

---

# WHAT WILL WORK HERE

---

## INFERENCES FROM EVALUATIONS IN COMPLEX ECOLOGIES

Robert Klitgaard

Claremont Graduate University

American Evaluation Association Annual Meeting, Oct. 27, 2012

In our rapidly changing world, many policy problems and possible solutions are in flux. History is happening all around us, with changing politics, shifting social mores, and new institutions. New technologies are creating new alternatives and enabling new levels and forms of participation. In issues ranging from public health to environmental protection to economic transformation, multiple actors from government, business, and civil society will have to collaborate in the improvements we need. Evaluating *what will work here* often takes place in complex ecologies.<sup>1</sup>

In such settings, evaluation needs broadening and deepening. Beyond “the study” for “the decision” by “the policymaker,” a central activity of evaluation may be the convening of multiple actors to understand their complex ecologies, define issues, reconsider objectives and alternatives, digest promising practices, and forge new relationships. A *convening* brings together

- facts, examples, and frameworks from outside with
- local knowledge and creativity.

One hoped-for result: better inferences about what will work here.

Ten years from now, convenings may be the grist of evaluations in complex ecologies.

---

## THE EXAMPLE OF CORRUPTION

---

Let me give a brief example of evaluation in a complex ecology. In June 2010, Benigno Aquino III was elected president of the Philippines. His campaign slogan was *Kung walang korap, walang mahirap*—“When there is no corruption, there will be no poverty.” His inaugural address focused on fighting corruption. Five weeks after he took office, an all-day cabinet meeting was convened to discern implications from international experience about “what will work here.”

---

<sup>1</sup> Annex 1 contrasts complex policy ecologies with more routine settings for policy choice. The two settings create different challenges and opportunities for evaluation.

How might we think about roles for evaluation in such a meeting? Typically, corruption is not approached analytically. People respond almost reflexively by declaring that corruption is a problem of morals and cultures, sometimes with the fatalistic implication that nothing can be done in the short run. What is more, because corruption is entwined with idiosyncratic local politics, social structures, economic conditions, and traditions, people may conclude that trying to discern lessons from elsewhere is useless.

And yet, we know a lot about corruption and how to reduce it. International data, such as the Corruption Perceptions Index of Transparency International and the multiple indicators of governance in the World Economic Forum's Global Competitiveness Report, can be evaluated statistically and improvements linked to outcomes such as investment, job creation, and growth. Case studies explore how various cities, ministries, and countries have successfully reduced corruption. Theoretical models and practical experience highlight two features of most forms of corruption:

1. The economic choices of the bribe-givers and bribe-takers are a function of the risks and rewards they face.
2. Those risks and rewards are themselves a function of the institutions that audit, enforce, prosecute, and judge corrupt transactions.

From these points "a checklist for policymakers" has been devised. It delineates such big alternatives as selecting agents, enhancing incentives, improving information flows, and raising the moral costs of corruption. Many policymakers have found this metaphorical formula useful: Corruption equals Monopoly plus Discretion minus Accountability (Klitgaard 1988).

These pieces of analysis—facts, examples, and frameworks—can form the basis for a convening. In Manila, at an all-day meeting on a Saturday in August 2010, the Philippines cabinet diligently considered facts about the extent and social costs of various kinds of corruption in the Philippines and elsewhere. In the style of a two-part business school case, they analyzed a country that had successfully reduced corruption. They worked through a framework for policy analysis. The problems of corruption are complex; the "ecology" of the Philippines is unique. And yet, the data, the case study, and the theoretically informed framework became, with these participants in this process, a catalyst for creativity. By the evening, they devised an outline of a national strategy, which they refined on Monday and briefed to the President on Tuesday. A plan of action followed—actions that so far have been successful in raising investment, improving public services, enhancing public trust, and increasing the popularity of the president.

In many other cities, ministries, and countries, I have seen such convenings-with-evaluation-research make dramatic differences in the way leaders look at corruption (a crime of calculation, rather than [just] a moral failing), in the alternatives they consider, in the partnerships they pursue, and in the choices they make.

