Hot Issues on the M&E Agenda

Report from an Expert Seminar with Dr. Michael Quinn Patton
March 23, 2012

Irene Guijt
Cecile Kusters
Hotze Lont
Irene Visser

Wageningen UR *Centre for Development Innovation* (CDI) works on processes of innovation and change in the areas of secure and healthy food, adaptive agriculture, sustainable markets and ecosystem governance. It is an interdisciplinary and internationally focused unit of Wageningen University & Research centre within the Social Sciences Group.

Through facilitating innovation, brokering knowledge and supporting capacity development, our group of 60 staff help to link Wageningen UR's expertise to the global challenges of sustainable and equitable development. CDI works to inspire new forms of learning and collaboration between citizens, governments, businesses, NGOs and the scientific community.

More information: www.cdi.wur.nl

Innovation & Change

Ecosystem Governance

Adaptive Agriculture

Sustainable Markets

Secure & Healthy Food

Context, international cooperation (Utrecht, the Netherlands) is a development organisation understanding itself as a social enterprise. We are innovators who use business principles that aim to achieve fundamental social change and generate revenue. Currently Context has 14 staff members, in addition to a number of external Programme Associates. We position ourselves between the development and academic communities, and try to facilitate dialogue between policymakers, academics and development practitioners. Context staff members and Associates have a sound academic background as well as working experience in development practice. For more information: www.developmenttraining.org.

Learning by Design is the company name under which Dr. Irene Guijt operates. She works independently on assessing results in social change, such as through the BetterEvaluation initiative that is being beta-tested and the Big Push Forward. She advises and researches on, and facilitates learning processes and systems in rural development and natural resource management, particularly where this involves collective action. Work over the past twenty years with multilateral organisations and many development NGOs has focused largely on strengthening critical reflective thinking to strengthen pro-poor development.

Vis-à-vis advies is the company name owned by Irene Visser. She works as a consultant in the management of natural resources, climate change, landscape architecture and Monitoring and Evaluation.

Hot Issues on the M&E Agenda
Report from an Expert Seminar with Dr. Michael Quinn Patton

Irene Guijt (organiser, editor)
Cecile Kusters (organiser)
Hotze Lont (organiser)
Irene Visser (documentation)

Hot Issues on the M&E Agenda
Report from an Expert Seminar with Dr. Michael Quinn Patton

Guijt, I.
Kusters, C.S.L.
Lont, H.
Visser, I.

March 2012
Centre for Development Innovation, Wageningen University & Research centre

This report summarises the discussions and presentations of the Expert Seminar 'Hot Issues on the M&E Agenda', which took place in Wageningen on March 23, 2012. The Expert Seminar was organised by the Wageningen UR Centre for Development Innovation in collaboration with Learning by Design and Context, international cooperation.

Preface

Dr. Patton is one of the founding fathers of the evaluation profession in the 1970s and to which he has dedicated his life. He is former president of the American Evaluation Association and author of *Utilization-Focused Evaluation*, now in its fourth edition, plus one book on essentials of Utilization Focused-Evaluation, and *Qualitative Research and Evaluative Methods*, now in its third edition. His 2006 book *Getting To Maybe*, with Frances Westley and Brenda Zimmerman, introduces Developmental Evaluation. Michael Quinn Patton*, Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*, Guilford Press, 2011, is his latest work.

On March 23rd 2012 over 50 people in international development working on evaluation debated the hot issues in monitoring and evaluation with Dr. M.Q. Patton. Before the meeting, participants had been asked to identify their critical M&E issues. These issues were clustered around three broad issues: the politics of evaluation, rigour in evaluative practice, and the future of evaluative practice.

Dr. Patton started off by highlighting his 'personal' top 10 of hot M&E issues, grounded in his vast experience as evaluator in the US and abroad, and his close engagement with the American Evaluation Association.

Chatham house rules applied to the meeting, so direct quotes have been made anonymous.

For the organisers and participants the connection between Patton's experience and our processes were very enriching.

Dr. A.J. Woodhill
Director Wageningen UR Centre for Development Innovation

Acknowledgements

Table of contents

Executive summary

This report summarises the discussions and presentations of the Expert Seminar 'Hot issues on the M&E agenda', with Dr. Michael Quinn Patton, which took place in Wageningen on March 23, 2012. The Expert Seminar was organised by the Wageningen UR Centre for Development Innovation in collaboration with Learning by Design and Context, international cooperation, GIZ, Hivos, Cordaid and ICCO.

The report describes the hot issues on the M&E agenda, divided into a global top 10 of hot issues, the politics of evaluation, the battle field of rigour, and the future of evaluative practice.

List of abbreviations and acronyms

AEA   American Evaluation Association
AfrEA   African Evaluation Association
CDI   Wageningen UR Centre for Development Innovation
EES   European Evaluation Society
JCSEE   Joint Committee Standards on Educational Evaluation
M&E   Monitoring & Evaluation
NGO- IDEAS   NGO Impact on Development Empowerment and Action
UF-E   Utilization Focused-Evaluation
Wageningen UR   Wageningen University & Research centre

1   Patton's top ten of global M&E issues

This section contains a summary of Dr. Patton's lecture on the top ten M&E issues. These are not direct quotes as such. Dr. Patton identifies these as the topics around which the M&E debate and innovation are intense.

10. Professional identity & status

Dr. Michael Scriven, a globally renowned evaluation specialist, contributed many important ideas to the evaluation profession. He was also at the forefront of pushing evaluation as trans discipline, arguing that there is a small number of trans disciplines: philosophy, statistics, ethics and evaluation. Every other discipline depends on those disciplines.

The development of the status of evaluation as both a profession and a trans discipline evolved in parallel with the evaluation community itself. In the 1970s the Australasian, Canadian and American Evaluation Associations were founded, in the 80s the European, AfrEA in 1999, and so forth. There is also NGO-IDEAS, an umbrella organisation of the national and regional evaluation associations. Despite some regional and national differences, these associations are critical for professional development.

Patton described a course he once gave on evaluation consulting. One of the exercises for participants was to develop an elevator pitch about why they were good evaluation consultants. No one mentioned in their pitches that they are or feel part of a profession. Patton's own elevator pitch starts: "I am part of an international evaluation community, with standards for evaluation and we deal with hot topics...." and continues" I am part of a profession that has standards, journals, professional development, associations, communities of practice to nurture new evaluators as professionals."

Part of this professionalisation is certification and licensure (processes that are strong in Canada, New Zealand, and Japan, for example) based on an 'essential skills' framework. Ethical sensitivity to deal with cross-cultural issues is also part of the professional identity.

Patton stressed the importance of evaluation as a profession and using this collective identity in national and regional professional bodies for legitimization of issues, capacity development and networking.

9. Utilization-focused evaluation (UF-E)

When you focus on evaluation use, you need to keep its misuse in mind. Stemming from work in utilization-focused evaluation (UF-E), others such as Malcolm Gray and Bradley Cousins, Patton developed a matrix on use and misuse, hypothesising that if use went up, misuse also went up (see Figure 1 below).

