Journal of MultiDisciplinary Evaluation
Volume 11, Issue 24, 2015

JMDE
Journal of MultiDisciplinary Evaluation

ISSN 1556-8180
http://www.jmde.com

Question-Driven Methods or Method-Driven Questions? How We Limit What We Learn by Limiting What We Ask

E. Jane Davidson
Real Evaluation Ltd.

Insidious forces are at work in the evaluation space. The "methodologically manic-obsessive" thinking in our evaluation profession (Patton, 2012) and the metrics- and measures-obsessed lay users of evidence have managed to seriously limit the value of what we learn from evaluation. The evaluation questions that are asked at the front end are limited by the askers' narrow understanding of what is possible methodologically at the back end. This, alongside the political forces so eloquently outlined by Chelimsky (2011, 2012) and various psychological forces working against real evaluation, are major drivers of the single narrative thinking that pervades the formulation and evaluation of national and local government policies and initiatives.

In this editorial I outline some practical suggestions for how we need to ask real evaluative questions, and also bring them back up to the strategy and policy level. Part of the solution is to ask the big-picture questions that really need to be asked, and show how evaluation is able to step up to the plate, methodologically and otherwise.

The Seductive Attraction of Low-Level Questions

It's easy to get lost in the interesting and important question of which (and whose) values and interests should be reflected when defining and evaluating "improvements" in outcomes. It is quite another – conceptually and methodologically – to ensure we are even asking the right questions in the first place, let alone answering them.

Let me start with the notion of "improvement". This term is used throughout a lot of evaluation and applied research work, especially in the formulation of questions to guide the study. By far the most common expression of these questions seems to be: "Has [insert measurable outcome] improved?"

To the lay person, it seems intuitive and sensible enough. After all, the purpose of the policy, program, or project was to improve certain outcomes, so it makes sense to ask whether or not this has occurred.

To the evaluator or researcher steeped in the mores of applied social science research, we love a question like this because it is that most wonderful of things – operationalizable. It is measurable, testable, and answerable. One look at this question and we know exactly how to tackle it. Find a good validated measure of said outcome, measure it before and after the program or policy has been put in place, then run the right statistical significance test to ensure the measured change is most likely real and not due to chance. Et voilà, we have our answer – yes or no.

The policy makers, managers, and politicians are happy too. Finally, a straight and simple answer from the social scientists! We got improvement, so the policy worked. Right?

Wrong.

Measures and Metrics as Answers – Not!

It's a prevalent way of thinking that pervades public discourse about what's working; the media's way of discussing it; and sensemaking in the management ranks of just about any organization in any industry or sector, public or private. It's the notion that measures and metrics are answers in their own right. Guess what – they're not.

Evaluators may in their hearts know this, but we do, as a profession, tend to fall into one of two fatal traps here – massively oversimplifying things and massively overcomplicating things.

Oversimplification

Why do we so often see evaluation questions phrased in terms like "Has X improved?" Somewhere in the evaluator's subconscious echoes the voice of his or her professors to make sure any research questions are answerable, and this usually invokes the idea of testable hypotheses. So when we talk about improvements in outcomes, the social science-trained mind gravitates toward statistical significance testing. What is testable? The significance of the difference between pre and post. And so that's what we ask.

It's a classic case of the tail wagging the dog. The methods are driving the questions rather than the questions driving the methods. We need to get back to the basic notion of asking the most important questions even if they are the hardest to answer. We all know it is better to get even approximate answers to the truly important questions rather than seek precision on the measurable but trivial – so why don't we do it?

Overcomplication

I recall attending an industrial and organizational psychology conference several years ago, watching a presentation on the effectiveness of an organizational change initiative. The presenters had mapped out an elaborate theory of change on which they had measured every single variable and

put all their data through a structural equation modeling analysis. The presentation consisted largely of showing us the fit indices for various different alternative models.

And why had this approach been chosen by the researchers? Well, announced the lead presenter, it is only a matter of time before you won't be able to get an article accepted for a journal unless you use structural equation modeling. And why was this? Because structural equation modeling allows us to infer causation, he said.

Quite apart from the irrational belief in structural equation modeling's magical powers, what struck me in this presentation was how completely lost in the details the presenters were, and how utterly unable to draw any big-picture conclusions about whether this initiative had, in fact, been worth implementing. They had done an elaborate job of evaluating the theory of change and some more specific hypotheses related to it, but had completely forgotten to evaluate the initiative itself!