And note this: in none of these "dramatic differences" was there a clear line connecting an evaluation to the decisions actually taken in the short- and long-run by the relevant

policymakers and stakeholders. The logic of use and usefulness did not run from a study with powerful recommendations to their acceptance by an authoritative decisionmaker.

Rather, as the result of a process of encountering statistical and economic analysis of data about their country in comparative context, case studies from other countries, and a framework for policy analysis, policymakers and stakeholders engaged in creative problem-solving, forged new relationships, and undertook concrete initiatives that I believe no outside analyst would have been able to conjure up, much less turn into action.

### COMPARATIVE ADVANTAGE IN EVALUATING WHAT WILL WORK HERE

Like corruption, many problems are complex and ecological, with too many causal loops and intervening variables to make us confident that what we learn from an evaluation of a problem in one setting can be transferred straightaway to a seemingly similar problem in a quite different setting. We often have theory uncertainty, measurement uncertainty, and modeling (or econometric) uncertainty, coupled with samples of countries and ministries and cities and projects that are too small to explore potentially important interaction effects.

What to do? How can knowledge about the relevant complexities of the local ecology be evoked and applied? Locals have tacit knowledge about local settings. But they often lack access to the best models and theories, the best data and techniques of data collection, and the appropriate comparisons for them to calibrate their challenges. They may lack examples of what has worked elsewhere, not so much to copy as to inspire. The challenge becomes how to combine forces: how to bring what they know best (local objectives, constraints, alternatives, intervening variables, etc.) with what outsiders may offer (facts, examples, frameworks).

*Convening* tries to bring these forces together.<sup>2</sup> Those convened have different if overlapping objectives, different if sometimes overlapping capabilities, and different if overlapping information about the state of the world and about if-then relationships (such as treatment effects). The stakeholders are strategically connected, in the sense that what one party does often affects the outcomes of what other parties do. They are not fully aware

---

<sup>2</sup> Convening, in fact, is a term of art in the literature on consensus building and dispute resolution. Its meaning there is more limited than the one I have in mind here. "In this context, convening typically involves

1. Assessing a situation to determine whether or not a consensus-based approach is feasible;
2. Identifying and inviting participants to ensure that all key interests (i.e., stakeholders) are represented;
3. Locating the necessary resources to help convene, conduct, and support the process; and
4. Planning and organizing the process with participants, or working with a facilitator or mediator to do so.

It may be helpful to think of convening as Phase 1 in a consensus building process, which is followed by Phase 2, the actual negotiating or consensus building phase" (Carlson 1999: 169).

of each others' objectives, capabilities, or information sets; they do not fully understand their strategic interrelations.

The kinds of convenings of interest here are *those that build on evaluation*. In particular, I have experience with research-based convenings that provide stakeholders with:

- Data, especially data that helps people “get on the same page” about the nature of the problems, if’s and then’s, funding, and costs. Data-rich discussions, in my experience, also help build trust, particularly about controversial issues where someone may be suspicious of being sold an ideological or political argument.
- Examples of success in similar problem areas, which spotlight goals, alternatives, if-then relationships, and partnerships. These examples are based on an evaluation of what they achieved and theory-based speculation about how.
- Frameworks for understanding goals, alternatives, if-then relationships, and/or strategic interdependence. A framework may be a grand theory, a program theory, or a heuristic. The framework draws upon social science, policy analysis, and evaluation.

---

## WHAT CONVENINGS CAN DO

---

Convenings help evaluations become practically meaningful, for seemingly standard evaluations as well as those taking place in complex ecologies. One may know the quantitative results of the evaluation; one may have assimilated the analyst’s recommendations. But what do the results mean to the various people who are involved in the system or who will have to make use of the results in their decisions and practices?

For example, suppose that in an educational intervention, an evaluation shows that the crucial success factor is parental participation, measured in two ways, a time variable (hours/month) and an intensity variable (“how involved did you feel?”). What do officials, parents, and students say about this result? What meaning do they give it? What examples of changes in someone’s participation do they see as resulting in large, small, or no changes in the student’s or program’s later success?