What do we do about the misuse of evaluation, which accompanies the need to ensure more useful evaluation practice? We see instances of substantial misrepresentation, cherry picking among findings and portraying of imbalanced views. Some kind of public response to use and misuse is needed.

AEA (the American Evaluation Association) took a strong stand on two major issues on public policy: the misuse of mandates for randomized controlled trials and high stakes testing. These are very controversial positions in AEA.

Figure 1	The higher the use, the greater chance on misuse

Source: Patton
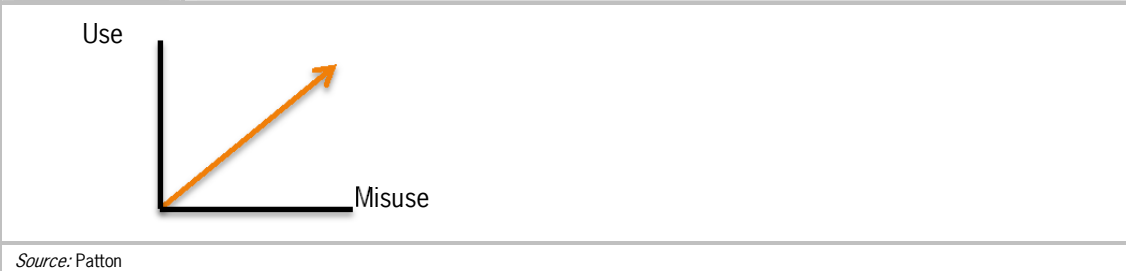
8. Politicization of accountability

If evaluation language becomes part of the political rhetoric (with terms such as "fair and balanced reporting"), then it loses its meaning. The rhetoric around the Vietnam War in the 1960s and 1970s was not a debate about indicators, but about accountability. In contrast, rhetoric around Iraq and Afghanistan, was infused with evaluation language, "what are the indicators", "what are benchmarks that indicate it is good to leave", and so forth.

Once language of evaluation has become part of the political rhetoric, it opens up space for hypocrisy, emptiness, and hollowness of concepts. Politicians are now running platforms on motherhood and accountability as twin goods, for everyone except themselves.

How do we ensure that evaluation remains meaningful? Substance, of knowledge about boundaries, what is possible and not possible, and what we have learned about how accountability gets corrupted.

The notion of 'evidence-based practice' has political cachet, hence its popularity. People use it as in 'I call it evidence based, then I favour a certain political stance'. The same happens with the language of the private sector and business that is increasingly visible in evaluation or in policies. For example, the business sector's version of return on investment has emerged as 'Social Return on Investment'.

7. Relationship between M&E

This hot issue is about the subparts, different kinds and different types. Clients should be clear about what kind of evaluation they need, evaluators should be clear about the approaches they can offer. The relation ship between kinds of evaluation is epitomized by the relationship between M&E.

Surfacing the relationship between 'M' and 'E' is important. One of the remarks on the seminar on Developmental Evaluation on March 22, was that it seems is that Developmental Evaluation seems to be about much good M(onitoring), with occasional E(valuation). Under the umbrella of the relationship between M&E, the larger issue is within E, among the different kinds of evaluation functions (e.g. like in the distinction formative, summative, developmental evaluation). Also, relationships between monitoring, evaluation, auditing, inspecting and management are sometimes difficult to understand. In the latest edition of his Utilization Focused Evaluation book, Patton distinguished 150 approaches of evaluation. Also Scriven (1991) as keeper of the evaluation thesaurus is good at giving names to evaluation. This issue (#7) goes back to the issue (#10) around the professionalization of evaluation, M&E as a profession.

6. Capacity building

Capacity building of evaluators and organisations in evaluation is critically important. But it does not always need to be in courses labelled 'evaluation'. For example, Patton gave a course on evaluation, but under the label of 'leadership', in which the focus was capacity building of individuals and organisations, to incorporate evaluative thinking in their practice.

Much work needs to be done to build in capacity, which is not easy or straightforward. This shows in the winning text of an AEA conference T-shirt slogan contest: "Evaluation is an unnatural act". Capacity building is about the whole notion of what the field of evaluation is, what it means to engage in evaluative thinking, the relation between M&E, the capacity of people in developing countries, of organisation all of this is a huge hot issue.

5. Attention to process-use

Evaluation is not only useful via its findings but via its process. The way in which evaluation process has influence on the people engaged in the process, its findings, and use is the fifth hot topic for M&E. The third edition of the Utilization-Focused Evaluation book discusses this in detail.

People are often focused on evaluation outcomes or findings. But it is often the way in which the evaluation *process* unfolds that has an influence on people's understandings as well as on the findings. An example is the choice of whether or not to involve people in deciding what indicators they consider to be important. Before any data are collected, discussions are about what matters and what the main activities are ('what gets measured gets done'). And thus, this becomes a focus exercise, an example of process use.

This hot topic relates to capacity building, in the sense that as we build capacity, to engage in M&E more generally, we are paying attention to M&E as a process and the utility of that process. How M&E is framed and how analysis is undertaken has an impact on overall utility.

More attention is needed to link effective management and effective M&E, for example, work with philanthropic leaders and their strategic thinking that is fed by evaluation practice.

Cultural differences and power play a role in making the process useful. Sully Gariba, ex-president of AfrEA, refers to 'Africans ways of knowing' that includes a way of asking/listening based on culture (Gariba, 2012). Within hot topic #10, cross-cultural aspects are part of professional identity. The AEA has a statement on cross-cultural competences. (AEA, 2011)

4. 'From studies to streams'

Nicoletta Stame and Ray Rist (2011) published 'From studies to streams: Managing evaluative systems (comparative policy evaluation)'. They note the trend away from producing lengthy reports towards on going data collection, real time data, harvesting data from many different official and unofficial sources, tracking and coding, etc. The information age has substantially changed how we deal with information and learn: "we can't read long reports anymore but just bullet points. This leads also to ever decreasing attention spans. There is neurological evidence for the incapacity to focus, not being able to read a new report.

This trend relates to the increasing trend on visualisation (see hot issue # 2 below), as we search, as a profession, for new ways of sharing data in order to have influence.

3. Rigour

The debate that was quite vivid in the 1990s around qualitative versus quantitative methods has been resolved by agreeing on the use of the term 'mixed methods'. The debate then was very harsh, with the nastiness of the debate around qualitative methods being professionally embarrassing, not a matter of personal integrity and intelligence.

Now the rigour issue has reignited but it focuses on 'attribution' versus 'contribution'. An even harsher debate is felt around the question or assumption on how knowledge is best generated. The RCT, the 'golden standard', is presented by some as the real set, the other one of 'everything else' is supposedly not the good standard. See also Section 3 of this report that focuses on rigour.

