

EVALUATIONS
CONNECTIONS

CONTENT

A Message from the President	1
Refocusing Evaluation: The Road to Dublin An editorial	2
Evaluation credibility and methodology: What counts as credible evidence and for whom?	3
The cornerstone RCT: norm, mistake, or exemplar?	4
White privilege – the flipside of inequality	6
Gender equality in development evaluation: lessons learned from British and Swedish practice	7
Strengthening Evaluation for Development: Do we need some 'Grand Challenges'?	9
Reframing Evaluation Capacity Development in Africa	11
Evolution of UNDP's approach to National Evaluation Capacity Development	13
Albert Hirschman as evaluation thinker	14
The authors	15

A Message from the President

Dear members and colleagues,

This is the last issue of our Newsletter before the next EES biennial conference in Dublin (October 1–3, 2014). The variety of articles presented in this issue offers only a small glimpse of the rich diversity of topics that will be evoked by paper presentations, panels and posters in Dublin. Yet they all relate to the overall theme of the conference, i.e. *Evaluation for an equitable society: independence, partnership and participation*. For they focus on strengthening evaluation and fostering participative evaluation; underline the imperatives of independence and credible evidence; interrogate political considerations in the use of randomized controlled trials, promote methods to integrate gender in mainstream evaluation; and add building blocs to evaluation capacity development through the incorporation of contextual political economy analyses.

The peer review of proposals for paper presentations, panels and posters for the Dublin conference has now been finalised. We received 552 submissions representing an increase of 24% as compared to the 2012 Helsinki conference. From the abstracts, it appears that a myriad of case studies, research results, experiments and developing theories will be presented and debated, covering a wide range of subjects from evaluation methods and effectiveness to governance and ethics in evaluation. I have no doubt that Conference delegates will address the intractable dilemmas and urgent concerns currently faced

by evaluation commissioners, evaluators and evaluation users in our troubled world and that they will trigger exchange of experiences and the building of relationships among our members.

Finally I am proud to inform you that the EES Conference will again provide a platform for a meeting of the Network of Evaluation Societies in Europe and that we will host the launch of the Faster Forward Fund (FFF) sponsored by Michael Scriven (one of our keynote speakers). FFF is designed to promote innovative evaluation in the public interest. We will also wish to share with you the results of EES workshops that were supported by EVALPARTNERS among which one that recommended a pilot initiative focused on voluntary evaluator peer review. Your insights on this proposal and your active involvement in the deliberations of current and future Thematic Working Groups are very much expected.

We have so much to discuss in the conference that three days will hardly suffice! The grounds are set for lively and buoyant exchanges of experience and knowledge within our ever-increasing community, and I am very much looking forward to meeting you in Dublin. Your fulsome engagement is needed so that your Society can respond to your expectations and so that we can work together to meet the challenges of evaluation in Europe and beyond.

Claudine Voyadzis, EES President

REFOCUSING EVALUATION: THE ROAD TO DUBLIN

An editorial

Robert Picciotto

This issue of Connections displays diverse ideas and trends within our far flung evaluation community. Once again it demonstrates a discipline in full bloom. It presages the next EES Biennial Conference to be held in Dublin (October 1–3, 2014). The overarching Conference theme (*Evaluation for an Equitable Society*) has struck a chord: it has attracted a record number of submissions for papers, panels and posters.

The Conference will address a fundamental issue: given that most people and institutions are driven to protect their own interests how should evaluators contribute to the design and implementation of policies and programmes that serve the public good? To be sure the basic evaluation mandate has not changed: evaluators are still tasked to dissect facts to reach conclusions about merit, worth and significance. But public perceptions about society have evolved given massive and growing inequalities.

By generating and disseminating knowledge, informing public debate and more than ever by amplifying the voice of the most disadvantaged in society evaluators help to shape the incentives framework within which policy decisions are made. However evaluation exerts influence only if the evidence it produces is credible. Hence the question treated by Maria Bustelo's leading article lies at the core of evaluation quality: what counts as credible evidence and for whom?

In mobilizing evidence for improved public policy there is no gold standard and no

silver bullet despite the hullabaloo associated with randomized controlled trials. This is the theme of William Faulkner's article which deconstructs what may well be the most celebrated evaluation ever carried out – the IFPRI Progresa evaluation. Despite the numerous defects of this evaluation (inadequate sampling, contamination and attrition) decision makers shrewdly exploited the mystique associated with experimental methods to build political support and mobilize financial resources for a major federal program in Mexico.