Second, what are the implications of the evaluation? Continuing with the previous example, participants can be asked, “What experience do you have with ways to increase parental participation (time, sense of involvement)?” Someone may suggest an idea that seems to have worked in one context; someone else may second that idea; and a third person may say, “But that didn’t work in our school.” The ensuing discussion may help fine-tune what it is that works—and suggest interactions with treatments and settings.

Third, what should we together do now? No doubt people will wish to know even more (especially the people who carried out the evaluation!). We may push for more than research and discussions about research. We may also move from “these changes or treatments would be desirable” to “what would they cost?”, “who would have decide what,

do what to make them happen?”, and “therefore, what should we do now to explore these possibilities?”

Let me provide good news and bad news about convenings. The good news is that I have seen convenings work.

The bad news is that those who do convenings, including me, have not evaluated convenings with the same rigor and usefulness we would expect from, say, a program evaluation. (Annex 2 provides some examples from a much longer review of many forms of group problem-solving.)

For one thing, those who are gifted in facilitation, say, are not always motivated to evaluate. For decades, Herbert C. Kelman of Harvard organized problem-solving workshops to help influential people from Israel and Palestine work together defining issues, understanding obstacles, and overcoming the obstacles. The goals were “to produce changes—in the form of new insights and new ideas—in the workshop participants; and to transfer these changes into the political process and the political culture of the two societies” (Kelman 2008: 44).

What did Kelman’s workshops accomplish? “The time has come for me to confess that, in the more than 30 years that I have been engaged in developing interactive problem solving, building its theoretical foundations, and practicing it, I have not engaged in systematic research designed to evaluate its effectiveness.” (Kelman 2008: 39)

Even with the best of will, evaluation would not be easy. Convenings like Kelman’s take place in complex ecologies that change over time. An evaluation of the effects of his workshops would face another version of the evaluation problem described above.

More generally, if we were trying to figure out what kind of convening might work on a certain kind of problem in our particular setting, we could not simply copy what Kelman (or anyone else) has done; and we would be unwise to say, “It all depends, we can’t learn anything.” We would want to garner what has been learned from experience elsewhere with convenings and digest it with regard to our particular situation.

We would be asking a version of “What will work here?” To create practical, relevant answers, we may need ... a convening.

---

### A CONVENING ABOUT CONVENINGS?

---

Let me ask your advice about how we as evaluators might learn more about convening. Annex 2 describes some disappointing results in evaluating what sorts of convenings work how well (defined how?) in what “complex ecologies.” Suppose we took a different tack. Imagine now that we and others interested in evaluation could bring in experts of various kinds to consider what kinds of convenings work where. Might we have a convening about convenings?

Here are some beginning ideas about what such an event might do.

First, we would assemble and consider some relevant facts. What kinds of convenings do people carry out on what kinds of problems? Consider workshops, multi-stakeholder groups, group decisionmaking, even conferences. How might we characterize these convenings? What is the nature of the problem? What are people trying to learn exactly? Who are the participants? What is the decision, or is it a flow of decisions, a partnership to be formed? What are the roles of evaluation?

Then, we would consider examples. I am partial to success stories (it is all too easy to identify failures and explain them away). We would consider it in the fashion of a two-part business-school-style case. In part A, we face the problem those in the case confronted. What kind of convening would we design, with what facts, examples, and frameworks; what participants; and so forth? Then in part B, we would see what happened in the case and how. In the ensuing discussion, we would extrapolate the example to situations we know well, and learn from each other.

Then, we would consider frameworks relevant to the design and management of convenings. We might play with a framework like mine, with facts, examples, and frameworks (for an example, see Klitgaard 2010). We might employ the framework of Raoul Blindenbacher (2010, see Annex 2 below) or the work of Michael Quinn Patton (2011, and his 2012 online AEA workshops at the beginning and intermediate levels).

The combination of facts, examples, and frameworks would enable us to learn from experience elsewhere and from each other, and then to reconsider how convenings might help people improve evaluation in complex ecologies.