2. Reporting

Not only the speed or rhythm (hot issue # 4 above) but also the nature of reporting has changed. There is increased attention to visualization, mapping, video, photos, streaming, poems, (documentary) dramas (engaging people), etc. This year, the American Evaluation Association initiated a TIG (Topical Interest Group) on data visualization. TIGs emerge when they are hot topics in the evaluation profession.

Examples of this include a wide range of ways for more artistic and evocative data reporting, such as conveying findings as plays or documentary dramas, written as poems. Jennifer Green (president of AEA in 2011) had the theme of 'Values' for that conference and included several more artistic examples in the special presidential stream presentations. This hot topic involves engaging stakeholders in diverse ways to report research differently, such as the Vagina Monologues did.

1. Systems and complexity

The evaluation field sees a movement away from evaluating projects and programs, with a trend focusing on new units of analysis: evaluation of strategy and evaluation of systems. These different units of analysis are both a response to, and driving, changing systems. Within AEA, a TIG on 'systems and evaluation' was formed four years ago and is very prominent at the AEA conferences.

Evaluation practice focusing on indicators do not help, particularly when they become high stakes indicators, they can become corrupted. Data have to be talked about and interpreted through dialogue.

Examples of recent work on this shift include an article on Evaluating Strategy in a recent volume of New Directions in evaluation (edited by Patricia Patrizi and Patton) with 'strategy' as the unit of analysis. This work was a response to CEOs of philanthropies wanting evaluations not of specific grants or grant clusters but of their strategies, to help them articulate and approve them. The evaluators were methodologically at a loss. This type of focus requires systems and complexity thinking.

2   The politics of evaluation

2.1   Introduction

Dr. Patton stated that the evaluation profession was slow to see the political nature of evaluation. Early textbooks were worded in terms of the 'observed value free views'. Not until the late 80s was evaluation perceived as a political activity. Patton, in 1988 the president of American Evaluation Association (AEA), (also a US presidential election year), chose as the AEA theme for that year 'the politics of evaluation'. This focus included an essay contest with the theme: "Given that evaluation is political, what aspects of or under what conditions is evaluation not political?"

The winning essay was quite short. 'When no one cares about the program, when no one knows about the program, when no money or power taken is at stake (and no one is sexually active)..., then evaluation is not political" (Patton, 2010).

2.2   Discussions on the politics of evaluation

Irene Guijt introduced the topic to initiate the group discussions.

Knowing what works for whom in which contexts is crucial for all working in international aid. Hard evidence, rigorous data, conclusive proof, value for money, and evidence-based decision making suggests clarity. This promise of clarity encompasses once and for all about what works and what should be funded in international development. But definition tussles, values, priorities and worldviews are in the way. While social interventions are difficult to measure, certain ways of knowing change are considered more legitimate than others for those who hold the purse string.

Therefore the context is that evaluation is political, with issues of power with high stakes. Arenas of tensions include the Dutch 2nd Co-financing Policy Framework (MFS II), the British Value for Money policies, the imposed impact evaluation approaches that make learning/participation harder, the RCT bonanza, etc. How do we deal with this?

Participants discussed current evaluation dilemma/concern around two questions:
1. What makes evaluation political?
2. What have you done to create political space you feel is needed for good quality work that meets real needs?

Participants discussed in eight groups on, for example, the rise of the notion of 'Value for Money' and the need to build people's ability to resist political pressure when asked to do the impossible (e.g. 2 weeks, 1 person to assess impact on poverty in three countries). Another group approached the question differently by saying 'let's acknowledge that we are all engaged in politics. So how can evaluation help us do politics?' Attractively packaging many types of results - including failure reports is one strategy.

Comments that stood out from the group work included:
a. The issue of funding for evaluation underlies political issues.
b. Incentives to maintain the status quo. The 'aid' system wants to maintain itself and keep existing. So what interest do large development organisations have to know and tell the truth about results and changes achieved?
c. Money being allocated for certain types of evaluation, namely (quasi) experimental os crowding out other types of evaluation, although this varies between development organisations.

d. 'Value for money' is being used by decision makers who attempt to use evaluation results to justify decision making resource allocation and expenditure,

e. Co-optation of terminology about rigour (note that the 2010 Utrecht conference " Evaluation revisited originally was originally called 'Reclaiming rigour' but this was deemed too 'political' by the organisers)

f. Two realities are emerging: one responds to political pressures to satisfy funding, the other to find out what is happening. Sometimes, we try to satisfy both sides, by commissioning two studies or combining two types of methodologies.

2.3 Patton's take on the politics of evaluation

2.3.1 Professional evaluation standards

Patton noted the importance of referring to professional evaluation standards when dealing with the politics of evaluation. The evaluation profession has adopted principles and standards to guide professional practice.

For example, the AEA has developed a core set of evaluation principles. The order of these principles does not imply priority among them; priority will vary by situation and evaluator role.

a. **Systematic Inquiry:** Evaluators conduct systematic, data-based inquiries about whatever is being evaluated.

b. **Competence: Evaluators** provide competent performance to stakeholders.

c. **Integrity/Honesty: Evaluators** ensure the honesty and integrity of the entire evaluation process.

d. **Respect for People:** Evaluators respect the security, dignity and self-worth of the respondents, program participants, clients, and other stakeholders with whom they interact.

e. **Responsibilities for General and Public Welfare:** Evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare.

Another important point of reference is the set of program evaluation standards (3rd Edition) developed by the Joint Committee of Standards for Educational Evaluation, of which the AEA is member (JCSEE, 2011 a and b) (see Appendix 2).

a. **Utility standards (U1-U8):** Intended to increase the extent to which program stakeholders find evaluation processes and products valuable in meeting their needs.

b. **Feasibility standards (F1-F4):** Intended to increase evaluation effectiveness and efficiency.

c. **Propriety Standards (P1-P7):** Support what is proper, fair, legal, right and just in evaluations.

d. **Accuracy Standards (A1-A8):** Intended to increase the dependability and truthfulness of evaluation representations, propositions, and findings, especially those that support interpretations and judgments about quality.

e. **Evaluation Accountability Standards (E1-E3)** (meta-evaluation; new): Encourage adequate documentation of evaluations and a meta-evaluative perspective focused on improvement and accountability for evaluation processes and products.

Note that for this set of standards, the sequence is important as it reflects relative priority (see Section 4 below on rigour). Politics is mainly embedded in the 'feasibility standards'.

Another point that Patton stressed was the need to move away from generalisations, generic statements about the politics of M&E as politics is different for different kinds of evaluation (internal, external, formative, summative, etc.)

2.3.2 Include political dimensions in template for evaluation reports

Patton was impressed by the presence at the seminar of some evaluation commissioners and professionals setting guidelines for evaluations. Their presence opens the possibility of including in the template for evaluation reporting, a section on the political context/dimension of the evaluation.

