Hawai'i-Pacific Evaluation Association
Keynote Address, September 11, 2009

THE NEXT BIG THINGS IN EVALUATION: THE HAPPINESS INDEX AND A FEW MORE

Lois-ellin Datta[1]
Datta Analysis

Three points.

First, evaluation is less than 50 years old, yet many next big things from the 70s, 80s, and 90s have grown from seedlings to sequoias. They continue to engage us, they are still works-in-progress, and they are still debated. These ideas are infused, or well along to being infused, however, rather than cutting edge.

Second, we may have some emerging next big things in evaluation. We can find them in a new book we recognize as a future classic. We can detect them in a journal article that is five wows. We can hear them in a great panel, in the reports of an innovative project, in a new evaluation voice.

Third, we may be too new a field to over-quantify whether we have the most adaptive proportion of leading edge, lifting surface, and following edge ideas for our plane to fly. We can look, none-the-less, at some indicators such as AEA sessions and workshops compared with what we know of evaluation needs. My take is we have a marvelous lifting surface, but perhaps need more leading edge.

The next sections discuss these ideas further. Your crystal balls may be clearer, you stand tall in evaluation practice today, and I invite your participation. Our friends are passing out 3 x 5 cards now. What is your candidate for the next big thing in evaluation? We'll gather your advice in a few minutes and share it later.

<u>Seedlings to Sequoias:</u>

Over the past twenty years, our evaluation landscape has been transformed. The top ten ideas that were leading edge yesterday and are infused in our practice today may include:

1. Stakeholder involvement

2. Program logic models and theory-driven evaluation

3. Performance and implementation measurement

4. Mixed methods, including integration of process, implementation, and outcome data

[1] Lois-ellin Datta may be reached at datta@ilhawaii.net or 78-7054 Kamehameha III, #1304, Kailua Kona, HI. 96740

5. Formative and summative evaluation

6. The importance of context

7. Cultural competency, culturally appropriate evaluations; evaluation for and by indigenous peoples

8. Understanding multiple ways of establishing influence, attribution, and causality

9. Empowerment, participatory, democratic, values-engaged, utilization-focused, consultative, responsive, appreciative inquiry, deliberative democratic, fourth generation, and critical friend stances

10. Rigorous evaluation and evidence-based practice; Government Performance and Review Act (GPRA) connections between evidence of results and budgets; accountability

By infused, I mean these concepts are part of our every-day toolkit—ours, as in everyone who does evaluation.

Much of our work, as indicated by conference presentations and journal articles, involves strengthening our grasp of these concepts, polishing our skills in what is needful to apply them, and improving our understanding. This is what I mean by our lifting surface (see., e.g., Donaldson, Christie, and Mark, 2009).

For example, almost all evaluators are not likely to be experts at the best ways to assess extent and quality of stakeholder involvement. Almost all evaluators are, however, likely to be aware of the importance of stakeholder involvement both for evaluation utilization and because it is ethically right. Further, most of us have our own repertoire for achieving appropriate stakeholder involvement. Yet, appropriately, we continue to learn more about different approaches to stakeholder involvement, more about its connections with other aspects of evaluation, more about its limits and limitations.

We continue to debate the circumstances under which a randomized design may fit best and where one might better use propensity scores to create a counter-factual, where a subtractive design best addresses an evaluation question and where an additive design does.

We continue to advance the awareness, practice, and theory of culturally appropriate evaluation including the primacy of evaluation for and by indigenous peoples.

We share practice-based knowledge of possibilities and pitfalls, and this is good. It is particularly good because we still can hear those who created our field, such as Bob Boruch, Eleanor Chelimsky, Brad Cousins, David Fetterman, Karen Kirkhart, Donna Mertens, Michael Quinn Patton, Patricia Rogers, Michael Scriven, Will Shadish, Tom Schwandt, Carol Weiss, and others. They are among our shiningest stars and I'd walk many a mile to hear them. They are our kupuna and our story. In our interest in them, we honor them and their work, as it should be.