Enough for now. Let us turn to commentators Stewart Donaldson and Tom Schwandt for their thoughts, and then ask you for your ideas about convening as a method for evaluating what will work here.

## ANNEX 1. STANDARD EVALUATION VS. EVALUATION IN COMPLEX ECOLOGIES

---

Category	Standard	Complex Ecologies
Policy problem	Well defined	Poorly understood
Objective function	Given	Murky
Alternatives	Given	Perhaps incomplete
Program theory (or production function)	Well understood	Poorly understood
Information	Appropriate data on inputs, alternatives, and outcomes are available and valid. Understood process of data generation.	May be abundant but a bush; may also be inadequate in many ways. Process of data generation poorly understood.
Context	Standard.	Highly variable and perhaps complex.
Decisionmaker	Single. One clear decision point.	Multiple. No single decision or even clear decisions.
Relationship between client and evaluator	Arm's length, emphasizing objectivity.	Trusting, emphasizing problem clarification and learning in both directions.
Incentives	Decisionmaker's incentives are aligned with organization's and nation's objectives.	Incentives and accountability may be weak or even absent.
Feedback loops	If evaluation is used and valid, decisionmaker and evaluator will eventually know this.	Weak or absent feedback on decisions.

Source: Author. See also Patton (2011).

## ANNEX 2. EXAMPLES OF RESEARCH ON VARIOUS KINDS OF CONVENINGS

---

In evaluation generally, experts lament how few well-documented evaluations of evaluation exist, even when excellence is defined in terms of craft and not impact.<sup>3</sup> Similarly, in the varied and non-overlapping literatures on group problem-solving—from conferences to workshops, from meetings to consensus building—I have encountered plenty of prescriptive advice and how-to lists but few evaluations or careful case studies.

What positive outcomes (and negative ones) occur for what kinds of convenings (done how) with what populations in what settings?

Judith E. Innes aptly described the challenge.

Those who study public policy making and organizational development need to understand how and why consensus building is similar to or different from the alternatives, develop theory and evidence to define *best practices*, and establish when and why consensus building is (or is not) successful. Evaluations can help meet all of these needs. (1999: 632)

In principle, yes; but in practice, evaluations haven't met these needs, at least not yet. "The evaluations of consensus building conducted to date have not adequately assessed the full set of long-term and secondary effects..." (636), nor were the variety of settings and indeed treatments taken into account.<sup>4</sup>

One of the only systematic evaluations had, in Innes' judgment, fallen victim to selection effects. Cary Coglianese (1997) found that a consensus-building variant called "negotiated rulemaking" produced federal regulations more slowly than the traditional method and engendered the same amount of later litigation (a proxy for the clarity and quality of the regulation). Innes plausibly notes but does not prove that the instances where consensus-building techniques were used may have involved more difficult and contentious regulatory issues, which would make Coglianese's conclusions a problematic guide to the choice of methods (638).

---

<sup>3</sup> Chris L. S. Coryn, Lindsay A. Noakes, Carl D. Westine, and Daniela C. Schröter reviewed twenty years of "theory-driven" evaluations from 1990 to 2009. One of their conclusions: with only a few exceptions, "very little empirical evidence exists to buttress the numerous theoretical postulations and prescriptions put forth for most evaluation approaches, including theory-driven forms of evaluation. Yet, for many years, evaluation scholars have urged the evaluation community to carry out empirical studies to scrutinize such assumptions and to test specific hypotheses about evaluation practice" (2011: 215).

<sup>4</sup> By 2008, one author could lament too much emphasis on the "process effects" and not enough on the quality of the resulting policies or decisions. "The bias toward evaluating the process and its social outcomes has resulted in a gap in knowledge of the impact collaborative environmental planning and management has on changing environmental conditions." (Mandarano 2008: 456).