I have seen evaluators fall into the same kinds of traps time and time again. Don't get me wrong; I'm a frequent user of theory-based evaluation and many of the other useful methodologies we employ in our craft. But, as I have said before, we do need to keep our minds on the main task at hand. We're in the business of evaluating programs and policies, not theories (Davidson, 2007).

The Important Questions – and Why They Go Unanswered

Asking about "improvement" is marginally better than asking about "change" because it shows that we have at least determined the direction in which the change would need to go in order for things to be "better". But asking "Has X improved?" or even "How much has X improved?" doesn't go nearly far enough.

We need to ask (a) whether it changed substantially enough and quickly enough to be considered worthwhile; (b) how well it got people or communities to at least some minimal level of functioning; (c) the extent to which said "improvements" were at all attributable to the policy or program; (d) whether these were in fact the most needed improvements, and what else was also needed; (e) how sustainable and long-lasting any improvements were; (f) whether it was all worth the time, effort, and investment plunged into it; and so on.

Bottom line: the questions we need to ask are *evaluative* in nature, not merely descriptive in the value-neutral sense of the word. Evaluation questions also need to incorporate the notion of a causal link. In other words, they need to ask how much the policy or program has influenced (caused, catalyzed, or contributed to) important outcomes.

Although I can only base my observations on the evaluation work I see, hear, and review, I have found that it is still *astonishingly rare* for evaluations to even be asking these kinds of truly evaluative questions in the first place. Why is this? Michael Scriven, Syd King, and I explored this very question at the American Evaluation Association conference last year (Scriven, Davidson, & King, 2012). The main reasons we identified were:

1. evaluation anxiety
2. a lack of the evaluative attitude
3. value-free thinking
4. a lack of confidence and expertise to be able to answer them validly

1. Evaluation Anxiety

We often think of evaluation anxiety – an irrational negative response to drawing real evaluative conclusions – as an affliction on the client side (Scriven, 1991). But the reality is that evaluators themselves are just as affected by this instinct to – as Scriven puts it – "avoid the war zone".

Drawing evaluative conclusions means saying something explicit about the merit, worth, or significance of something. And that means living dangerously. People may not like what we conclude. And when they don't, you can be sure they will turn the microscope up on our evaluative reasoning, our methods, and our evidence. Our work will be mercilessly scrutinized, criticized, and questioned.

Not surprisingly, this makes evaluators extremely nervous, and there is a strong tendency to resile from drawing any real evaluative conclusions at all. The preference instead is to retreat back into our applied research comfort zone of simply presenting the evidence and perhaps pointing out a few minor strengths and weaknesses.

I am not able to cite a published paper that has examined the behavior and thinking of evaluators with respect to the truly evaluative parts of our craft. I am not sure anyone has ever studied this. Instead, I base my conclusions on two things. First, my own personal experience has repeated itself *every single time* I have collaborated with or engaged with evaluators or researchers *other than* those who deliberately seek to do genuine

evaluative work – and by that I mean people who ask and aim to directly answer questions about merit, worth, and significance. Second, the evaluations I have read, along with my observations at conferences and in online discussions, have revealed that the vast majority of those who identify as evaluators tend to stop well short of drawing an explicitly evaluative conclusion. If pressed for one, there is seldom a clear or direct answer.

2. A Lack of the Evaluative Attitude

The evaluative attitude, defined as "the relentless pursuit of the truth about quality and performance", was originally identified as a key element of a high-performance organizational culture (Sathe & Davidson, 2000). Based on experience working with large numbers of evaluators within an evaluative quality assurance framework, Syd King and I have also come to see this as an essential element of the evaluator psyche, the additional piece needed beyond evaluation knowledge, skills, and tools (King & Davidson, 2012; Scriven et al., 2012).

Whether evaluation is conducted in participatory/collaborative mode or as an independent undertaking, if the evaluative attitude is missing or weak, there is a tendency to just go through the motions, and not to use evaluative tools and frameworks to their most powerful effect. The evidence is gathered but the evaluative conclusions remain unclear or are not drawn at all. Real evaluation requires a mindset of seriously going after answers to questions about quality and value, and making sure they are well-reasoned and well-evidenced.