But the concepts I've mentioned are not our leading edge. New understanding on a old topic does not necessarily mean big or the next big thing.

Consider, for example, a well-done, thoughtful review of research on evaluation use in the

September 2009 issue of the American Journal of Evaluation (Johnson et al.) Its conclusion?

"Findings point to the importance of stakeholder involvement in facilitating evaluation use and suggest that engagement, interaction, and communication between clients and evaluators is critical to the meaningful use of evaluations." (p. 389).

This is what I mean by strengthening our grasp and deepening our understanding. It is not what I mean by "the next big thing."

As another example, Patricia Rogers will give a workshop at the American Evaluation Association Conference on purposeful program theory. She notes,

"[Attention will be given] to three particularly important issues: improving the quality of models by drawing on generic theories of change and program archetypes; balancing the tension between simple models which communicate clearly and complicated models which better represent reality; and using the model to develop and answer evaluation questions that go beyond simply meeting targets."

That is terrific, incorporating continuing lessons learned from Rogers' extensive practice, and it should be a standing room only session. This also is what I mean by strengthening our grasp and deepening our understanding. It is not what I mean by "the next big thing."

As another example, thanks in great measure to our Maori colleagues such as Linda Smith and Fiona Cram, and to leaders such as Kathy Tibbetts, Joan LaFrance, and Hazel Symonette, organizations are supporting culturally appropriate evaluation on a wide scale. See for example, the materials from the Administration for Health Care Quality, the Annie E. Casey Foundation, and the National Center for Cultural Competency at Georgetown University, together with the recent handbook on cultural competency in evaluation by Denzin and Lincoln (2008). There is much to be done, much to be learned, but this also is not what I mean by "the next big thing."

How much do these strengthening and deepening efforts engage us? One indicator is the workshops offered for the 2009 American Evaluation Association conference.

Lifting surface presentations on quantitative methods and theories, such as non-parametric methods --21%
Lifting surface presentations on qualitative methods and theories, such as empowerment evaluation -- 21%
Lifting surface presentations on basics such as an introduction to evaluation and report writing -- 23%
All the rest--35%

About the same distribution was found in recent articles in the American Journal of Evaluation. Omitting book reviews and letters to the editor,

Quantitative methods and theories -- 21%
Qualitative methods and theories -- 36%
Foundational such as evaluator roles -- 29%
All the rest -- 14%

So what about the possible next big things in evaluation?

The Next Big Things

A big thing changes how we do or think about evaluation. As examples, "Empowerment Evaluation" did this (Fetterman and Kaftarian, 1995). "Utilization Focused Evaluation" did this (Patton). "Experimental and Quasi-experimental Designs for Research" did this (Campbell and Stanley). "Fourth Generation Evaluation" did this (Lincoln and Guba). "Decolonizing Research" did this (Smith, 1999). "Unobtrusive Measures" did this (Campbell and Secrest).

By this definition, how does one pick a winner or set of winners here? By what criteria? Mine are

-- approaches that might meet an important, unmet need, that evoke “Wow” when I read about them

--concepts which have charismatic, prolific advocates, because advocacy seems to play a role in what survives among many new ideas, as a big thing

--issues which are receiving increasing attention, even if the evaluation thing hasn’t caught up yet and

-- hunches

By these criteria, here are my candidates for the next big things, starting with those whose tops are visible above the waves.

1. *Systems concepts and methods in evaluation*, particularly the work of Bob Williams of Australia, generously made available through his excellent web site and his commentaries on Evaltalk. Williams and Iman's 2005 book may prove to be the equivalent of Fetterman’s Empowerment Evaluation. Systems concepts already have traction in how we talk and think, infusing, for example, much of Michael Quinn Patton's recent articles and speeches.