Just as treatments vary, even when putatively identical, so do convenings vary. The quality of the conveners—and the participants—may determine the effect of the treatment on the outcome. Lawrence E. Susskind, Boyd W. Fuller, Michele Ferenz, and David Fairman (2002) reviewed the effects of “multistakeholder dialogue at the global scale.” “While there are examples of successful MSDs contributing to official policy-making, too many multi-stakeholder dialogues founder because the participants are inadequately prepared, the processes are managed ineffectively, and expectations are unrealistic” (abstract). Innes (2004) asserted that if done well, convenings could yield positive results. But she noted that if done badly, they might result in unintended consequences, even negative net effects. For example, consensus building would lead to “lowest common denominator solutions” and, because of power imbalances, apparent agreements would be fleeting. Robert Deyle and Carissa Schively Slotterback (2009) reviewed the challenges of estimating “group learning in participatory planning processes,” without yielding reliable lessons or rules of thumb. Rory Truex and Tina Søreide (2011) studied “multi-stakeholder groups” in oversight and implementation and concluded that the barriers to success are substantial:

The laundry list of potential challenges includes: poor participation among members due to time constraints or conflicts of interest; problems reaching consensus on key decisions; imbalances of power and capacity across stakeholder groups; a lack of broader social and political legitimacy; difficulties obtaining needed inputs; and insufficient time as a result of external deadlines. (479)

---

### HERBERT KELMAN ON PROBLEM-SOLVING WORKSHOPS

---

For decades, Herbert C. Kelman organized problem-solving workshops to help influential people work together defining issues, understanding obstacles, and if time remains, to overcoming the obstacles. The goals were “to produce changes—in the form of new insights and new ideas—in the workshop participants; and to transfer these changes into the political process and the political culture of the two societies,” which in turn would lead to the resolution of a dispute (Kelman 2008: 44). The key features of his process were “(a) confidential dialogue, (b) facilitated discussion of underlying needs and fears, and (c) joint problem solving by the parties to the dispute.”

Kelman’s focus was the Palestinian conflict. Year after year, he invited citizens (not policymakers or political leaders) from Palestine and Israel to his workshops. His method structured their interactions in a distinctive way.

To break the ice, participants would share information. “Each party is asked to talk about the situation on the ground and the current mood in its own community, about the issues in the conflict as seen in that community, about the spectrum of views on the conflict and its solution, and about participants’ own positions within that spectrum.” (Kelman 2008: 35)

Then the participants interacted in four segments:

1. Clarifying goals. Each side discusses “the fundamental needs that would have to be addressed and the existential fears that would have to be allayed if a solution is to be satisfactory to them.” (35)
2. Imagining possible alternatives. “What participants are asked to do in this phase is to develop, through an interactive process, ideas about the overall shape of a solution for the conflict as a whole or, perhaps, a particular issue in the conflict, that would address the needs and fears of both sides. They are given the difficult assignment of thinking of solutions that would meet not only their own side’s needs, but the needs of both sides.” (35-36).
3. Identifying obstacles. “Discussion of the political and psychological constraints within the two societies that would create barriers to carrying out the ideas” (36)
4. Overcoming obstacles. “Finally, depending on how much progress has been made and how much time is left, we ask the parties to engage in another round of joint thinking—this time about ways of overcoming the constraints” (36)

Note that the process did not incorporate policy analysis or evaluation. The relevant evidence was about perceived feelings and needs, and it was provided by the participants themselves. Participants did not consider together case studies of solutions in other cases of international conflict.

Nor did Kelman and his colleagues use the workshops’ results to create better evaluations and policy analyses. They did not assemble the goals, solutions, obstacles and ways to overcome the obstacles so that others could learn from the “creative problem solving.” They take the analyses of many workshops over the years about the “same” conflict, to see for example how the participants views of needs and alternatives and constraints changed over time.

In both these aspects—not using policy analysis and evaluation to help the participants, and not chronicling the emerging insights and policy proposals that emerged from the workshops—Kelman’s problem-solving workshops differ from the convenings discussed here.

What did Kelman’s workshops accomplish?