3. Value-Free Thinking

Scriven (1991, 2013) has long talked about the value-free doctrine and its insidious influence on evaluation. It initially prevented evaluation from being recognized as a legitimate activity at all. Although we have come a long way since these dark days, it is abundantly clear that value-free thinking is still alive and well. I have had comments from graduate students taking my workshops that what I say to them – about evaluative questions and answers being the absolute essence of what evaluation is – is diametrically opposed to what their professors are telling them. Keep values and "value judgments" out of your work, they say. Leave that to the decision makers.

Most who identify professionally as evaluators would laugh at the notion of value-free social science. We all have values, they say. Our values, our perspectives, and our biases influence our choices of research question, of instruments, of analysis method, of the way we choose to conduct evaluations.

But there is something deeper than this; it is the difference between "process values" and "deep values" in evaluation (Davidson, 2010). Process values are the values we draw on in determining how an evaluation should be conducted. This includes whether it is done in participatory or independent mode. In contrast, "the application of 'deep values' refers to the deliberate and systematic inclusion of…values in the very definitions of 'quality' and 'value' used in an evaluation, and in the evaluative interpretation of evidence" (p. 206).

In essence, most evaluators reject the notion of value-free social science, but they reject it for the least important reason. It is true that we all have lenses and perspectives that are rooted in our cultural and disciplinary identities and our personal experiences, and these do influence the way we design and conduct evaluations. But from that viewpoint, it is a very easy and seductive slippery slope to conclude that all values are arbitrary, idiosyncratic, and highly personal.

We might also call this value relativism, which is the belief that quality and value exist *only* in the eye of the beholder; it's all relative to one's own cultural biases and individual preferences and opinions. If this were true, there would be no such thing as a demonstrably "good" outcome or a "well-designed" program or "great value for the taxpayer dollar." Seriously? In my view, value relativism is fundamentally at odds with taking up evaluation as a profession at all. If we are not here to ask and answer questions about quality and value, then how are we different from some of our disciplinary relatives such as applied sociologists, applied psychologists, educational researchers, or statisticians?

It's been my observation that the prevalent way of thinking about values is to wave a hand at the notion that we all have them, but then to avoid any deliberate infusion of them in our evaluative work. Apparently the irony is lost on the many who identify professionally with a discipline whose essential core is answering questions about quality, value, effectiveness, and so forth, but who scrupulously avoid doing anything of the sort! It is somehow seen as more scientific to merely present the non-evaluative facts and to let stakeholders – who are allowed to have values because they are not scientists – infer for themselves what is

valuable and what is not, what is good and what is bad. The implicit message to the client wanting to know how good some of their outcomes were or how worthwhile the program was overall is "You work it out!" The problem with this, though, is that *it is actually our job* to draw these evaluative conclusions. Throwing that back to the client is basically shirking our responsibilities.

There are implications here for validity, credibility, and justice. Part of our job is to work out how 'quality' and 'value' should be defined for a particular evaluation, and this often involves gathering and synthesizing multiple value perspectives (e.g., taking into account needs, aspirations, unrealized potential, cultural and contextual considerations, and other relevant standards and values). Of particular importance is how we represent the needs and aspirations of those whose voices may not be clearly heard in the everyday life of the program. Provided we infuse these values carefully, systematically, and transparently, the evaluative conclusions drawn should be demonstrably valid and defensible. Contrast this with the conclusions drawn by a specific stakeholder with a particular set of interests and values and with no expertise in how to draw valid evaluative conclusions. They are likely to give less-than-adequate consideration to values and perspectives other than their own, particularly those whose voices usually go unheard, and they may draw erroneous conclusions due to this or to faulty evaluative reasoning.

4. Lack of Confidence and Expertise to Answer Evaluative Questions

My thesis so far has been that evaluators basically lack the mettle to take the evaluative bull by the horns and get the job done. But actually, there's another reason why we don't do this, and it is critically important.

It is all very well to ask evaluative questions – such as outcome-related questions about not just what effects emerged but whether they were substantial enough, fast enough, important enough to consider worthwhile. But it is no trivial undertaking to actually answer such questions in a systematic, transparent, valid, and defensible way (Davidson, 2012). Answers to real evaluative questions cannot be answered with the usual mix of qualitative and quantitative methods; they actually require evaluation-specific methodology (Davidson, 2005).

Several years ago (in about 2001-2002), as part of the groundwork for designing Western Michigan University's Interdisciplinary Ph.D. in Evaluation, I had occasion to look at the curricula of the available doctoral programs in evaluation offered around the world. Virtually none of them offered any coursework in evaluation-specific methodology. It is therefore not surprising that most evaluators simply do not know that such a skill set exists and do not know how to apply it. Consequently, they do not feel confident in drawing evaluative conclusions because they are not able to do so in any systematic, transparent, and valid way.