The conceptual framework of systems approaches to evaluation needs, however, to be followed up with more show-us and how-tos. We need more applications establishing how much greater its benefits in what evaluation situations, compared to evaluations ignoring systems concepts. Systems approaches in evaluation may shake out to be a variant of taking context into account, but I believe it could be far more, in taking program evolution and complex, interacting influences into account.

2. *Applying evaluation concepts, methods, and practice beyond our human services roots, outside of our comfort zones and boxes*. A recent New York Times article discussed a major study intended to shine the spotlight on innovative and cost-effective approaches to health care provision. The methodology was looking at care providers nominated as outstanding to identify the common and presumably distinctive factors.

Gawande et al. (2009) claim:

"There is a lot of troubling rhetoric being thrown around in the health care debate. But we don't need to be trapped between charges that reforms will ration health care and doing nothing about costs and coverages. We must instead look at the communities that are already redesigning American health care for the better, and pursue ways for the nation to follow their lead." (p.A27)

Ouch, major ouch. This came, as I understand it, from a medical research and policy group. An evaluator, I think, would have studied a similar number of providers identified as failures with regard to costs. This would help establish that the common success factors weren't equally characteristic of failure. This concern applies, of course, to Promising Practices clearinghouses and approaches. One wants factors that differentiate as well as agglutinate!

If this is typical, not just illustrative, in areas such as agricultural sustainability, energy conservation education, and failed nations where evaluations increasingly are required, we evaluators have a lot to contribute, as well, obviously, as a lot to learn.

The 2008 AEA conference had some first-rate sessions on these beyond-our-usual-areas topics, such as energy evaluations (Catrambone, 2009). The 2009 conference builds on these (Jordan, 2009, for instance). Not a charismatic super-star yet or something which adequately brings the products of this cross-fertilization together, but there are sessions with next new thing potential.

By searching out emerging leaders on emerging topics, inviting them to participate in our larger and smaller conferences, we may gain the benefits of a well-connected archipelago of evaluation in dispersed areas. A useful indicator of such spacious vision is whether an article's references are exclusively in its field and mostly those of the author or whether there's a connecting archipelago-mind at work.

3. *Following the money.* Another next big thing is incorporating the concepts and methods for evaluating return on investments, costs, cost-effectiveness and cost benefits in the human services areas. Many of these come from economics and utility studies in areas such as health.

An example: a drug, Sutent, has been in the news lately when a man died in Great Britain after being denied the expensive drug. Efficacy studies show Sutent costs above $50,000 (US) a treatment year and it yields less than six additional months of a Quality of Life Year (QALY) for advanced kidney cancer patients. Since 1972, Great Britain has quantified the situations under which the government will pay for treatments. A Quality of Life year is the months of anticipated perfect health remaining after subtraction of anticipated months of life with severe disabilities (DALY) as estimated in efficacy and efficiency trials, considering both mortality and morbidity (National Institute for Health Care and Clinical Excellence, Great Britain, 2009).

Other countries use QALY and DALY (and variants such as ED-ORD) for health care decisions (EuroQol Group, 2009). We do this but more indirectly. For example, the National Institutes of Health Consensus Panel reviews determine whether or not Medicare should pay for certain drugs and treatments, based on efficacy and effectiveness data. Quantifying the value of extended months of survival is not part of these decisions. It could be and probably should be. While not perfect, there's a wealth of experience in the QALY/DALY system on quantifying the value of life, a topic we shy away from (Joa and Libetkin, 2009; Sassi, 2006; Singer, 2009).

Some, I appreciate, will say "And we should shy away from this possibly slippery slope. Any life is worth unlimited medical care even for an extra day." I would agree if resources were unlimited and if we were willing to pay first the costs of good food, good housing, good education, and good health care for all our children. Since we are apparently not, and nations are looking hard at returns on investments, I see mindful, respectful, wise development of QALY/DALY-like measures and application of cost/utility analyses as a likely next big thing in evaluation.