No surprise here: evaluation has always been about politics. Inevitably the ideological lenses through which different stakeholders interpret evidence shape their judgments about the policies and programmes being evaluated. Diverse ways of seeing generate different premises about what constitutes merit, worth and significance. Hence independent evaluators keen to make a difference take explicit account of stakeholders' passions and interests.

From this perspective, Kate McKegg calls on evaluators to be "responsive in practice and in form to the aspirations, perspectives and views of those for whom inequality is a way of life" in a spirited article about "white privilege". Similarly, in an article informed by hard won evaluation experience Julia Espinosa recognizes that the mainstreaming of gender equality, an international commitment, "tends to evaporate" in the face of cultural resistance and institutional rigidity. As for Zenda Ofir she calls on creative minds in the south and

the east to achieve breakthrough advances by rising to grand evaluation challenges.

Underlying the cultural concerns and vested interests that impede policy reform are structural impediments built into the fabric of country authorizing environments. Applying a political economy lens to the evaluation experience of five African countries, Osvaldo Feinstein and Stephen Porter reframe the evaluation capacity development enterprise as a tool for helping countries "sail against the wind towards democracy".

Thus improving the enabling environment for evaluation emerges as an imperative. This is the purpose of south-south cooperation promoted by the United Nations Development Programme through regular International Conferences on National Evaluation Capacities. They are aptly described by Roberto La Rovere and Ana Rosa Soares. In this context, social research and evaluation should work hand in hand. They should not treat each other as rivals to be denigrated or colonized.

Instead social scientists and evaluators should operate as partners in a shared enterprise. When done well evaluation and social research converge. Illustrating this tenet the last article in this issue shows that Albert O. Hirschman, arguably the most influential and original economist of his generation, was also an evaluation pioneer. All of these ideas and many more will be explored at the evaluation event of the year – the 11th EES Biennial Conference.



Convention Centre Dublin, Ireland, 1–3 October 2014
Pre-conference workshops 29–30 September 2014

11th EES Biennial Conference
**Evaluation for an
Equitable
Society**

www.ees2014.eu

Independence, Partnership, Participation



EVALUATION CREDIBILITY AND METHODOLOGY: WHAT COUNTS AS CREDIBLE EVIDENCE AND FOR WHOM?¹

María Bustelo

Many issues affect the credibility of an evaluation – the expertise and independence of the evaluators, the degree of transparency in the evaluation process and the quality of outputs, to name just a few. The cultural context is also important – the values on which an evaluation rests, the way that evidence is deemed credible, the institutions that support evaluation systems and structures, the people that contribute to an evaluation, and how the evaluation is shared, communicated and reported. I developed these issues at the 2013 NEC Conference in Sao Paulo. Here I present the issue of credibility and methodology and the way in which the evaluation community conceives credible evidence.

Evaluation credibility is related to the data that should be gathered empirically to answer evaluation questions. This has a first level, which is related to the scope of information to be gathered and the reliability of the information acquired by an evaluation. The quantity and reliability of information, along with lack of bias, is what we aim to get through third persons or secondary sources. Where and from whom did we get information? Did we gather it from the whole spectrum of stakeholders? Were we honest and impartial in the selection and interpretation of the evidence? Did we report successes as well as failures?

Credibility is inevitably related to the methodological perspective as well as to how empirical evidence is gathered and interpreted. It is related to methods (for example, questionnaires and interviews) and methodology (for example, case studies, surveys and experiments). But it is also related to a more philosophical question of social inquiry, about the nature of reality, about what constitutes knowledge and how it is created, that is, about epistemology and ontology. More-

over, what is considered credible evidence is clearly mediated by key philosophy of science notions, such as the concept of paradigm. The debate on what it is considered credible evidence comes from an old and recurring discussion on how best to study social phenomena: the quantitative-qualitative debate.

In the evaluation field, the question about what constitutes credible evidence used to support claims relating to the impact of a practice, programme or policy, fiercely reappeared when some international organizations, networks and federal departments in the USA identified the randomized controlled trial (RCT) as the “gold standard” design for generating ‘scientific’ evidence of evaluated programmes or policies. This created much discomfort in the evaluation community, and generated responses from several evaluation societies, including from American Evaluation Association (AEA) in 2003 and the European Evaluation Society in 2007.² In 2006, Claremont University organized a symposium on “What Counts as Credible Evidence in Applied Research and Evaluation Practice?” in which eminent evaluation academics, both in the experimental-quantitative and the non-positivistic-constructivist and qualitative approaches participated. It led to the production of an edited volume (Donaldson, Christie & Mark, 2009), organized around social inquiry paradigms that helped to frame the debate regarding credible evidence.