The same problem is mirrored when one looks at evaluator competency inventories from around the world. The vast majority, and particularly the most widely referenced competency lists, do not include evaluation-specific methodology or evaluative reasoning. For example, the *Essential Competencies for Program Evaluators* developed by Stevahn, King, Ghere, and Minnema (2005) lists "Makes judgments" but does not refer to any evaluation-specific knowledge or skills required to do this. Thus, this is implicitly listed as a cognitive task rather than a methodological one. This is not to say evaluative reasoning does not require cognitive skills, but in order for it to be systematic and transparent, it also requires evaluation-specific methodology. Similarly, the Canadian Evaluation Society's (2010) *Competencies for Canadian Evaluation Practice* lists "Draw conclusions and makes recommendations", for which the conclusions piece is unpacked as "Formulate conclusions for each evaluation question using inductive reasoning including answers that are implicit in the analysis, and potential answers that are consistent with, but not necessarily implicit in, the analysis" (p. 10).

There are some notable exceptions – but these are due to the influence of a few. Western Michigan University's Interdisciplinary Ph.D. in Evaluation lists evaluation-specific methodologies as required competencies (Western Michigan University, 2004). Michael Scriven has published "The Something More List" (Scriven, 2003) and "The Real Evaluation Competencies" (Scriven, 2006). Finally, the Aotearoa New Zealand Evaluation Association is, I believe, the only evaluation association to have placed values (including evaluation-specific methodology) at the core of its list of evaluator competencies, alongside and infused with cultural competence (Wehipeihana, Bailey, Davidson, & McKegg, in press).

I recently participated in an online discussion where I mentioned evaluation-specific methodology, the methodologies that were specific to evaluation. Given that the vast majority of

evaluators have been trained to believe that evaluation is little more than applied social science research with some consulting skills thrown in, it was not surprising when an evaluator with literally decades of experience asked, "What methodologies are these?"

As I replied in a blog post (Davidson, 2013), examples of evaluation-specific methodologies include:

- needs and values assessment
- merit determination methodologies (blending values with evidence about performance, e.g., with evaluative rubrics)
- importance weighting methodologies (both qualitative and quantitative)
- evaluative synthesis methodologies (combining evaluative ratings on multiple dimensions or components to come to overall conclusions)
- value-for-money analysis (not just standard cost-effective analysis or SROI, but also strategies for handling VfM analysis that involves a lot of intangibles, for example)

In contrast, the following would not count as evaluation-specific: statistics or any of the standard research methods (interviews, observations, surveys, content analysis, or even causal inference methodologies). Although evaluators clearly draw on these and use them a lot, they are not distinctive to evaluation because they are not specifically about the "values" piece. In other words, one could use these (non-evaluative qualitative and quantitative research methods) and still *not* be doing evaluation.

Without evaluation-specific methodology (ESM), either we are skipping the whole evaluative conclusions piece, or we are getting to it by logical leap (e.g. "I looked upon it and saw that it was good"). ESM is what allows us to get systematically and transparently from evidence about performance to evaluative conclusion, by weaving in the values ("how good is good") piece.

As an aside, it is true that several disciplines use evaluation-specific methodologies (e.g. industrial & organizational psychology uses cost-effectiveness analysis). That doesn't make them "not evaluation-specific" any more than statistics becomes psychology just because psychologists use it.

Unanswered Questions at the Strategic Level

The discussion thus far has been about the need to lift evaluation questions up from the indicator level and make them higher-level and explicitly evaluative. In other words, we need to ask and answer the most important questions about quality and value with respect to each program or project being evaluated.