A good example is the evaluation of the impacts of the implementation of the Paris Declaration on Aid Effectiveness. The outputs of this important international joint evaluation include a synthesis report, 21 country evaluations, seven donor studies and several thematic reviews. The evaluation was managed by an independent reference group (OECD, 2011). Important in the document was the explicit inclusion of political contexts in which the Paris Declaration was implemented.

2.3.3 Every evaluation is political and should be held accountable

Evaluators need to transparently acknowledge that every evaluation is political. Taking the notion of 'value for money' and applying this to evaluation means that evaluation should render sufficient value to justify the money spend on evaluation. A compliance exercise is not only a tragedy but also unethical. A euro spent on M&E is a euro not spent on feeding the poor or dealing with Aids. Evaluations should demonstrate utility and contribute to informed decision-making. The idea of 'meta-evaluation' is important for the profession.

To ensure a legitimate set of standards, the JCSEE revises the standard at least every 10 years. In the latest recently completed revision of the standard, a section on accountability and meta-evaluation was included (see Appendix 2)

Meta evaluations are a way to recognise the political pressures under which evaluators operate (see for example Patton and Gornick, 2011). This (meta evaluation) process ensures that we, as evaluators, have somebody looking over our shoulders, rendering credibility to our independence and speaking truth. In this quality check, lies the future of the profession. We cannot continue to assume that evaluation is inherently good, if we cannot show value for money.

2.3.4 Role of professional associations: speak truth to power professionally and jointly

As discussed above, one of the roles for professional associations and evaluation societies is to develop evaluation quality standards based on values. As a professional, we are not alone, and can take a stand by drawing on those standards. Instead of individual battle fighting, we can assert: " I am part of a profession that has standards, this is what the profession stands for, I am proud to be a part of that profession that has values, with decisions taken on positions. The profession has said this is what high quality is." Patton urges those involved in evaluation to recognise that evaluation is political, be sophisticated about this, in part by applying existing standards. We are quite weak as individuals, but have strength as a community. Speak truth to power.

Figure 2	Speaking truth to power

*Source:* Pattons' powerpoint

2.3.5   Propriety…

Politics also occur within an evaluation team. Take, for example, standard A-6 about sound designs and analyses: "evaluations should employ technically adequate designs and analyses that are appropriate for the evaluation purposes". 'Adequate' is subjective, so in the team one might expect a battle, on what is adequate, technically appropriate.

3 Rigour in evaluative practice

3.1 Beyond method as the focus of evaluation quality and rigour

Figure 3	Questions for group discussions

Source: organisers
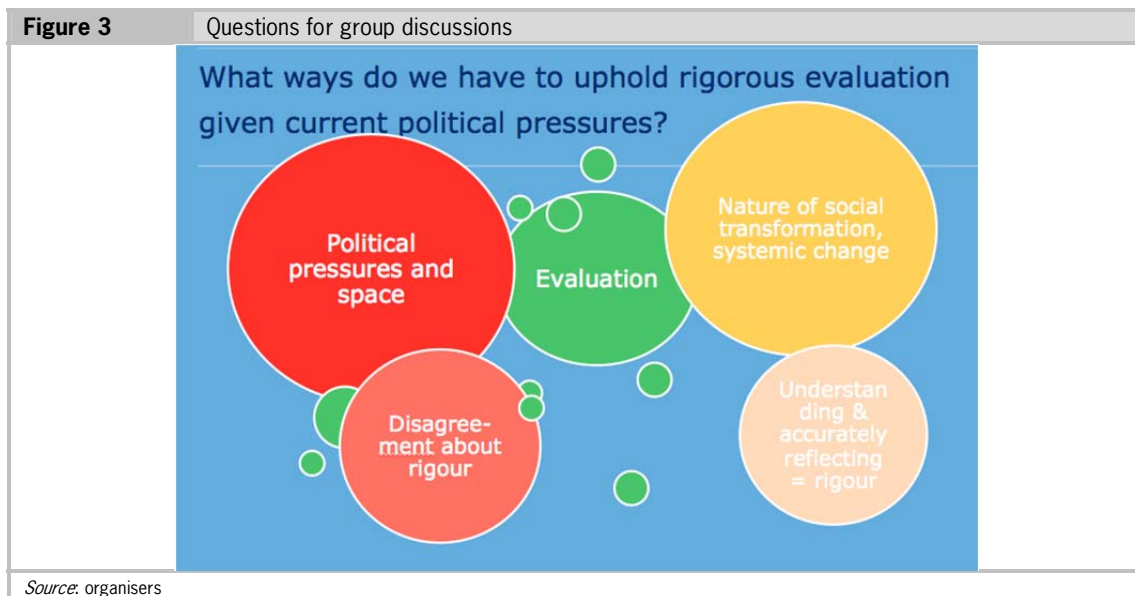
Cecile Kusters introduced the second 'hot topic' – rigour and the core question 'How do we uphold rigorous evaluation given current political pressures'. Figure 3 shows the different dimensions of this topic. First is that evaluation needs to contribute to understanding social change processes, often systemic changes and not help ascertain the numbers of tangible items built, disseminated, used, etc. What rigour is needed and possible to help us understand and accurately reflect social change? Second, we operate in a context with many political pressures – decreasing funding, suspicion of 'the development project', shifting economic power and more. How can we deal with evaluating social change processes within our particular context, influenced by these diverse political pressures? How evaluation is carried out is also influenced by the scientific or evaluation background from which we come. Also, some processes that are valued in international development, such as 'capacity development' and 'civil society strengthening' are very difficult to measure. How do we deal with the rigour around such topics?

Participants' discussions ranged widely in relation to this topic. One group discussion focused on the ubiquitous 'RCT' – the elephant in the room with claims by some in international development that they are the highest standard of methodological rigour. We need to see rigour more broadly, with other methods also being able to lay claim to this term. Maybe the "gold standard" problem is not that big as its claim is not universal. It claims applicability to large n cases. It does not claim rigour/applicability to small-n cases, like lobby and advocacy. Part of the problem may reside in the fat that RCT proponents are not aware of the fact that the development/social change field includes many small N cases. Court case/criminal investigation (n=1), single unit of analysis, can be rigorous investigation process.

Another group questions to what extent 'rigour' is connected to truth. It is clearly connected to 'replication' about which there is often an incorrect assumption about generalizable conclusions. Yet another group discussed the aspect of rigour not being about 'methods' as such but about evaluation standards.

3.2   Patton's take on 'Rigour': rigorous thought and analysis

Patton started his reflections by bringing us back to the Latin origin of rigour. Rigour and rigidity are related with 'rigor mortis' = stiff (dead) body. Part of the issue is about not getting rigid in our efforts to be rigorous within evaluation. Rigour is about rigorous thinking, not about rigorous methods.

The quality of evaluation is improved when the thinking is critical and rigorous, rather than driven by mechanistic and methods driven rigour. The quality of thought can be seen in the ability of dealing with different perspectives and making evaluation transparent.