Even in areas where we ought to be bolder, such as the cost and safety issues of nuclear waste disposal, there is a lot of "it is too complex to estimate the value of a human life; we just can't do it."

Further, we shy away from candid discussions of savings from human services programs such as special education compared to their costs. I am evaluating a program whose claims---to get the grants---anticipated savings. Now---understandably considering the context of the Hawaii Felix decision---the cost data are being refused, so my analyses are going to be cut, paste and estimate.

If I had a magic foundation wand, I'd set up a $50,000 prize for the best concept and practice in estimating the full, true costs and values for human service programs. We already know, from the extraordinary leverage of the High/Scope longitudinal studies showing that a dollar in the High/Scope early childhood programs yielded $7 in savings and benefits, that such information can be well-understood by legislators and budget analysts.

Perhaps we might begin with quantifying the costs and full benefits efforts to sustain and enhance Native Hawaiian culture, including programs such as those supported by the Kamehameha Schools and the Native Hawaiian Education Council. Ormond Hammond and I have talked, for example, of the benefits of these initiatives to Hawaii's economy and our visitor industry. The Native Hawaiian Educational Council has made some fascinating advances on the journey toward a meaningful, common language for understanding holistic values and programs, that before long, I hope, will have indicators and measures aligned with these.

The methodology for cost analyses is available (see, for example, the work of Yates and of Levin). Much thoughtful analysis has been completed on nuances such as rate of return estimates and appropriate value for decisions (see Ludwig and Phillips, 2007). This is alluded to in some of our foundational texts in evaluation; Yates usually presents at least one workshop at the American Evaluation Association but the costs, relative to benefits, has not yet been given lifting surface attention. I think it is a leading edge issue whose time may be coming soon.

4. *Evaluation quality.* We have anecdotes but not data on evaluation quality. Lord knows there's a huge evaluation industry in training non-evaluators how to carry out adequate evaluations, including "The Nuts and Bolts of Program Evaluation," the Center for Disease Control books, the National Science Foundation guides on science education evaluation (Frechtling, 2002), and the United Way's evaluation guidebooks.

There are workshop satisfaction surveys but few if any assessments of knowledge, skills and abilities. There are some stories about the quality of the evaluations the workshop participants go forth and do but few systematic studies. One exception is for national evaluations carried out by professional evaluators. The study concludes the process, formative, and implementation components are good, and the results, outcomes, impact components are not so good.

Again, with the magic foundation wand, I'd put $50,000 into a comprehensive study of all evaluations submitted in Hawaii to the United Way, to the Hawaii Community Foundation, Castle and Cook, and others. How good are the hundreds of evaluations completed yearly in Hawaii alone as part of funding requirements? If they are not good enough, let's face it and consider what we should do.

Poor evaluations are not free. We often write of evaluation as being a good investment in better programs. I hope this is so. It is a painful truth, however, that evaluations do cost program director and staff time and they do reduce resources available for service delivery, in addition to direct costs. Direct costs can include both development evaluation time for demonstration programs and evaluation service time for operational programs. The indirect and opportunity costs for the evaluations can be as great as these and considerably greater.

Poor quality evaluations leading to illusory accountability and unachieved program improvement can be worse than no evaluations at all. Are these even or usually read? Does something happen in the grants office besides a tick on the "evaluation report received" list? Perhaps among the next big things in evaluation will be a courageous examination of the quality of practice and the actual use of these hundreds of reports..

5. *Shining the Compatibility Light and Fewer Word Warriors.* Another coming attraction may be shining the light on evaluations that demonstrate the compatibility among different evaluation stances. We spend so much time in dramatic stand-offs between word warriors. These boxing-ring like debates can illuminate, they can teach, but they also can set up an artificial belief in incompatibility among stances. Patton recently and wistfully complained that the formats --- set by the debate sponsors---force a bundling of ideas into utilization-focused evaluation that artificially excludes empowerment evaluation, which is not what he means.