But there is one more level evaluation needs to go to, and this is up to the strategic or policy level. Several years ago I outlined the key characteristics of genuine strategic evaluation:

A strategic evaluation asks (and answers) questions that go beyond program improvement or overall program quality, addressing the value of the program as a contributor to the broad strategic mix of initiatives and interventions. For example, to what extent does this particular initiative make a unique contribution to the strategic vision? How well does it fit with other initiatives in that respect? Are there any unnecessary overlaps? Is the initiative fundamentally consistent with the organization's overarching values? How much is it helping the organization make progress toward achieving its vision? (Davidson & Martineau, 2006, p. 438)

Several years later, Nan Wehipeihana and I developed this thinking further into two workshops on strategic policy evaluation (Wehipeihana & Davidson, 2010, 2011). We improved on the earlier list of high-level questions that often need to be answered for policy makers, but are seldom asked:

1. To what extent does this particular initiative make a *unique contribution* to the strategic vision?
2. Is the initiative fundamentally consistent with the organization's overarching *values*?
3. How well does each initiative *fit with and complement* the other initiatives that make up the strategic policy mix? Are there any unnecessary *overlaps*?
4. What is the *collective value* of the suite of initiatives to achieve a particular strategic outcome?
5. Have we got *the 'right mix'* of initiatives to deliver on the key strategic outcomes?

Why do such questions seldom get asked? The following observations are based primarily on my own context, New Zealand central government, so I welcome your reflections on the extent to which

they are also issues in the sectors and countries where you work.

A primary reason in my own context (New Zealand) is that evaluation is generally only commissioned for what we call "new money" projects, never for the billions of dollars' worth of work being performed under the so-called baseline budget, the ongoing funding for an agency's core business. As a result, these big ticket items are never put under scrutiny; only the new ideas. As a taxpayer, this strikes me as bizarre.

Another major barrier to evaluation being seen as a valuable contributor to high-level strategic decision making is that much of our work is commissioned at the program or project level. Evaluation budgets tend to be associated with projects and programs, and resource is seldom devoted to asking high-level cross-project strategic evaluation questions about policy. Instead, we tend to evaluate everything that moves, no matter how small, and we keep attempting to do so with the tiniest of budgets (Wehipeihana & Davidson, 2010, 2011). These shoestring evaluations have serious limits on the kinds of questions they can feasibly answer, and many of them turn up very basic findings that we already knew from evaluating similar projects. Time after time, the same important questions are left dangling because no single evaluation ever has the resources to tackle them.

Many higher-level cross-project questions are never addressed because evaluation budgets are not associated with an entire policy or strategy encompassing multiple interrelated components. And so it is rare to see evaluation seriously adding value to policy at this level. We never really get to ask and answer questions about, for example, the *collective value* of a suite of initiatives; whether we have the right mix to deliver on key strategic outcomes; the relative contributions of each policy instrument; and so forth.

Time for Evaluation to Step Up to the Plate

Clients and stakeholders who are not trained evaluators may be forgiven for wanting simple and direct answers to their questions. But too often they believe evaluation is simply a matter of picking a few indicators and measuring them or choosing some "gold standard" design that will miraculously give them all the answers.

Part of our job as professional evaluators is to guide our clients, expand their horizons, and broaden their expectations of what evaluation can deliver. In order to do that, we need to have the right mix of tools in our repertoires. If we don't, we end up limiting the questions we ask to just the ones we can answer with the methods we know.

If we stick with the applied social science research approach to evaluation, or the "just measure a few indicators" approach, the end result is that evaluation is less courageous, less important, and less valuable than it might be. In many circles it is seen as irrelevant, trivial, lost in the details, the work of backroom boffins with no real sense of what policy makers and other decision makers really need to know.

It's time for evaluation to step up to the plate. What does this mean, specifically?

Delivering on the Essence of Evaluation – the Real Evaluative Stuff

We need to add evaluation-specific methodology to our repertoires; purge value-free thinking from our minds; get a large and powerful dose of the evaluative attitude; and get a grip on our evaluation anxiety. We need to face the fear and do it anyway, because that's our job, and no-one else is better equipped.

When we finally do step up to the evaluative plate, what we can deliver is of far higher value to programs, their management, their staff, and their recipients. Decision makers can't get to the "now what?" (action planning) simply on the basis of knowing "what's so?" (just the non-evaluative facts about what happened). In order to identify what needs attention and to prioritize what's urgent, they also need the "so what?" (How well have we done so far? Where, for whom, and why are the outcomes truly excellent and where are they most disappointing? Which are the most important and urgent weaknesses in design and implementation?).

Delivering Better at the Strategic and Policy Level

To step evaluation up to the strategic evaluation plate, we need evaluation budgets to be attached to multi-project strategies covering multiple initiatives seeking to realize the same broad strategic intent. Next, we need cross-project strategic policy evaluation questions that are explicitly evaluative.