Although one could argue that the very differentiation between experimental and non-experimental approaches is somehow normative and tends towards the condition defined–defining ‘the other’ by the absence of that condition – this was the first time that the credibility and the concepts of evidence and impact were debated openly and in depth from a methodologically plural perspective

not exclusively related to a single epistemological stance, such as experimentalism.

From my perspective, there was no other way, because the discipline of evaluation has evolved from single narratives to joint perspectives and multiple methods and approaches that accommodate very different evaluation models and traditions. Due to its practical and applied nature, and the need for credible evidence from different perspectives to answer varied questions in diverse contexts, evaluation has been one of the first fields in which “quantitative” and “qualitative” researches and evaluators have exchanged views, reached out across networks and engaged with one another. This is reflected in the widespread embrace of the mixed methods approach. In our “era of paradigm pluralism” (Greene, 2013: 111) and tolerance of different social inquiry perspectives, this approach is now broadly accepted in the evaluation community.

The mixed methods approach applies not only at the methods level, but also at methodology and epistemology levels. As Donna Mertens and Sharlene Hesse-Biber state, “it is important to understand that mixed methods is not just about (mixing and combining) methods. The use of any given method or set of methods in an evaluation is also tightly linked to specific epistemologies, methodologies (theoretical perspectives), and axiological assumptions, as well as being connected to particular stakeholder perspectives” (Mertens & Hesse-Biber, 2013: 5–6).

This claim emphasizes the need to be conscious of and explicit about theoretical perspectives and assumptions. Thus Jennifer Greene argues that mixed methods evaluators should be “explicit about the paradigmatic assumptions that frame and guide

¹ This is a summarized extract of the Guiding Thematic Paper on Credibility I wrote and presented at the 3rd International Conference on National Evaluation Capacities: Innovative Solutions to Challenges Linked to Independence, Credibility and Use of Evaluations, Sao Paulo (Brazil), 29th Sept–2nd Oct. 2013.

² See EES statement on methodological diversity at www.europeanevaluation.org/library.htm.

their work”, and that “careful explication of just what is being mixed in a mixed methods study contributes to the subsequent warrant for and thus credibility of results”. For her, “it is a critical responsibility of the *inquirer*” to make explicit assumptions, such as the “nature of the social world, what counts as warranted knowledge, defensible methodology, and the role of social inquiry in society”. They should also “justify the values they invoke – values of distance, engagement, inclusion, objectivity, generalizability, contextuality, social action”, and so forth. This is particularly important in evaluation contexts, because they are saturated with value” (Greene, 2013: 111–112).

In sum, credibility, evidence and impact are not concepts exclusively valid for positivist stances, so they should be explored and defined by other paradigmatic perspectives. Positivism has been the dominant paradigm

for many years, but this is not necessarily the case anymore – as the methodological and paradigmatic pluralism in the evaluation community has demonstrated. Mixed Methods evaluators propose advancing the debate of credible evidence by making explicit values as well as ontological, epistemological and methodological choices.

Therefore, paradigmatic and methodological transparency is needed for credibility. We should insist that this transparency be exercised by all evaluators. For the sake of credibility, ‘classical’ and dominant understandings, such as experimentalism, should not be taken for granted, and the paradigmatic and methodological choices that drive particular perspectives should be explained and justified. This would acknowledge that there are other legitimate modes of inquiry. In this way, methodological pluralism would ensure that a single perspective would cease to be

perceived as the ‘norm’, while other ‘alternative’ choices have to be justified.

References

Donaldson, S., Christie, C. & Mark, M. 2009. *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* Thousand Oaks, CA: Sage Publications.

Greene, J. C. (2013). Reflections and ruminations. In D. M. Mertens & S. Hesse-Biber (Eds.), *Mixed methods and credibility of evidence in evaluation. New Directions for Evaluation*, 138, 109–119.

Mertens, D. M., & Hesse-Biber, S. (2013). Mixed methods and credibility of evidence in evaluation. In D. M. Mertens & S. Hesse-Biber (Eds.), *Mixed methods and credibility of evidence in evaluation. New Directions for Evaluation*, 138, 5–13.

THE CORNERSTONE RCT: NORM, MISTAKE, OR EXEMPLAR?