I'm reading now an evaluation on railroad safety innovations that combines utilization and stakeholder focused approaches, a systems perspective, randomization, excellent use of implementation and process data, an appropriate diversity of process, context and outcome measures, and narrative methods. It isn't a hodge-podge, but a well-integrated study.

Another evaluation, this one conducted over seven years, also combined many of these elements. It is a national study of the efficacy and effectiveness of abstinence education to prevent teenage pregnancy and sexually transmitted diseases. The results, while not methodologically pristine due to post-randomization issues, are sturdy enough to have affected policy and practice, and no one, as far as I know, sees anything but good in the integration of stances.

This may be a next big thing in evaluation. I'd like to see it hastened by some thorough analysis of where it succeeded and where it didn't, to better help understand when stances may and may not be compatible...in deeds, please, not words.

There are also some fascinating new books and articles coming along. Both deal--courageously again--with the common but not much talked about experience of both program and evaluation failures.

6. *Recommending program closures.* The September 2009 issue of the American Journal of Evaluation includes an article (Eddy and Berry, 2009) on a model for decision-making and professional responsibility in recommending program closure. There is no shortage of more general discussions of whether evaluators should make recommendations. This article is different. Using the image of a scale, it identifies specific weights to be empirically determined that incline toward program improvement--- flexible factors about which the community can do something such as staff changes, program recruitment, and allocation of expenditures. The weights toward closure are the immutabilities such as saturation of services, legislative or bureaucratic barriers, and lack of qualified staff. Together with inhibitory and facilitative characteristics, these can provide a disciplined analysis supporting program improvement

versus program closure when under-performance is determined. The context is not failing to respect stakeholders or failure to celebrate small successes. Rather, if a recommendation is wanted and an appropriate, sensitive, respective evaluation shows the program is under-performing, how to take into consideration factors leading to improvement or closure.

This is, to me, at least a three WOW! article in the directness of the analysis, in taking head-on a really important but too often silent issue in evaluation, and in its timeliness. I'd look for the authors among our rising stars.

7. *Dealing with surprises*: In the past year, the unexpected and unanticipated have caught me again and again and again:

--a trusted colleague has been unable to transcribe the notes from key interviews because due to meltdown from some tough personal situations. The handwriting on the wall says I can kiss these data goodbye.
--efforts to help graduate students develop scoring rubrics for their assessments of student learning worked well in most instances but one graduate student, whose classes were a large N for this study, used an essay format thinking the rubric could come later and found the question led to unscorable results
--a new program similar to the one being evaluated has come to town, leading to many new demands on staff time, serious confounding with regard to the nature of the treatment, and no small challenge to the evaluation.

These little aggravations are hardly newsworthy. There are, it is said, only two kinds of sailors. Those who have already gone aground and those who are going to go aground. What is newsworthy is that Jonny Morrell (in press) has written a good book on evaluation surprises, how to avert them and how to deal with them, that is disciplined, systematic, theory-based, and practice-grounded. I'm not sure this will be as discipline-altering as say participatory evaluation, but it is worthy of inclusion among the next big things.

8. *The National Happiness Index and Other Affirmative Indicators.* When people are miserable and with scant control over their fates, they may appreciate a little gaiety and happiness. The films that entertained us in the economic boom years tended to explore the dark nights of the soul. During the 1930s Depression, it was Fred Astaire and Ginger Rogers who danced up that stairway to Heaven.

Indexes have been around a long, long time as guides to progress and policy. The Gross National Product, for example, was established in 1934 as a snapshot of how we were doing in the Mother of All Depressions. Child Count, that marvelous profile, was developed around 1980, drawing attention to children's unmet needs and states' progress.