3.2.1   There is more to evaluation quality than 'rigour'

One of the issues concerns the relative value of rigour. There is more to evaluation quality than 'rigour'. Rigour is not a synonym for quality. Rigour is only one dimension of quality. Other dimensions are: robustness, ownership, ethics, usefulness, transparency - all of which are important.

The aforementioned five agreed evaluation standards are: utility, feasibility, propriety, accuracy and evaluation accountability. The sequence is not accidental. In the recent review process by the JCSEE, Patton said that the most intense discussions were between those keen to elevate 'accuracy' to the first standard and those waving the flag to maintain 'utility'. Utility won out in that space of the professional evaluator. Yet, in other spaces the 'accuracy' standard wins. Maybe there are ways there to expand the discussion on rigour to one on 'quality' – one that encompasses the other standards as well.

The European Evaluation Society has taken position not allowing impact evaluation to be defined by a particular method (European Evaluation Society Statement on methodological diversity, 2007).

Important to note also is that there is a difference between research and evaluation: researchers are only measured against accuracy standards whilst for evaluation the five evaluation standards are important.

Patton referred to a screening process when one is asked to do an evaluation. The first question is: Can we do anything useful here? What is feasible? Can we do it ethically? Can we be transparent? Only then is it useful to focus on the issues and design the studies and methods.

Patton made another point on the importance of 'valuing'. In the recent volume of New Directions for Evaluation 'Promoting Valuation in the Public Interest: Informing Policies for Judging Value in Evaluation' by George Julnes', one of the papers comments on another evaluation weakness, namely that of insufficient rigour about what we mean by 'valuing'.

3.2.2   Analytical rigour in intelligence: Rigour Attribute Model[1]

Patton shared the example from the intelligence community in the USA in relation to the non-existent weapons of mass destruction for which there was supposedly so much evidence. The intelligence community was seriously discredited. In response, a community of practice emerged to redefine 'rigour'. This resulted in the Rigour Attribute Model by Zelik, Patterson and Woods (2010). This model challenges the focus on prescribed standards and tight adherence to this, linking this practice to a failure of 'intelligence'. The authors look at different sources of risk around evidence, and offer eight attributes of analytical rigour.

---

[1] This section is based on two blogs (Guijt 2012; Davies 2012) that were inspired by this section of Patton's talk.

These attributes of analytical rigour, along with guidance on recognising the extent to which each criterion has been met, are summarised in Table 1 below (Zelik et al 2010).

Table 1	Rigour attribute model

1. ***Information Search*** relates to the depth and breadth of the search process used in collecting data (triangulation)
2. ***Hypothesis Exploration*** – extent to which multiple hypotheses are considered in explaining data. In a low-rigor process there is minimal weighing of alternatives (looking at alternative explanations)
3. ***Information Validation*** details the level at which information sources are corroborated and cross- validated (triangulation).
4. ***Stance Analysis*** is the evaluation of data with the goal of identifying the stance or perspective of the source and placing it into a broader context of understanding.
5. ***Sensitivity Analysis*** considers the extent to which the analyst considers and understands the assumptions and limitations of their analysis.
6. ***Specialist Collaboration*** describes the degree to which an analyst incorporates the perspectives of domain experts into their assessments.
7. ***Information Synthesis*** refers to how far beyond simply collecting and listing data analysts went in their process.
8. ***Explanation Critique*** is a different form of collaboration that captures how many different perspectives were incorporated in examining the primary hypotheses.

Source: Zelik, Patterson & Woods, 2007, 2010

Take attribute of 'information synthesis' – a task that challenges the most experienced of evaluators at times. Zelik et al define this as: "*the extent to which an analyst goes beyond simply collecting and listing data in "putting things together" into a cohesive assessment*". Moving from low rigour to high rigour for this attribute leads them to question the lack of insight that accompanies much data (low rigour). High rigour occurs when individuals are involved "*who are "reflexive" in that they are attentive to the ways in which their cognitive processes may have hindered effective synthesis*" (ibid).

This means that it is about transparency about the rigour of thinking rather than rigorous methods. Rigour is not understood by assessing deviations of process, but rather, **by assessing the contextual sufficiency of many different aspects of the analytic process.** "*Across information analysis domains, it is often difficult to recognize when analysis is inadequate for a given context. A better understanding of rigor is an analytic broadening check to be leveraged against this uncertainty. The purpose of this research is to refine the understanding of rigor, exploring the concept within the domain of intelligence analysis. Nine professional intelligence analysts participated in a study of how analytic rigor is judged. The results suggest a revised definition of rigor, reframing it as an emergent multi-attribute* **measure of sufficiency rather than as a measure of process deviation**. *Based on this insight, a model for assessing rigor was developed, identifying eight attributes of rigorous analysis. Finally, an alternative model of briefing interactions is proposed that integrates this framing of rigor into an applied context. This research, although specific in focus to intel analysis, shows the potential to generalize across forms of information analysis.*" (Zelik et al, quoted by Davies).

Cognitive Systems Approach: While professional intelligence analysts can make perceptive assessments about quality of an analysis based on product quality, these perceptions are apt to change with insight into the analytic process.

"*The proliferation of data accessibility has exacerbated the risk of shallowness in information analysis, making it increasingly difficult to tell when analysis is sufficient for making decisions or changing plans, even as it becomes increasingly easy to find seemingly relevant data. In addressing the risk of shallow analysis, the assessment of rigor emerges as an approach for coping with this fundamental uncertainty, motivating the need to better define the concept of analytical rigor.*" (Zelik et al website: quoted by Davies)

Figure 4	What is evidence?

Source: cartoon 1: internet, cartoon 2 Mail today 2008
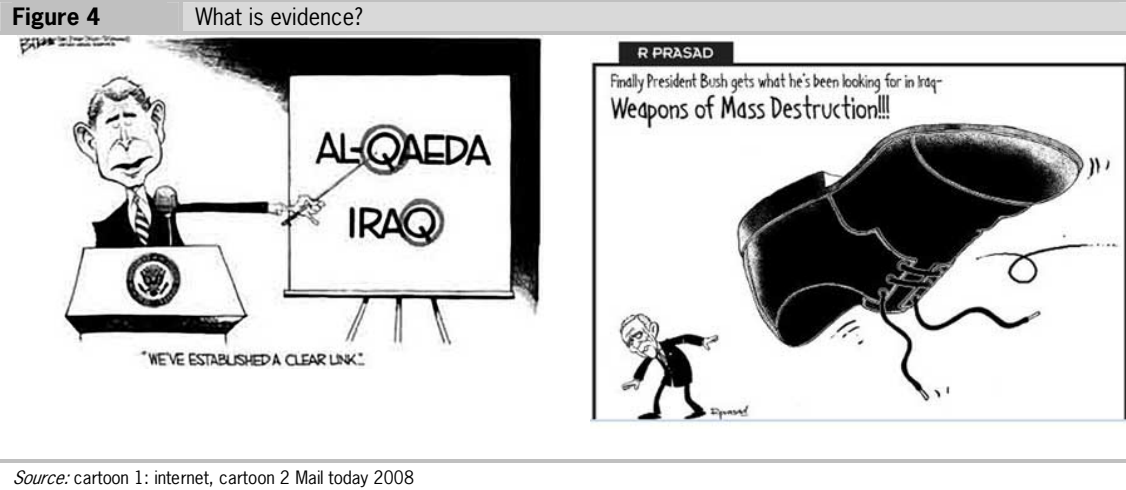
4   The future of evaluation

4.1   Changes in the world that affect future of evaluation

Participants were asked to identify their personal most important change in the world. Then the asserted changes were shared and clustered (see Table 2). Discussion followed around the question why these changes are important for the future of evaluative practice.