William N. Faulkner

The mainstream narrative of perhaps the most famous randomized-controlled trial (RCT) in history ignores gaping omissions, moral quandaries and sociopolitical currents which riddled its undertaking and dissemination. This short article summarizes a study (Faulkner, 2012) which addressed three aspects of the IFPRI’s¹ evaluation of PROGRESA², (1997–2000) then and still Mexico’s largest social program: sampling, attrition and contamination. The first-order goal of the study was to help fill some significant lacunas by profiling them against (a) the intellectual context of RCTs and (b) the political realities of social policy development in Mexico during the late 1990s. Its main message is that what superficially appeared as an airtight laboratory-style trial turned out to have had significant flaws. Yet given the sociopolitical and intellectual context of the evaluation, the smooth, glossy narrative

is the one which has survived. It still echoes throughout the evaluation world today as a paragon of RCT’s powers as extolled by its proponents and held aloft before its critics.

Sampling: Sifting through the vast body of articles discussing the IFPRI-Progressa evaluation reveals that one of its most important and celebrated features (methodologically speaking) remains shrouded in mystery. Specifically, IFPRI-Progressa’s sampling methodology appears to be recorded only in personal communications and fading memories. Scattered sources, mostly in the grey literature, offer sparse and contradictory references in appendices and footnotes. They provide tantalizing hints of its true complexity (Faulkner, 2012: 88–89). The earliest of these sources seems to indicate that either a clustered semi-experimental design, or a matched-pairs method was implemented, but none

provides enough detail to fully understand the process. The authors who tested the statistical equality of the treatment and control groups at baseline found a mix of evidence as to whether or not the sample was randomly distributed.

Attrition: Selective sample attrition was strongly present: “close to 35% (45%) of households (individuals)... in one or more waves,” “differ[ing] significantly between treatment and control groups.” (Rubalcava and Teruel, 2003: 7). As opposed to selection bias, which features prominently in the IFPRI-Progressa documentation, only two (of 17) final reports mention attrition bias, one only as a hypothetical. A handful of later pieces analyze attrition in the sample, but the evaluation reports which dramatically influenced Mexican social policy and bestowed the project with celebrity status

¹ International Food Policy Research Institute.

² *Programa de Educación, Salud, y Alimentación* (Education Health, and Nutrition Program). In 2001, the program was renamed under its current moniker, *Oportunidades* (Opportunities).

strategically omitted direct mention of this critical weakness.

Contamination: The term ‘contamination’ refers to several different concepts. In this article, I define the term as the possibility that “individuals from control localities or other localities [could have] immigrate[d] to treatment group localities in order to receive program services” (Behrman and Todd, 1999: 3). Unlike attrition, not even the presence of contamination is verifiable since the evaluation did not follow up with out-migrants or interview in-migrants. This is worrisome, given that “control communities often were literally surrounded by [treatment] communities,” (Fiszbein and Schady, 2009: 311) and that the experiment had to be ended sooner than planned due to pressure from local authorities in control communities (Parker and Teruel, 2005: 211)

So why have multitudes of influential economists, donors, and policymakers eulogized IFPRI-Progresa over the past fourteen years? Why would so many trained eyes risk throwing in the towel with such a technically dubious study? Textual breadcrumbs that lead back to answers to these problematic issues are not too difficult to follow.

While several authors have delved back into the IFPRI-Progresa experience, a systematic and thoroughly critical lens was never used. As a result, especially in its early years, an oversimplified narrative served to engender broad political changes within Mexico and abroad. Domestically the IFPRI evaluation helped Progresa gain political hardiness and facilitated access to funding. It also helped turn evaluation into a legally codified piece of Mexican federal social programming. Given the myopia that allowed the caveats of IFPRI-Progresa’s methods to go unacknowledged, the evaluation paved the way for a technocratic coup d’état that jolted Mexican social policy out of a longstanding vicious cycle of political patronage and frequent overturn.

Internationally, the World Bank and the Inter-American Development Bank extensively and successfully used the example of IFPRI-Progresa to propel greater use of Conditional Cash Transfers (CCTs) and to extol the virtues of evidence-based policy (via ‘rigorous’ evaluation of social programs). A single, high profile study provided these institutions with a foundational exemplar that

allowed them to solidify both trends over the following decade.

Complicating the narrative of IFPRI-Progresa’s findings by fully explicating their potential weaknesses would have paralyzed both operators and funders. Any apparent weakness of the representativeness of findings might have been used to ensnare the emergence of the evidence-based policy strategy while still in its infancy. Such nit-picking would have flagged the momentum made towards spreading the CCT tool and tarnished the image of a project which actively and seriously promoted governmental transparency and the de-politicization of social programs.