In 1972, to some derision from economists, King Jigme Singye Wangchuck of Bhutan established the National Happiness Index. (For comments, pro and con, see for example, Kahnman et al., 2004; Revkin, 2005; New Economist, 2005). Such a title can suggest a national smiley faces survey but the NHI is more complex and spiritually harmonious. The Index reflects the Bhutanese belief that material and spiritual development reinforce each other. What counts---or at least what is counted in Bhutan---are (1) promotion of sustainable development, (2) preservation and promotion of cultural values, (3) conservation of the natural environment, and (4) establishment of good governance. In the 1990s, wellness indexes helped remind us the aim was well-being, not only smallpox avoidance.

Recently, a Happiness Index variant that involves empirical data on how our time is spent and how happy we are with these activities has been proposed by Alan Krueger, a much-lauded economist located at Princeton. It has been adopted by the U.S. Bureau of Labor Statistics as part of its surveys. In Nova Scotia, a similar concept, the Genuine Progress Indicator (GPI) is routinely reported (Pannozzo and Coleman, 2008). National, the U. S. Department of Energy Benefits Analyses (2009) and internationally, the Yale University Environmental Performance Index (2008) are instances of trying to count what matters. If your car breaks down and has to be repaired, that is a plus in our Gross National Product, for example; if your car goes for ten years without replacement, that doesn't register or is a negative. Let's hear it for Bhutan!

Dannemiller (2009) reports that during a planning meeting, he was told not to bring up the Bhutanese National Happiness Index as too frivolous. OK; he didn't. But after the workshops were over, the four elements of the [imagine the state name] index for planning for 2020 had the four Bhutanese elements.

9. Where Evaluation Leadership Now Lives

This may be lifting surface or leading edge. I'm not sure. Like the Big Bang, where it seems likely that a singularity in space time exploded and is dispersing at an accelerating rate creating the universe as it goes, evaluation leadership is dispersing at an accelerating rate and it lives internationally.

National and regional evaluation associations are wonderful. From them, come tomorrow's shiningest of stars. In them, emerging ideas are debated. Through them, we connect so the whole of our efforts can be more than its parts, particularly in our region where distances are vast and evaluators, few.

However, the intellectual drivers, in my view, may no longer our national associations. From what I see in the programs and journals, the functions of these organizations becoming U-shaped. They offer great opportunities for early career and idea development coupled with training, consolidation, and improvement for more established concepts...the lifting surface, the trailing edge.

The American Evaluation Association, for instance, now and again really catalyzed a next big thing in evaluation, such as the Seattle conference on international efforts in evaluation, led by Eleanor Chelimsky and Arnold Love of the Canadian Evaluation Association. Recently, fresher bolder more visionary initiatives in evaluation are emerging in international groups such as 3ie, from leaders such as Howard White, and from more specialized groups such as those concerned with energy and environmental evaluation.

A recent International Initiative for Impact Evaluation (3ie) newsletter, for example, reported that a synthetic review of preventing waterborne disease shows providing soap is more effective than education campaigns; invites proposals for looking at impact evaluations of interventions for ageing populations, invites another proposal for experiences in impact evaluation, summarizes the UK Government White Paper and the conservation party Green Paper on international development and impact evaluations, gives links to an interview (in Spanish) with Gonzalo Hernandex-Licoma, Ale Dean, and Mauricio Santa Maria, gives links to a working paper on designing impact evaluations from different perspectives, connects the reader to two new world bank impact evaluation reports, one on methodologies for evaluating the impact of large scale nutrition programs and the other on smart policy case studies of evaluations that influenced policies, plus connections to over 10 on-going international discussions and debates on evaluation issues. International Initiative for Impact Evaluation, 2009; International Development Research Council, 2009).

The implications for us? Stay attuned to the excellent website from 3 ie, where one can download new requests for proposals, read fine papers and articles, participate in world-wide dialogs, and be at the leading edge in our field. The internet is our friend, with links to emerging ideas and international resources in evaluation, fast, free, and often wise. And hey---maybe invite Howard White to come here, and see if one of the 3 ie meetings might be held in Hawaii or New Zealand or…

As for the tenth next big thing? Well, I've shared already a few of my ideas for the next big thing in evaluation. What are yours?