“The time has come for me to confess that, in the more than 30 years that I have been engaged in developing interactive problem solving, building its theoretical foundations, and practicing it, I have not engaged in systematic research designed to evaluate its effectiveness.” He blames this lack on “the inherent difficulties in applying standard evaluation models to this case, to the ethical and methodological issues raised by intrusive evaluation procedures, and to my own need to set priorities.” (39)

He can imagine how his workshops might be evaluated, even though he and his colleagues never did it:

Thus, for example, if workshop participants enter into the discourse of interactive problem solving, manifest changes in attitude, inject their new ideas into the

political process, and then become personally involved—or advise others who are involved—in decision making and negotiation, we have some justification for concluding that the workshop experience, along with many other inputs, made some contribution to the agreement that is finally negotiated. This, I propose, is the closest that evaluation research can come to demonstrating the impact of unofficial efforts on the macro-process of conflict resolution. (51)

And so, Kelman concludes, this particular version of convening was plausibly successful but one cannot be sure. “Problem-solving workshops have two operational purposes: to produce changes, in the form of new insights and new ideas for resolving the conflict, in the individuals who participate in the workshops; and to transfer these changes into the political debate and the decision-making process in the two societies. This dual purpose is related to the ultimate goal of contributing to the achievement of a durable, high-quality negotiated agreement. The contribution of interactive problem solving to the achievement of this ultimate goal, however, cannot be demonstrated in any definitive way... The precise contribution of workshop products to the ultimate political decisions and negotiation outcomes can generally not be traced in detail” (54).

---

### MARGARET MEAD ON CONFERENCES

---

The goals of Kelman’s events contrast with the “small conferences” described by Margaret Mead: to produce new ideas that participants would not have had on their own, ideas that after the workshop might be disseminated widely. “In the conference the kind of intellectual communication which depends upon the linearity of script is also eliminated ... The ideas that grow under such conditions are different from the ideas which any participant would have had working alone, or even working with one or two colleagues” (Mead & Byers 1968: 6).

Mead’s small book (with photographer Paul Byers) about small conferences presents her idiosyncratic lessons from thirty or more years of conference-going. She focuses on whom to invite, where to meet, who should be the chair, and the characteristics of participants. For example, on the facilitator or chair she recommends:

The choice of a chairman is closely bound in with the whole conference style. Individuals who make fine parliamentary-type chairmen, or fine chairmen in a policy-making group, and fine impartial chairmen of groups charged with hostility and factionalism are not, except in rare and probably accidental cases, ideal chairmen for substantive conferences. The conference chairman must be committed to advancing the ideas that are developing, not to being fair, or to establishing a reputation for impartiality, or to demonstrating a good memory for the sequence in which people have asked for the floor. Instead he has to attend to the direction of the thinking, hold in mind points made earlier, dropped and now recurring, and sense which of the proffered contributions will advance the flow of thought... He has to be able to weave together—in summaries and interpolations—the strands of the discussion, conscious of the pattern, and yet willing to leave the

shape of the next patterned unit open and undetermined. He must be unconcerned with such points of parliamentary etiquette as the circumstances under which a chairman should speak and be able to relate his behavior to other types of roles, such as that of discussion leader or clarifier. At the same time he must have the authority to quell attempts to disrupt the conference atmosphere by those who seek to convert the meeting to some other style of group behavior such as majority rule, consensus, or *laissez faire* free discussion, or to establish an agenda. It takes, if anything, more skill to prevent habitual but inappropriate types of behavior from reasserting themselves than it does to keep a group working along established lines (18) ... He should have no extraneous motives, no *hidden agenda* as these have come to be called in group research jargon, and no notion that he knows how the discussion is going to turn out ... Such chairmen are exceedingly rare, and seldom if ever self-selected (19).

For organizers of conferences, Mead provides many down-to-earth recommendations about food, drink (it should be readily available to avoid “glut and famine phenomena”), and luxury (facilities should be very luxurious or, for the right cause, Spartan).

Although she does cite, in footnotes, disguised examples of disastrous conferences, her evaluation of what works does not present examples of excellent conferences leading to important insights. Nor are her often-amusing generalizations, which may still have practical value, supported by anything like formal evaluations.