In 2014, more than a decade later, the political status of IFPRI-Progresa and its worldwide relevance have faded dramatically. We now see that the rise in evidence-based policy triggered an intense, widespread debate over whether RCTs should be the preferred suppliers of the evidence base. While the Progresa evaluation was a striking example of evaluators’ power to facilitate social reform, it also contributed to this unnecessary dispute. Today, filling in the story of IFPRI-Progresa by unveiling its socio-historical complexity provides important lessons for the evaluation community. It should also help to bring RCT proponents and their critics closer together since the case vividly displays the demonstrable flaws and the obvious strengths simultaneously.

The “RCT-as-gold-standard” agenda still commands a significant swath of the networks which commission and undertake social policy evaluations. Yet the sharp edge of the “randomista” movement (as Martin Ravallion dubbed them) is largely composed of academic micro-economists. Most of them, along with the staff in their organizations, have only a vague awareness of evaluation as a discipline in its own right. On the one hand, some academic economists (e.g. Lant Pritchett) have begun to more openly condemn the “randomista” agenda. On the other hand, some prestigious voices advance technical solutions to the ailments of the policy-relevant RCT production machine (e.g. Miguel et al., 2014). Meanwhile, the public relations staffs of the randomistas’ organizations are still permitting the breathless overselling of RCTs to broader audiences (e.g. Benko,

2013). Clearly, there remain opportunities for evaluators to be more active and exigent about holding the peripheral and politically influential edges of the field accountable. If ever such savvy evaluators were searching for an ideal centerpiece with which to begin a grounded, fruitful discussion of the rewards and risks of experimental evaluation they could find it in IFPRI-Progresa.

Full publication: Faulkner, W. N. 2014. “A Critical Analysis of a Randomized Controlled Trial Evaluation in Mexico: Norm, Mistake or Exemplar?” *Evaluation* 20 (2) (April 10): 230–243. doi:10.1177/1356389014528602. <http://evi.sagepub.com/cgi/doi/10.1177/1356389014528602>.

Sources:

Benko, Jessica. 2013. “The Hyper-Efficient, Highly Scientific Scheme to Help the World’s Poor.” *Wired Magazine*. <http://www.wired.com/2013/11/jpal-randomized-trials/>.

Camerer, Miguel, E. C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., Glennerster, R., et al. 2014. “Promoting Transparency in Social Science Research.” *Science* 343 (6166): 30–31. <http://emiguel.econ.berkeley.edu/research/2/>.

Faulkner, William N. 2012. “Negotiated Perspectives: A Complexity Approach to IFPRI-Progresa in Mexico”. Tulane University. http://media.wix.com/ugd/33c277_6d608b30854c413b8edb76a1bbd88667.pdf.

Fiszbein, Ariel, and Norbert Schady. 2009. “Conditional Cash Transfers: Reducing Present and Future Poverty.” *Journal of Banking & Finance*. Vol. 28. Washington, DC. doi: 10.1016 / S0378–4266(03)00124–9. <http://linkinghub.elsevier.com/retrieve/pii/S0378426603001249>.

Parker, Susan W., and Graciela Teruel. 2005. “Randomization and Social Program Evaluation: The Case of Progresa.” *The ANNALS of the American Academy of Political and Social Science* 599 (1) (May 1): 199–219. doi: 10.1177 / 0002716205274515. <http://ann.sagepub.com/cgi/doi/10.1177/0002716205274515>.

Rubalcava, Luis, and Graciela Teruel. 2003. “The Effect of PROGRESA on Demographic Composition”. 285. Economics Working Paper. Mexico City.

WHITE PRIVILEGE – THE FLIPSIDE OF INEQUALITY

Kate McKegg

For those of us who consider evaluation to be a democratizing practice – with the goal of social betterment – it is a critical time. We appear to be losing ground on most democratic principles we hold dear (McKegg, 2013). How can we justify that one-fifth of people alive on the planet today experience growing wealth, longer lifespans, more leisure, greater comfort, whilst the remaining four-fifths are losing on all fronts (O’Riordan, 2012)? Why does inequality keep rising and the wealthiest in society are pulling further away (Weissman, 2014)?

Our evaluative efforts are usually focused on supporting programme and policy improvements targeted towards ameliorating the plight of needy or ‘target populations’. We recognize the centrality of problems associated with poverty and inequality. And yet we are not making much headway. A game changing strategy is needed. More of the same just will not do. What if instead of a focus on inequality we focused on “white privilege”? What might our evaluations look like if we paid attention to white privilege instead of targeted inequalities? As Anne Milne (2009) reminds us when we talk about inequality or diversity “the background color stays white”.