10. ______________________________

The Tenth Next Big Thing

The tenth next big thing, as seen by the H-PEA participants is---cultural competency. That is, about 40% of all responses had to do with cultural and indigenous evaluations, cultural indicators, models created by indigenous peoples for indigenous settings, the self as central to evaluation, cultural evaluations with ethnically same evaluators who have experience living the culture, and regarding cultural considerations as more important than technical considerations. This is an area I considered as lifting surface: more work to do, but not a leading edge area. Not so to our H-PEA participants!

About 30% of the responses focused on an area I've neglected but which seems to me very much a next big thing: evaluations utilizing the power of emerging technologies. These ideas included technological advances in data collection, multi-media techniques, digital story-telling, online interactive focus interviews, surveying via social networking, real time text polls, and advanced statistical techniques. The enthusiasm for this new big thing may have been stimulated by a well-attended workshop on new technologies in evaluation but it sure sounds right on.

There were some fascinating next big things that had wow! potential offered by one or maybe two people. These included getting to outcome indices that address both federal and community outcomes, incorporating different epistemologies, going from results of how things actually happen to developing new theories and (my favorite) developing program *illogic* models challenging the status quo as illogical.

Finally quite a few H-PEA participants listed as the next big thing areas I'd considered trailing edge or lifting surface, such as stakeholder involvement, utilization focused evaluation, appreciative inquiry, listening to participants' voices and qualitative measures.

One possible conclusion: we in H-PEA are diverse in our evaluation experience. What is excitingly new to some (such as stakeholder involvement) is infused in our every-day practice to others. Another conclusion: for many of the participants, the theme of the H-PEA 2009 conference, Cultural Competency, sounded a deep chord, and is perceived as urgently, significantly leading edge.

REFERENCES

Administration for Health Care Quality (2009) Guidelines on Cultural Competency. http://www.ahrq.gov/about/cods/cultcompded.htm.

Aldridge, M. L., et al. (2009) Improving data quality in large-scale, performance-based evaluations. *American Journal of Evaluation*, 30 (3) 426-436.

Annie E. Casey Foundation (2009) Race Matters Toolkit. http://www.aecf.org/KnowledgeCenter/PublicationsSeries/RaceMatters.aspx.

Campbell, D. and Stanley, J. 1xxx *Experimental and Quasi-Experimental Designs for Research*

Catrambone, J. C. (2009) Nonparametric Statistics: What to do when your data is Slewed or your sample size is small. Workshop, American Evaluation Association Conference, October 2009.

Chamber, R. (2008) Making the Poor Count: Using Participatory Methods for Impact Evaluation. Paper presented at the 3ie Cairo Conference.

Dannemiller, J. (2009) Personal communication.

Davidson, E. J. (2004) *Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation.* Menlo Park: Sage.

Denzin, N. K. and Lincoln, Y. S. (Eds.) (2008) *Handbook of Critical and Indigenous Methodology*. Menlo Park, CA: Sage

Donaldson, S. I., Christie, C. A., and Mark, M. M. (Eds.) (2009). *What counts as credible evidence in applied research and evaluation practice?* Los Angeles: Sage

Eddy, R.M. and Berry, T. (2009). The evaluator's role in recommending program closure: A model for decision making and professional responsibility*. American Journal of Evaluation* 30 (3), 363-376.

EuoQol Group. (2009) EQ-5D. http://www.euroqol.group

Fetterman, D. and Kaftarian, S. J. (1995) *Empowerment Evaluation: Tools for self-Assessment.*

Frechtling, J. (2002) *The User-Friendly Evaluation Handbook*. Washington, D.C.: National Science Foundation.