Table 2	Clusters of changes as perceived by participants		
– Competition for resources – Less economic growth inthe 'West' – Financial crisis – Neo liberalism – Fragile states – Climate change – Changing aid sector new global framework (post MDGs; China, India, Brazil, philanthropy), shifting power relations, changing values, corporate sector / business – Demise-of-Northern-development agencies as bureaucracy – Equity in development	– Interconnectedness – Social media, electronic media, ICT in data collection, social connectedness – Transparency – Increasing speed of change – Apprehension about the extinction of quantum physics, kindness and complexity	– Antonym of rigour – More people doing evaluation – More people doing evaluation poorly – Growing body of knowledge and expertise on evaluation and the power they can bring (e.g. evaluation associations)	– Developmental evaluation – More awareness on the political dimensions of evaluation – Demand for more evaluations – Design-evaluation connection – Polarisation: accountability *versus* learning or accountability *and* learning?
Source: Individual and group work			

Development aid as sector is going to disappear, for example because civic support for development is diminishing (but still present as a sense of solidarity) due to the financial crisis. The nature of aid flows is going to change, for example from aid to business development. The private sector has a role and can bring in cost efficiency and new ways of monitoring and tracking. An example is the Global Reporting Initiative.

The new social media and the use of ICT in data collection make information more available for more people, also at a rapid rate in developing countries. The increased transparency does not necessarily bring about good things.

The fast growing BRIC economies (Brazil, India, China) cause a shift in power relations, in which money is sourced more locally. There is also a bigger role for private small initiatives as well as finance stemming from big philanthropic organisations.

There is competition for resources, that might mean that there is more money for evaluations available since politicians will need evidence on how the money will be spent. Also the more profound and accelerating changes might mean more work for evaluators, with different levels of interaction between stakeholders.

4.2   Evaluators as futurists

Patton said that in a few years time we will know whether these changes identified by the participants were on the right track or not. He mentioned that there is evaluation literature about futuring with fascinating reviews of different disciplines. Futurists are always almost wrong. Part of the trick is to make lots of predictions so that at least some 1-2 are correct.

Evaluators are inevitably called upon to be 'futurists'. The moment we make recommendations, we become futurists. See Figure 5 for a schematic representation of futurist thinking. Futurists have learned not to make linear predictions but think in terms of the likelihood of scenarios (see for example work of the World Future Society). Futurist thinking as a skill set involves scenario constructing; not getting locked in a new 3-year plan based on mono-dimensional recommendations. Most evaluations are based on the straight line (scenario 2 in the figure), which is the projection of the past into the future. Whatever the baseline is we made history… Straight line is projections from the past- and most evaluations are based on that!

Figure 5	Futurist thinking
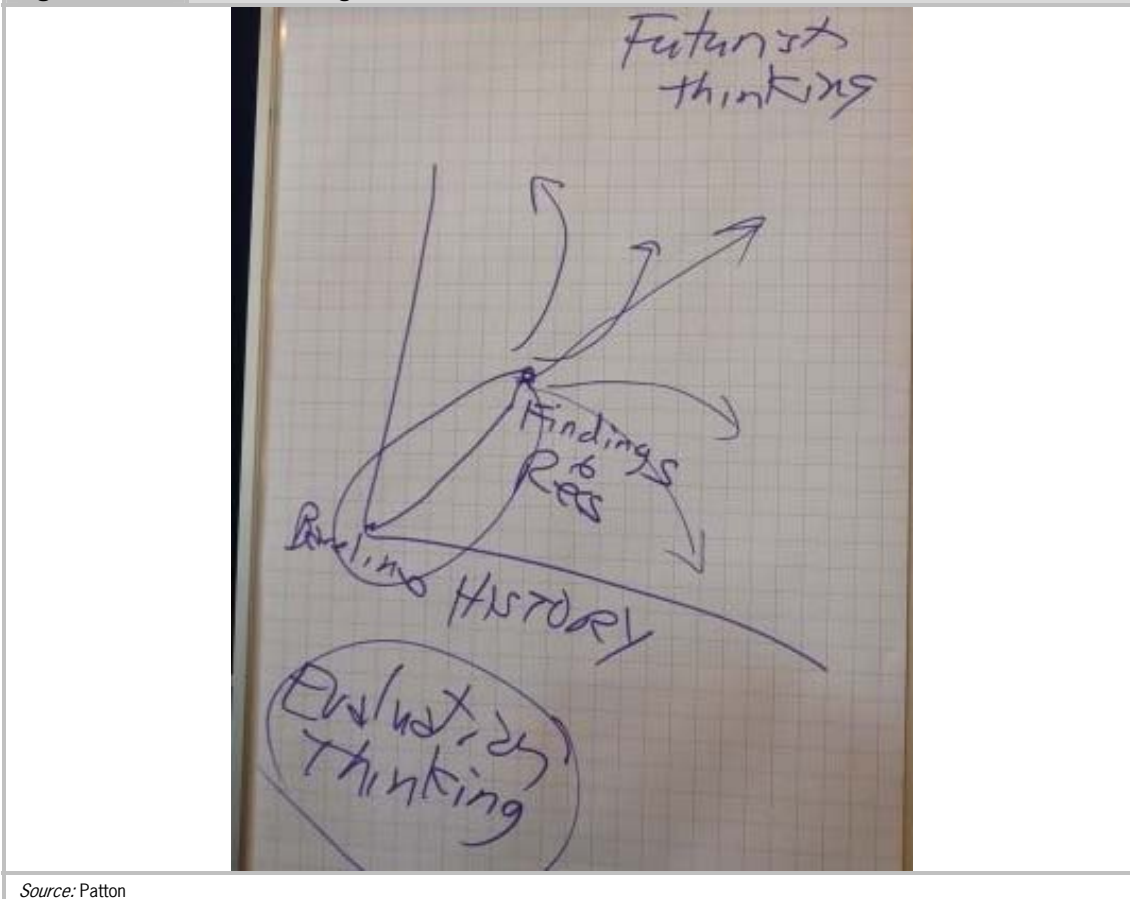
Source: Patton

Evaluators would do well to pay attention to conditions and under what conditions different scenarios could unfold. This is similar to the way of working in developmental evaluation (see seminar report for March 22, 2012). Evaluators might probe in the following manner: "This set of factors appears to have gotten us here, as you move forwards, this are the factors to monitor. So if this set of conditions prevail, these are the consequences of this." The most valuable contribution of this way of thinking is a better understanding

that led to the results that we see. Those need to be monitored closely, every 3-6 months[2] to see how events are unfolding. Then we observe how the conclusions and recommendations are unfolding.