Perhaps, if we acknowledged that privilege exists among the white, majority population of western democracies, a pattern “held in place by internal processes of majoritarian democracy and by an ideological consensus of a benign, and inevitable, colonization” (Hugens, 2011) we would provoke sufficient discomfort to create the conditions for change. For if we are to redesign the system, we have to recognize its “colossal unseen dimensions” (McIntosh, 1988; quoted in Milne, 2009).

White privilege as a concept boils down to identifying benefits and rights that accrue to the dominant white majority (to which I belong without any deliberate effort on my part). Below are a few examples of ways white privilege manifests itself in my own everyday life (McIntosh, 1988):

1. If I should need to move house, I can be pretty sure of renting or purchasing a house in an area which I can afford and in which I would want to live
2. I have no difficulty finding neighborhoods where people approve of me or my household
3. I feel welcomed and ‘normal’ in the usual walks of public, institutional and social life
4. Whether I use checks, credit cards or cash, I can count on my skin color not to work against the appearance of my financial reliability
5. I can arrange to protect my children most of the time from people who might not like them
6. I do not have to educate my children to be aware of systemic racism for their own daily physical protection
7. I can be pretty sure that my children’s teachers and employers will tolerate them if they fit school and workplace norms; my chief worries about them do not concern others’ attitudes toward their colour
8. I am never asked to speak for all the people of my racial or ethnic group
9. I can be pretty sure that if I ask to talk to the ‘person in charge’, I will be facing a person of my colour
10. I can be late to a meeting without having the lateness reflect on my colour or race.

A focus on such white privilege is about switching the gaze and acknowledging that it exists. It is the flipside of inequality – just as day is to night, as hot is to cold. Opening our eyes to white privilege is a crucial aspect of critical, transformative evaluative thinking and practice. Naming white privilege for what it is, rather than obfuscating the reality under the guise of doing good for those ‘others’ is something, albeit uncomfortable, we must now do. Most of the ‘targeted’ communities I have ever worked with do not need us to determine their needs; or to try and save them. They need those of us with white privilege to understand our privileged position and become supporters in their struggle (Smith, 1998; Hugens, 2011)).

Hugens (2011) has a useful way of framing the ‘work’ that white privileged evaluators need to tackle through their thoughts and actions. She suggests four categories of work:

Ideological

Ideological work involves critically revisiting *and then retelling* the history of our relationship with minority, indigenous or enslaved peoples in order to appreciate that western, colonial ideologies have shaped all our world views, and been self legitimizing for those of us with privilege. Ideological work challenges evaluators to critically question and unpack the assumptions that underlie needs assessments, evaluation design, implementation, outcomes and use.

Cultural

Cultural work involves critiquing those aspects of our identity, culture and tradition that will *not* serve social betterment or reducing inequality and oppression. Hugens (2011) argues that white privilege espouses values of equality, justice and human rights while tolerating indifference to the experience of others. This has serious implications for the roles we play on evaluation projects and teams. Privileging others’ voices, values and traditions is a stance we must take if we are to shift the balance of power.

Emotional

As we bring to the notice of those with privilege our collective complicity and ignorance, there is tough emotional work ahead as we deal with feelings of shock, fear, guilt etc. We should not expect help or assistance from those we have oppressed. In evaluation situations, our job is to challenge others like us about their power over an evaluation process, design, system or practice when it has implications for perpetuating inequalities and oppression.

Constitutional

This work involves the privileged majority being responsive in practice and in form to the aspirations, perspectives and views of those for whom inequality is a way of life.

This may mean moving beyond majoritarian processes towards constitutional changes. In evaluation contexts, our evaluation associations can play an important role in forging new representation and decision making processes that ensure there are many more voices around our board tables.

Although these notions are provocative and uncomfortable, if we are going to make inroads into inequality, we have to develop an evaluation praxis that “transforms ourselves as well as transforming reality” (Hugens, 2011). As evaluators we have to discover and adopt a practice of being white that works to dismantle current relationships and rebuild them in ways that do not oppress others.

References

Freire, P. (1975). ‘Cultural action for freedom, *Harvard Educational Review Monograph*, 1

Hugens, I. (2011). Developing a Decolonisation Practice for Settler Colonisers: A Case Study from Aotearoa New Zealand *Settler & Colonial & Studies* 2,1

McIntosh, P. (1988). *White Privilege and Male Privilege: A Personal Account of Coming To See Correspondences through Work in Women’s Studies* (Working Paper 189): Wellesley College Center for Research on Women, Wellesley MA.