Gawande, A. et al. (2009) Ten Steps to Better Health Care. New York Times, August 13, p. A27. http://www.nytimes.com/2009/08/13/opinion/13gawande.html?pagewanted=2&_r

Guba, E. and Lincoln, Y. (1989) *Fourth Generation Evaluation.* Menlo Park: Sage.

International Development Research council (IDRC) http://www.idrc.ca/en/ev-61364-201-1-DO _ Topic.html

International Initiative for Impact Evaluation (3ie) 2009 Newsletter. August. http://www.3ieimpact.org/admin/pdfs_papers/50.pfd

Joa, H. and Libetkin, E. I. (2009) The statewide burden of obesity, smoking, low income, and chronic diseases in the United States. Journal of Public Health, 10, 1093. http://jpublichealth.oxfordjornals.org/gi/content/abstract/fdp012v1. Downloaded, 8/9/2009

Johnson, K., Greenseid, L.O., Toal, S.A., King, J.A., Lawrenz, F., and Volkow, B. (2009) Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation,* 30 (3) 377-410.

Jordan, G. (2009) Energy Program Evaluation: The Growing Field of Energy Program Evaluation and Where the Opportunities Lie. Workshop. American Evaluation Association Conference, October 2009.

Kahnman, D. et al. (2004) Toward National Well-being Accounts. *American Economic Association Proceedings.* May, pp. 429-434.

Kiernan, N-E 2009 Advanced focus group moderator training. Workshop, American evaluation Association Conference, 2009.

Levin, H. and McEwan, P. J. (2001) *Cost-effectiveness Analysis Methods and Applications*. Thousand Oaks: Sage.

Ludwig, J. and Phillips,. D. (2007) The Benefits and Costs of Head Start. National Povetry Center, report 07-09, February.

Morrell, J. (in press) *Evaluation in the Face of Uncertainty: Anticipating Surprise and Reacting to the Inevitable.* New York: Guilford Press.

National Center for Cultural Competence, Georgetown University. http://www11.gorgetown.edu/research/gucchd/nccc/index/html

New Economist, October 25, 2005. Latest from the Happiness Industry. http://neweconomist.blog.com/new_economist/2005/10/latest_from_the.html. Downloaded 8/9/2009

National Institute for Health Care and Clinical Excellence. Recommendation to the National Health Service regarding offering Sutent for advanced kidney cancer.

Pannozzo, L. and Coleman, R. (2008) The 2008 Nova Scoia Genuine Progress Index. http://gpiatlantic.org

Reed, J. 2009. IEPEC and AEA. Posting to Evaltalk. August 17, 2009, 7:15 a.m.. See also IEPEC and AEA, Posting to Evaltalk. August 17, 2009, 4:44 a.m.

Revkin, A. C., (2005) A New Measure of Well-Being from a Happy Little Kingdom.

Rogers, P. (2009) Purposeful Program Theory. Workshop, American Evaluation Association Conference, 2009.

Sassi, F. (2006) Calculating QALYS, comparing QALY and DALY Calculations. *Health Policy and Planning* 21(5) 402-408.

Singer, P. 2009 Why we must ration health care. New York Times, July 15 Downloaded August 8, 2009 http://www.nytimes.com/2009/07/19/magazine/19healthcare-t.html?_r=1

Smith, L. T. (1999) *Decolonizing Methodology: Research and Indigenous People.* London: Zed Books

United States Department of Energy (2009) FY 2010 Benefits Estimates. http://www1.eere.energy.gov/ba/pba/2010_benefits.html.

Yale University. (2008) Environmental Performance Index. Yale Center for Environmental Law and Policy and the Center for International Earth Science Information Network. http://epi.yale.edu. Download 8/9/2009

Yates, B. (1996) *Analyzing costs, procedures, processes, and outcomes in human services*. Washington, D.C. Social Science Methods.

Williams, B. and Iman, I. (2005) *Systems Concepts in Evaluation*. Point Reyes: Edge Press.