Michael Scriven (who contributed to the evaluation profession with the formative-summative evaluation distinction) also stated that: "evaluators should not make recommendations". The skill set to value whether something did or did not work is different from recommending what should be done as next steps. It is better to present scenarios, not recommendations. The other advantage is less push back and resistance when you confront decision makers with scenarios rather recommendations. It is like saying: "Here are the things to consider, here are some implications, take into account the rest of your world that I do not know about. "

4.3 The future of evaluation

Patton's impression is that in large international aid agencies, most evaluations are undertaken without a single evaluator on the team. In World Bank evaluations, economists are often included and assumed to be 'the evaluation specialist'. Patton states it is crucial to include evaluation specialists in evaluations, along with (but not only) subject matter specialists. Without evaluation specialists, evaluations would be unethical and irresponsible.

The future of evaluation is bright and dynamic. You may get beaten down in one political context but being nurtured and valued in another. All these hard times are part of building the future of our profession.

---

[2] Even in three months things might change dramatically due to changes of government, the budget for development cooperation might be cut, which has implications for evaluation budgets.

References and resources

References

American Evaluation Association, *Guiding principles for evaluators,* 2004
   http://www.eval.org/Publications/GuidingPrinciples.asp

American Evaluation Association, *American evaluation association statement on cultural competence in evaluation,* 2011 http://www.eval.org/ccstatement.asp

European Evaluation Society Statement on methodological diversity, 2009
   http://www.europeanevaluation.org/images/file/Member_section/EES%20Statement%20on%20Metho
   dological%20Diversity/EES%20Statement.pdf

Davies, R. Blog, *Analytic Rigour in Information Analysis – Lessons from the intelligence community?* 2012
   http://mande.co.uk/2012/uncategorized/analytic-rigour-in-information-analysis-lessons-from-the-
   intelligence-community/

Gariba, S and Hoop, T. de *African theories of change: lost in translation?* 2012
   http://www.3ieimpact.org/en/blog/2012/02/06/african-theories-change-lost-translation/

Gamble, J.A.A., A Developmental Evaluation Primer, The J.W. McConnel Family Foundation, Canada, 2008
   http://tamarackcommunity.ca/downloads/vc/Developmental_Evaluation_Primer.pdf

Guijt, I. Blog *The Battlefields of Rigour,* 2012 http://bigpushforward.net/archives/1432

Joint commission on Standards for Educational Evaluation, *Program evaluation standards, synthesis*
   http://www.eval.org/EvaluationDocuments/progeval.html

Joint Commission on Standards for Educational Evaluation, *The Program Evaluation Standards, A Guide for Evaluators and Evaluation Users* Third Edition, Sage, 2011

Mintzberg, H. *Tracking Strategies: Toward a General Theory*, Oxford University Press, 2007.

Miranda, R*. Eva, the Evaluator*, Learningham Press, 2009

OECD, *Evaluation of the Implementation of the Paris Declaration*, 2011
   http://www.oecd.org/document/60/0,3746,en_21571361_34047972_38242748_1_1_1_1,00.html

Patrizi, P and M.Q. Patton *Strategy as the focus for evaluation.* In P. A. Patrizi & M. Q. Patton (Eds.),
   Evaluating strategy. New Directions for Evaluation, 128, 5–28, 2010.

Patton, M. Q., "*Developmental Evaluation*." P. 116 in S. Mathison (ed.) Encyclopaedia of Evaluation.
   Thousand Oaks, CA: Sage, 2005.

Patton, M.Q. "*Evaluation for the Way We Work*", Nonprofit Quarterly, Spring, 2006
   http://store.nonprofitquarterly.org/13evforwaywe.html

Patton, M.Q. *Utilization-Focused Evaluation*, 4th ed., Sage, 2008.

Patton, M.Q and J. Gornick, *Evaluation of the phase 2 evaluation of the Paris declaration. An Independent review of Strengths, Weaknesses, and Lessons,* 2011
   http://www.oecd.org/dataoecd/37/3/48620425.pdf

Patton, M.Q. *Developmental Evaluation: Applying complexity concepts to enhance innovation and use*, the
   Guildford Press, 2011, chapter 1 of the book http://www.guidlford.com/excerpts/patton.pdf

Patton, M.Q. *Webinar on Developmental Evaluation*, UNICEF, 2012
   http://mymande.org/index.php?q=content/developmental-evaluation

Patton, M.Q., *Essentials of Utilization-Focused Evaluation*, Sage, 2012
http://www.sagepub.com/books/Book233973

Scriven, M. various writings on types of evaluations http://michaelscriven.info/

Scriven, M. *Evaluation thesaurus*, 4th edition Sage, 1991

Snowdon, D. J. and M.E. Boone, *A Leader's Framework for Decision Making*, Harvard Business Review, November 2007

Stame N. and Rist, R. *From Studies to Streams: Managing Evaluative Systems*, Transaction Publishers, 2011

Vaginamonologues, 1996 http://en.wikipedia.org/wiki/The_Vagina_Monologues

Westley, F., Zimmerman, B. and Patton, M.Q. *Getting to Maybe: How the World Is Changed,* Random House Canada, 2006,

Zelik, D. J., Patterson, E. S., & Woods, D. D. *Measuring attributes of rigor in information analysis*. 2010 In E. S. Patterson & J. E. Miller (Eds.), Macrocognition metrics and scenarios: Design and evaluation for real-world teams. Aldershot, UK: Ashgate.

Zelik, D., Patterson, E. S., & Woods, D. D. *Understanding rigor in information analysis*. 2007

Modeling Rigor in Information Analysis: A Metric for Rigor Poster (PDF) This poster provides an overview of the rigor model, identifying the aspects of the attributes that contribute to low, moderate, and high rigor analysis processes. It also overviews the rigor metric as applied to the LNG Scenario study.

For further references see http://csel.eng.ohio-state.edu/zelik/research/Rigor.html

Appendix 1
Biography

Michael Quinn Patton

Dr. M. Q. Patton is an independent organizational development and evaluation consultant. He is former **President of the American Evaluation Association**. He is the only recipient of both the *Alva and Gunnar Myrdal Award* from the Evaluation Research Society for "outstanding contributions to evaluation use and practice" and the *Paul F. Lazarsfeld Award* for lifetime contributions to evaluation theory from the American Evaluation Association. The Society for Applied Sociology honoured him with the 2001 *Lester F. Ward Award* for Outstanding Contributions to Applied Sociology. He was the Gwen Iding Brogden Distinguished Lecturer at the 2008 National Conference on Systems of Care Research for Children's Mental Health.