We will also need better opportunities for evaluators to build skills in strategic policy evaluation. This includes lifting thinking up from program logic to policy logic and forming a strategic evaluation framework around that. Other

useful tools include the development of broad cross-project evaluative rubrics that can be used as a shared language for reporting results based on different kinds of evidence but in a form that can be synthesized.

Keeping it Simple – but not Simplistic

Evaluation faces challenges on both sides – forces that lead us toward more sophisticated and complicated evaluation designs, and forces that push us to simplify. The trick for truly adding value is to balance these two, guided with clarity of purpose and delivered with clarity of data visualization and communication.

Eleanor Chelimsky (2011, 2012) has made some important observations about the prevalence of single narrative thinking in both policy design – and the domino effect on evaluation. Narrow policy thinking leads to narrow evaluation questions and narrow requirements for the methods used to answer those questions. When powerful clients dictate terms like this, it is extremely difficult for contractors to push back because the cost of doing so is likely not being contracted at all.

My own experience has also been that some of the most unrewarding pieces of evaluation work are those that are already loaded with specifications for approach and methodology at the Request for Proposal stage. The evaluation team simply isn't given enough leeway to ask or answer what is actually most needed, and it is hard to negotiate for additional scope.

It is my own view – and observation – that a lot of this single narrative thinking and the resulting unidimensional evaluation questions are based at least partially in limited thinking about methods. Policy makers may narrow the evaluation agenda for political reasons, true, but they also tend to write RFPs based on what they believe is possible methodologically, around the questions they believe are answerable. Clients I work with often have no idea that it is possible to develop defensible answers to much higher-level evaluation questions. They simply have never seen it done.

Layered on top of these forces pulling us toward oversimplification are considerable forces in the direction of overcomplication. One is particular to those who seek to publish some of their evaluation work, or have it considered as part of academic performance review. Evaluators who work in an academic space will naturally pay attention to what top tier journals are looking for, which is often work that demonstrates methodological sophistication and prowess.

This problem is by no means unique to evaluation. The American Psychological Association's Task Force on Statistical Inference noted the tendency for researchers to choose sophisticated and complicated analytic methods to impress other academics rather than more elegant, simpler approaches that would be more appropriate and the findings easier to communicate (Wilkinson & The Task Force on Statistical Inference, 1999). It is not just ease of communication that is the issue; it is the potential of the work to be impactful:

> "[T]he evidence of history is clear that the research studies with the greatest impact in psychology are *breathtakingly simple* in terms of the questions posed, the methods and designs used, the statistics brought to bear on the data, and the take-home messages" (Peterson & Park, 2010, p. 398, emphasis added).

This is something that communication and data visualization experts have known for years. They continue to urge academics and others to improve their communication of findings so that they are more likely to be used (e.g., Evergreen, 2013; Hayes & Grossman, 2006; Heath & Heath, 2008; Morris, Fitz-Gibbon, & Freeman, 1987).

The programs and policies we evaluate are often complicated and/or complex, and there are many nuances in the questions we might ask about them. It is tempting to opt for very sophisticated designs - but these are what put us in danger of becoming overwhelmed with the details and unable to pull out the "knowledge nugget" or key insight decision makers really need. It is true that many evaluations require gathering huge amounts of information. The real problem is that our deeply detailed answers glaze over the eyes of our stakeholders. As a profession, I am afraid we are not too good at the synthesis task – pulling it all together (Davidson, 2005; Scriven, 1994).

There are four things we need to do to step up to the plate here. The first is to keep our eyes on the high-level questions needed to guide the evaluation and stop it from getting lost in the details. Second, we need to opt for the simplest possible approaches we can use to get the answers we need to the required level of certainty. Third, we need to get much better at synthesizing values and evidence from multiple different sources to draw clear evaluative conclusions and deliver direct answers to the evaluation questions. And finally, we need to use truly outstanding data

visualization and reporting strategies to convey the key messages clearly and succinctly

.

References

Canadian Evaluation Society (2010). *Competencies for Canadian evaluation practice*. v. 11.0 4 16 2010.

Chelimsky, E. (2011). *Evaluation and the single narrative*. Presentation at the meeting of the American Evaluation Association.

Chelimsky, E. (2012). Values, evaluation methods, and the politicization of the evaluation process. *New Directions for Evaluation, 133,* 77-83.

Davidson, E. J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Thousand Oaks, CA: Sage.

