacher evaluation, teacher improvement and school improvement in all five countries.

The Effects of Chronological Age and Information Media on Teacher Screening Decisions for Elementary School Principals by I. P. Young and F.P. Chounet

The article “The Effects of Chronological Age and Information Media on Teacher Screening Decisions for Elementary School Principals” was an experimental study that examined traditional versus modern methods of screening decisions made by elementary school principals. The study failed to detect any statistically significant multivariate effects for chronological age of teacher candidates; however, it did show that principals may well overlook many qualified teacher candidates who choose to use electronic mediums to communicate teacher credentials to organizational representatives, at the screening stage of the selection process.

Continuing Assessment of Teacher Reactions to a Standards-Based Teacher Evaluation System by H. G. Henemann III and A. T. Milanowski

“Continuing Assessment of Teacher Reactions to a Standards-Based Teacher Evaluation System” was a study that provided a description and evaluation of the first two years of district-wide implementation of a standards-based teacher evaluation system in the Cincinnati public schools. The authors concluded that many of the teachers viewed the four performance domains and accompanying standards as highly relevant to their job and consistent with their conceptions of good practice. However, others became dissatisfied and left the district because of fear that the evaluation results would be linked to pay increases in the future.

The Impact of Teacher Training on Student Achievement: Quasi-Experimental Evidence from School Reform Efforts in Chicago by Brian Jacob and Lars Lefgren

This study used a regression discontinuity strategy to estimate the effect of teacher training on math and reading performance of elementary students in 20% of Chicago's lowest-achieving elementary schools. It concluded that marginal increases in in-service training have no statistically or academically significant effect on either reading or math achievement. This suggests that modest investments in staff development may not be sufficient in increasing the achievement of elementary school children in high-poverty schools.

The Efficacy of Portfolios for Teacher Evaluation and Professional Development: Do They Make a Difference? by P.D. Tucker, J. H. Stronge, C. R. Gareis and C. S. Beers

This study, conducted in a small school district, introduced the use of portfolios as part of a new evaluation system to determine the efficacy of portfolios in evaluating teacher performance both for accountability and professional development purposes. The conclusion was that the portfolio system was found to be particularly useful in documenting teacher performance in the areas of assessment and professionalism. However, survey and focus group data indicated that while teachers and administrators viewed portfolios as fair and accurate measures of performance, both had concerns about their feasibility given the time demands of portfolio development.

Evaluating the Evaluators: Teachers' Perceptions of the Principal's Role in Professional Evaluation by S. Zimmerman and M. Deckert-Pelton

Eighty-six educators from five northwest Florida counties were surveyed to examine their perceptions of their principals as effective evaluators. The results suggest that teachers' perceptions of an effective evaluation process involve a focus on their principals' knowledge, skills, and abilities as both an experienced educator and as an educational leader. From the respondents' comments, four key domains emerged as pivotal components to a successful professional evaluation process: interactions between principal and educator; consistent evaluations; principal commitment to effective professional evaluation; and a principal knowledgeable in pedagogy, content, and evaluation.

References

- Delli, D.A. & Vera, E. M. (2003). Psychological and contextual influences on the teacher selection interview: A model for future research. *Journal of Personnel Evaluation in Education* 17(2): 137-155.
- Jacob, B. & Lefgren, L. (2004). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. *The Journal of Human Resources* 39(1): 50-79.
- Henemann III, H. G. & Milanowski, A. T. (2003). Continuing assessment of teacher reactions to a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education* 17(2): 173-195.
- Raths, J. & Lyman, F (2003). Summative evaluation of student teachers: An enduring problem. *Journal of Teacher Education* 54(3): 206-216.
- Teddlie, C, Springfield, S. & Burdett, J. (2003). International comparisons of the relationships among educational effectiveness, evaluation and improvement

variables: An overview. *Journal of Personnel Evaluation in Education* 17(1): 5-20.

Tucker, P.D, Stronge J.H., Gareis, C.R. and Beers, C.S. (2003). The efficacy of portfolios for teacher evaluation and professional development: Do they make a difference? *Educational Administration Quarterly* 39(5): 572-602.

Young, I. P. & Chounet, F. P. (2003). The effects of chronological age and information media on teacher screening decisions for elementary school principals. *Journal of Personnel Evaluation in Education* 17(2): 157-172.

Zimmerman, S., Deckert-Pelton, M. (2003) Evaluating the evaluators: teachers' perceptions of the principal's role in professional evaluation. *NASSP Bulletin* 87(636): 28- 36.