### “THE BLACK BOX OF GOVERNMENT LEARNING”

---

Raoul Blindenbacher began organizing “learning events” for various levels of government in Switzerland. Later, at the World Bank, he organized “about 150 documented learning events...in all sizes and shapes” (2010: 71). Lots of methods were used (“a full range of different models”): “their compilation shows a kaleidoscope of theory-based disciplines, which were used to describe and understand how governments learned” (71). The rationales of the different concepts were analyzed using “qualitative content analysis”: “Following this methodology, the deducted rationales were clustered around precisely defined subject matter and thematically organized around and sequenced into eight distinctive stages” (72). This analysis and its results are not described in any detail in the book or in any referenced publication. The eight stages are:

1. Conceptualization [what knowledge is needed, and what knowledge is available?],
2. Triangulation [multiple viewpoints must be represented and evoked],
3. Accommodation [people are made to feel comfortable],
4. Internalization [reflect on their own experiences in light of the new knowledge],
5. Externalization [share their individual reflections with other participants “in a natural and unobtrusive way”],

6. Reconceptualization [“overlapping reflections” evolve into “a new collective perception” and “a new normative frame of knowledge”],
7. Transformation [action plans are created “to fit the requirements of a given political reality”], and
8. Configuration [written output of results].

“When these stages are performed, a didactical process is established that encourages behavioral change in governmental institutions, their members, and representatives from involved nongovernmental organizations and interest groups” (xviii).

1. The conceptualization, triangulation, and accommodation stages are considered as the preparatory stages, where the knowledge to be learned is framed, the selection and invitation of the participants is completed, and an initial bond and a sense of trust between the learning actors and the event facilitator and between participants and the learning process is established.

2. The internalization, externalization, reconceptualization, and transformation stages represent the core of the didactical procedures, where the learning actors review and adapt the new knowledge according to their personal needs. Thereafter the actors change their individual and organizational thinking and behavior in an elaborate inter- and intrapersonal procedure accordingly.

3. The follow-up to the learning activity is organized in the final configuration stage, where all developed knowledge is made available and accessible to everybody involved in the learning activity as well as to a wider audience. This new knowledge further serves as the knowledge frame of the next spin of the Learning Spiral, as well as a feedback loop in the context of a new learning system. (85)

The design is based on an analysis of the knowledge to be learned, usually state-of-the-art or evaluation-based knowledge regarding the issue at hand. The design also takes into account the political environment where the event is taking place, and it requires a deliberate selection of the individual learning actors and the governmental and nongovernmental institutions involved. (xix)

The effects of the applied Learning Spiral are threefold: The primary effect is that governments’ access to the latest knowledge in democratic governance is enhanced and can be applied in concrete, practical action. A second effect is that—because of the iterative character of the learning process—the knowledge to be learned is always validated and updated in real time to include the latest existing experiences on the subject. And a third effect is that participation in the learning process evokes a sense of social belonging among the learning actors, which often leads to the creation of social networks, where governments continue to share their latest experiences and by doing so launch the next spin of the Learning Spiral. (xix)

Learning theories suggest:

First, a safe and comfortable learning environment has to be established to give the learning actor the opportunity to experience new knowledge, which may awaken his or her curiosity and interest. Second, the learner's motivation has to be further enhanced by setting up a deliberate didactic process in which he or she becomes rationally aware of the difference between the known and the unknown knowledge, as well as the importance of overcoming this knowledge gap. Third, the learning should be done in a social context, where the learning actor is exposed to different peer perspectives. This way he or she gets an opportunity to choose from a variety of different explicit or tacit knowledge and to think about how it may best fit into his or her individual social reality. And fourth, if actors are exposed to positive incentives and rewards, their motivation to engage in a learning process increases and supports the intended learning outcomes—a change of thinking and an intended change of behavior. (62)

Who should participate?