Davidson, E. J. (2007). Unlearning some of our social scientist habits. *Journal of Multidisciplinary Evaluation, 4*(8), iii-vi. http://jmde.com

Davidson, E. J. (2010). "Process values" and "deep values" in evaluation. *Journal of Multidisciplinary Evaluation,* 6(13), 206-208. http://jmde.com

Davidson, E. J. (2012). Actionable evaluation basics: Getting succinct answers to the most important questions [minibook]. Real Evaluation.

Davidson, E. J. (2013). Evaluation-specific methodology: The methodologies that are distinctive to evaluation. Genuine Evaluation blog. http://genuineevaluation.com/evaluation-specific-methodology-the-methodologies-that-are-distinctive-to-evaluation/

Davidson, E. J., & Martineau, J. W. (2006). Strategic uses of evaluation. In J. W. Martineau, L. Merritt, & K. Hannum (Eds.), *Leadership development evaluation handbook* (pp. 433-463). San Francisco, CA: Jossey-Bass.

Evergreen, S. D. H. (2013). *Presenting Data Effectively: Communicating Your Findings for Maximum Impact*. Thousand Oaks, CA: Sage.

Hayes, R., & Grossman, D. (2006). A scientist's guide to talking with the media. Piscataway, NJ: Rutgers University Press.

Heath, C., & Heath, D. (2008). Made to stick: Why some ideas take hold and others become unstuck. [Kindle edition] Cornerstone Digital.

King, S., & Davidson, E. J. (2012). *Making the complex doable: Practical evaluation frameworks that work*. Mini workshop presented at the Aotearoa New Zealand Evaluation Association conference, Hamilton.

Morris, L. L., Fitz-Gibbon, C. T., & Freeman, M. E. (1987). How to communicate evaluation findings. Thousand Oaks, CA: Sage.

Patton, M. Q. (2012). Contextual pragmatics of valuing. *New Directions for Evaluation, 133,* 97-108.

Peterson, C., & Park, N. (2010). Keeping it simple. The Psychologist, 23(5), 398-400.

Sathe, V., & Davidson, E. J. (2000). Toward a new conceptualization of culture change. In N. M. Ashkanasy, C. P. M. Wilderom, & M. F. Peterson (Eds.), *Handbook of organizational culture and climate* (pp. 279-296). Thousand Oaks, CA: Sage.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Thousand Oaks, CA: Sage.

Scriven, M. (1994). The final synthesis. *Evaluation Practice, 15*(3), 367–382.

Scriven, M. (2003). Evaluation in the new millennium: The transdisciplinary vision. In S. I. Donaldson & M. Scriven (Eds.), Evaluating social programs and problems: Visions for the new millennium. Mahwah, NJ: Routledge.

Scriven, M. (2006). *The real evaluation competencies.* Personal communication from Michael Scriven; also posted on the EVALWMU listserv, accessible at http://evaluation.wmich.edu/scripts/wa.exe?A2=ind0603&L=eval-wmu&F=&S=&P=2160

Scriven, M. (2013). *The good old days and the schizophrenic break in the history of evaluation*. Claremont Evaluation Center webinar, accessible at: http://www.cgu.edu/pages/10257.asp

Scriven, M., Davidson, E. J., & King, S. (2012). *Biting the evaluative bullet.* Demonstration session at the meeting of the American Evaluation Association, Minneapolis, MN. Online: http://comm.eval.org/EVAL/Resources/ViewDocument/?DocumentKey=e3a5bfb4-22f1-4f87-a139-2984da097a22

Stevahn, L., King, J. A., Ghere, G., & Minnema, J. (2005). Establishing essential competencies for evaluators. *American Journal of Evaluation, 26*(1), 43-59.

Wehipeihana, N., Bailey, R., Davidson, E. J., & McKegg, K. (in press). Values and culture at the core: New Zealand's evaluator competencies. *Canadian Journal of Program Evaluation.*

Wehipeihana, N., & Davidson, E. J. (2010). *Strategic policy evaluation: Answering macro-level cross-project questions*. Workshop presented at the Aotearoa New

Zealand Evaluation Association Regional Symposium, Wellington, New Zealand.

Wehipeihana, N., & Davidson, E. J. (2011). *More strategic policy evaluation: Answering macro-level cross-project questions*. Workshop presented at the Aotearoa New Zealand Evaluation Association conference, Wellington, New Zealand.

Western Michigan University (2004). *Interdisciplinary PhD in Evaluation Competency Assessment*. Kalamazoo, MI: The Evaluation Center.

Wilkinson, L., & The Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594-604.