He is the author of five evaluation books including a 4th edition of *Utilization-Focused Evaluation* (2008) and 3rd edition of *Qualitative Research and Evaluation Methods* (2002). These books have been used in over 500 universities worldwide. He is also author of *Creative Evaluation* (1987); *Practical Evaluation* (1982); and *Culture and Evaluation* (1985). He has co-authored a book on the dynamics of social innovation with two Canadians drawing on complexity theory and systems thinking: *Getting to Maybe: How the World is Changed* (Random House, 2006). His latest book is *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use* (Guilford Press, 2011).

After receiving his doctorate in Organizational Sociology from the University of Wisconsin, he spent 18 years on the faculty of the University of Minnesota (1973-1991), including five years as Director of the Minnesota Center for Social Research and ten years with the Minnesota Extension Service. He received the University's Morse-Amoco Award for outstanding teaching. He also served as a faculty member with the Union Institute & University for 16 years.

He was the keynote presenter for the launching of the Latin American Network in Peru in 2004, the African Evaluation Society in Nairobi, Kenya in 1999 and at the European Evaluation Society in Switzerland in 2000. He has twice keynoted the American, Canadian, and Australasian Evaluation Society conferences, as well as national evaluation conferences for the United Kingdom, New Zealand, Italy, Denmark, Japan, and Brazil. He is a regular trainer for the World Bank in Ottawa, the American Evaluation Association's professional development courses, and The Evaluators' Institute.

He has worked with organizations and programs at the international, national, state, and local levels, and with philanthropic, not-for-profit, private sector, and government programs. He has worked with peoples from many different cultures and perspectives. He is a generalist working across the full range of efforts at improving human effectiveness and results, including programs in leadership development, education, human services, the environment, public health, employment, agricultural extension, arts, criminal justice, poverty programs, transportation, diversity, managing for results, performance indicators, effective governance, and futuring. He uses a variety of evaluation and research methods, with special focus on mixed methods designs and analysis.

Appendix 2

Program Evaluation Standards Joint Committee for Standards on Educational Education

---

Utility Standards
The utility standards are intended to increase the extent to which program stakeholders find evaluation processes and products valuable in meeting their needs.
U1 Evaluator Credibility Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context.
U2 Attention to Stakeholders Evaluations should devote attention to the full range of individuals and groups invested in the program and affected by its evaluation.
U3 Negotiated Purposes Evaluation purposes should be identified and continually negotiated based on the needs of stakeholders.
U4 Explicit Values Evaluations should clarify and specify the individual and cultural values underpinning purposes, processes, and judgments.
U5 Relevant Information Evaluation information should serve the identified and emergent needs of stakeholders.
U6 Meaningful Processes and Products Evaluations should construct activities, descriptions, and judgments in ways that encourage participants to rediscover, reinterpret, or revise their understandings and behaviors.
U7 Timely and Appropriate Communicating and Reporting Evaluations should attend to the continuing information needs of their multiple audiences.
U8 Concern for Consequences and Influence Evaluations should promote responsible and adaptive use while guarding against unintended negative consequences and misuse.

Feasibility Standards
The feasibility standards are intended to increase evaluation effectiveness and efficiency.
F1 Project Management Evaluations should use effective project management strategies.
F2 Practical Procedures Evaluation procedures should be practical and responsive to the way the program operates.
F3 Contextual Viability Evaluations should recognize, monitor, and balance the cultural and political interests and needs of individuals and groups.
F4 Resource Use Evaluations should use resources effectively and efficiently.

Propriety Standards
The propriety standards support what is proper, fair, legal, right and just in evaluations.
P1 Responsive and Inclusive Orientation Evaluations should be responsive to stakeholders and their communities.
P2 Formal Agreements Evaluation agreements should be negotiated to make obligations explicit and take into account the needs, expectations, and cultural contexts of clients and other stakeholders.
P3 Human Rights and Respect Evaluations should be designed and conducted to protect human and legal rights and maintain the dignity of participants and other stakeholders.
P4 Clarity and Fairness Evaluations should be understandable and fair in addressing stakeholder needs and purposes.
P5 Transparency and Disclosure Evaluations should provide complete descriptions of findings, limitations, and conclusions to all stakeholders, unless doing so would violate legal and propriety obligations.
P6 Conflicts of Interests Evaluations should openly and honestly identify and address real or perceived conflicts of interests that may compromise the evaluation.
P7 Fiscal Responsibility Evaluations should account for all expended resources and comply with sound fiscal procedures and processes.

Accuracy Standards
The accuracy standards are intended to increase the dependability and truthfulness of evaluation representations, propositions, and findings, especially those that support interpretations and judgments about quality.
A1 Justified Conclusions and Decisions Evaluation conclusions and decisions should be explicitly justified in the cultures and contexts where they have consequences.
A2 Valid Information Evaluation information should serve the intended purposes and support valid interpretations.
A3 Reliable Information Evaluation procedures should yield sufficiently dependable and consistent information for the intended uses.
A4 Explicit Program and Context Descriptions Evaluations should document programs and their contexts with appropriate detail and scope for the evaluation purposes.

A5 Information Management Evaluations should employ systematic information collection, review, verification, and storage methods.

A6 Sound Designs and Analyses Evaluations should employ technically adequate designs and analyses that are appropriate for the evaluation purposes.

A7 Explicit Evaluation Reasoning Evaluation reasoning leading from information and analyses to findings, interpretations, conclusions, and judgments should be clearly and completely documented.

A8 Communication and Reporting Evaluation communications should have adequate scope and guard against misconceptions, biases, distortions, and errors.

Evaluation Accountability Standards

The evaluation accountability standards encourage adequate documentation of evaluations and a metaevaluative perspective focused on improvement and accountability for evaluation processes and products.

E1 Evaluation Documentation Evaluations should fully document their negotiated purposes and implemented designs, procedures, data, and outcomes.

E2 Internal Metaevaluation Evaluators should use these and other applicable standards to examine the accountability of the evaluation design, procedures employed, information collected, and outcomes.

E3 External Metaevaluation Program evaluation sponsors, clients, evaluators, and other stakeholders should encourage the conduct of external metaevaluations using these and other applicable standards.

Source: http://www.eval.org/EvaluationDocuments/progeval.html

This report summarises the discussions and presentations of the Expert Seminar 'Hot Issues on the M&E Agenda', which took place in Wageningen on March 23, 2012. The Expert Seminar was organised by the Wageningen UR Centre for Development Innovation in collaboration with Learning by Design and Context, international cooperation.

The report describes the hot issues on the M&E agenda, divided into a global top 10 of hot issues, the politics of evaluation, the battle field of rigour, and the future of evaluative practice.

More information: www.cdi.wur.nl