The relevant stakeholders of each unit need to be selected to make sure that the learning activity considers the different organizational perspectives as well as the existing governmental power structure. This strategic selection and representation of learning actors is important for the subsequent implementation of the envisaged policy changes. It is expected that these stakeholders will act as opinion leaders and support the information and persuasion process of individuals inside and outside of government who were not involved in the learning process. The overall goal is the support of existing change agents to help them ensure the implementation of the envisaged political measures, within the existing constitutional and legal framework. (66)

Learning from each other can be valuable, but one has to avoid negative dynamics. “Its practical application has to establish a knowledge-sharing dynamic, which is based on confidentiality, mutuality, noncompetitiveness, inclusiveness, partnership, truthfulness, and willingness.” (24)

What is known about how to design such convenings? Not much. “To date the contemporary literature offers few theory-based concepts to enhance participatory and interactive governmental learning” (33). And like the other authors we have briefly considered, beyond short-term evaluations by participants, impact evaluations are nonexistent. “Though the Learning Spiral has been developed and applied over the last decade, it has yet to be rigorously evaluated and grounded in a *results framework*.” (88)

\* \* \* \* \*

These deficiencies in evaluation are not inherent. They are signs of an emerging problematic. As the importance of convenings and their ilk grows over the next decade or two, they will generate more analyses and evaluations. For progress to occur, we may have to add to the usual evaluation research something like convenings ... about convenings.

## REFERENCES

---

- Blindenbacher, R. (2010) *The Black Box of Governmental Learning: The Learning Spiral—A Concept to Organize Learning in Governments*. Washington: The World Bank.
- Carlson, C. (1999) "Convening" in *The Consensus Building Handbook: A Comprehensive Guide to Reaching Agreement*, ed. Susskind, L., S. McKernan, and J. Thomas-Larmer. Thousand Oaks, CA: Sage: 469-497.
- Coglianesse, C. (1997) "Assessing Consensus: The Promise and Performance of Negotiated Rule Making" *Duke Law Journal* 46: 1255-1349.
- Coryn, C.L.S., L.A. Noakes, C.D. Westine, and D.C. Schröter (2011) "A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009" *American Journal of Evaluation* 32(2): 199-226.
- Deyle, R.E., and C.S. Slotterback (2009) "Group Learning in Participatory Planning Processes: An Exploratory Quasi-Experimental Analysis of Local Mitigation Planning in Florida." *Journal of Planning Education and Research* 29(1): 23-38.
- Innes, J.E. (1999) "Evaluating Consensus Building," in *The Consensus Building Handbook: A Comprehensive Guide to Reaching Agreement*, ed. Susskind, L., S. McKernan, & J. Thomas-Larmer. Thousand Oaks, CA: Sage: 631-676.
- Innes, J.E. (2004) "Consensus Building: Clarifications for the Critics," *Planning Theory* 3(1): 5-20.
- Kelman, H.C. (2008) "Evaluating the Contributions of Interactive Problem Solving to the Resolution of Ethnonational Conflicts," *Peace and Conflict*, 14: 29-60.
- Klitgaard, R. (1988) *Controlling Corruption*. Berkeley and Los Angeles: University of California Press.
- Klitgaard, R. (2010) *Addressing Corruption in Haiti*. AEI Working Paper on Development Policy #2. Washington, DC: American Enterprise Institute.  
<http://www.aei.org/papers/foreign-and-defense-policy/regional/latin-america/addressing-corruption-in-haiti/>
- Mandarano, L.A. (2008) "Evaluating Collaborative Environmental Planning Outputs and Outcomes: Restoring and Protecting Habitat and the New York—New Jersey Harbor Estuary Program" *Journal of Planning Education and Research* 27(4): 456-468.
- Mead, M., and P. Byers (1968) *The Small Conference: An Innovation in Communication*. Paris and The Hague: Mouton & Co.
- Patton, M.Q. (2011) *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: The Guilford Press.

Susskind, L.E, B.W. Fuller, M. Ferenz, & D. Fairman (2002) "Multistakeholder Dialogue at the Global Scale," Working Paper, Massachusetts Institute of Technology. September.  
<http://web.mit.edu/publicdisputes/practice/mtstkhld.pdf>

Truex, R., and T. Søreide (2011), "Why Multi-stakeholder Groups Succeed and Fail," *The International Handbook on the Economics of Corruption Volume II*, ed. Rose-Ackerman, S., & T. Søreide. Cheltenham, UK: Edward Elgar Publishing: 478-498.