McKegg, K. (2013). Evaluation Responsibility and Leadership in the Face of Failing Democ-

racies *American Journal of Evaluation* 2013 34: 579–582.

Milne, A. (2009). *Colouring in the White Spaces, Cultural Identity and Learning in School*. A research report prepared for the ASB/ APPA Travelling Fellowship Trust following a study tour in 2009.

O’Riordan, T. (2012) On social sustainability in a world of limits facing prolonged austerity. *Sustainability, Science, Practice and Policy*, 12, 1–2

Smith, L. (1998). *Decolonising Methodologies: Research and Indigenous Peoples* Auckland, New Zealand, Zed

GENDER EQUALITY IN DEVELOPMENT EVALUATION: LESSONS LEARNED FROM BRITISH AND SWEDISH PRACTICE

Julia Espinosa

Gender equality has been increasingly recognized as a crosscutting issue in development evaluation since the nineties. Since development policies have different impacts on women and men, the inclusion of a gender perspective in evaluation is now considered essential to achieve results, promote learning, enhance accountability, empower people and accelerate positive social change (Batliwala, 2011, Bustelo, 2011; Espinosa, 2013; Sielbeck-Bowen et al., 2002). However, the incorporation of a gender perspective in development and evaluation activities has not been an easy task. Although gender equality is now an international commitment, concern with gender issues has tended to evaporate in practice, due to problems and

to resistance encountered during the mainstreaming gender process.

This brief article, based on a meta-evaluation of policies, procedures and methodologies implemented by British and Swedish development agencies between 2000 and 2010¹, presents lessons learned on how to include a gender perspective and as a result improve evaluation practice. The integration of a gender perspective implies rethinking all the evaluation steps. It means collecting information on structural and systemic gender inequalities and identifying the measures needed to promote further gender transformations (Ligero et al., 2014; Bustelo, 2011; UNEG, 2011; UN Women, 2011; González

and Murguialday, 2004). It implies redesigning the main phases of the evaluation process – preparation, definition and development, dissemination and utilization – by exploring the systemic and structural nature of gender inequality in order to help design and carry out more transformative policies. In particular the inclusion of a gender perspective into evaluation requires measuring changes in gender relations with a view to foster greater equality between women and men and, thus, to improve policies.

How in practice can evaluation integrate a gender perspective? The lessons drawn from the analysis of the British and Swedish experience in development evaluation differ.

¹ This meta-evaluation examines how gender issues have been included in the following evaluation units and in their gender focused evaluation exercises: the Evaluation Department (EvD) of the Department for International Development (DFID), in the United Kingdom; and the Evaluation Department (UTV) of the Swedish International Development Cooperation Agency (Sida) and the Swedish Agency for Development Evaluation (SADEV), in Sweden.

Some are linked to the institutional context, others have to do with how gender equality results can be assessed and evaluations made more gender-sensitive.

As to the institutional context, evaluation policies and procedures that include gender equality as a central topic are crucial to promote the inclusion of a gender perspective. Specific tools and templates also help ensure gender mainstreaming. The incorporation of a gender perspective into evaluation depends ultimately on the willingness to prioritize gender issues, on the resources – time, funding and staff – available for the task and on the gender expertise and training of the staff involved in the process.

Concerning evaluation practice, the analysis of British and Swedish reports shows that there is no pre-defined model for assessing gender equality. The incorporation of a gender perspective implies paying attention to diverse gender issues, the characteristics of the evaluand and the diverse voices of stakeholders. In addition, it entails using all available experience relevant to gender equality taking into account gender and development theory. This forms the basis for defining evaluation criteria and questions.

All reports show the need for gender indicators in the formulation and monitoring phases in order to assess intervention outcomes. They emphasize the relevance of both process and outcome. Indicators should provide information on inputs, processes and results. Given the complexity of measuring gender change, designing indicators in a participatory manner is recommended. While the evaluation team cannot define all the indicators, it needs to be aware of the changes experienced by both women and men so that the voices of both sexes are heard to a similar degree.

Additionally, these evaluation reports combine quantitative and qualitative techniques

and use participatory tools to measure transformation in gender relations. This methodological approach responds to the complexity of measuring changes in gender relations and the limited existence of data disaggregated by sex. Moreover, it is linked to the wish to promote people's greater involvement in the decision-making process and greater ownership of evaluation results.

Finally, according to the analysis of these two donors, dissemination and use of evaluation results in a gender sensitive way to all stakeholders and their full engagement in the review of policies and their implementation are needed to ensure that evaluation results are actually used in decision-making processes.

In sum, the British and Swedish experience provides useful lessons as to how to include a gender perspective into evaluation and promote a gender-sensitive evaluation culture and practice. It confirms that mainstreaming gender in the policy cycle faces different types of challenges – ranging from institutional resistance to the lack of resources, training and sensitivity. Sustained efforts within public institutions, social organisations and evaluation teams will be required to give gender issues in evaluation the priority they deserve.

References

Espinosa, Julia (2012) 'Gender issues in the evaluation of international aid. The experience of British, Spanish and Swedish development agencies. Paper presented in the 10th European Evaluation Society International Conference, Helsinki.

Espinosa, Julia (2011) *La igualdad de género en la evaluación de la ayuda al desarrollo: Los casos de la cooperación oficial británica, sueca y española*. PhD Research. Madrid. Universidad Complutense de Madrid. <http://eprints.ucm.es/13206/>

Other references

Batliwala, Srilatha (2011) *Strengthening Monitoring and Evaluation for Women's Rights: Thirteen Insights for Women's Organizations*. Toronto. AWID.

Bustelo, Maria (2011) 'Last but not least: Gender-sensitive evaluation as a forgotten piece of the policymaking process'. Unpublished manuscript.

Espinosa, Julia (2013) "Moving towards gender-sensitive evaluation? Practices and challenges in international development evaluation". *Evaluation* 19(2), pp. 171 – 182.

González Gómez, Lara y Murguialday Martínez, Clara (2004) 'Evaluar con enfoque de género'. Cuadernos Bakeaz, n° 66. http://pdf.bakeaz.efaber.net/publication/full_text/128/CB66_maqueta_PDF.pdf

Ligero, Juan Andrés, Espinosa, Julia, Mormeneo, Carmen and Bustelo, María (2014) *Diferentes aproximaciones para hacer una evaluación sensible al género y al enfoque basado en derechos humanos para el desarrollo*. Madrid. Ministerio de Asuntos Exteriores y Cooperación.

Sielbeck-Bowen, Kathryn A., Brisolará, Sharon, Seigart, Denise, Tischler, Camille, & Whitmore, Elizabeth (2002) 'Exploring feminist evaluation: the ground from which we rise'. *New directions for evaluation*, 96, pp. 3–8.

UNEG (2011) *Handbook for Integrating Human Rights and Gender Equality Perspectives in Evaluations in the UN System*. UNEG. <http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/IOS/temp/HRGE%20Handbook.pdf>.

UN Women (2011) *UN Women's Guide to GE / HR Responsive Evaluation*. UN Women. http://unifem.org/evaluation_manual/index.html

STRENGTHENING EVALUATION FOR DEVELOPMENT: DO WE NEED SOME ‘GRAND CHALLENGES’?¹

Zenda Ofir

We live in extraordinary times. The rapid development of new technologies is taking society into uncharted waters; inequalities are impoverishing large swathes of populations in ostensibly prosperous nations; the long-dominant development model of capitalism is being convincingly challenged; and developed countries increasingly face unfamiliar uncertainties. On the other hand we can celebrate that for the first time in history, in countries until recently regarded as severely underdeveloped, hundreds of millions have been lifted out of poverty in record time – and almost exclusively through their own efforts.

Yet evaluation as currently practised and promoted has hardly played a role in these achievements. *We need to step back and seriously consider the implications for the profession.* It is my contention that in preparation for the exciting yet challenging time ahead, development evaluation requires a revitalised, innovated agenda nurtured by more visible, dynamic thought leadership from ‘developing’ countries, with special attention to critical issues at the development-evaluation interface. This will help us to focus on *evaluation for development*, rather than on the evaluation of development.

What is important at the development-evaluation interface? Over the past decade the distinctions between developed and developing countries have become increasingly blurred; some argue that making a distinction between ‘evaluation’ and ‘development evaluation’ is now moot. Yet there are still significant differences. The vulnerabilities of developing countries tend to be magnified: the poor are poorer; the vulnerable more vulnerable; institutions and systems more fragile or unstable; the powerful and powerless more so; contexts less predictable; and

those capacities considered essential to executing conventional development models, lower.

Moreover, ‘development’ is not a successfully executed project or programme that shows impact. As leading economist Ha-Joon Chang observes, a country can be called ‘developed’ only if its high income is based on superior knowledge embodied in technologies and institutions. He contends that *sustained* development requires effective, efficient institutions and productive enterprises supported by the collective accumulation and use of knowledge, and the expansion of those social and technological capabilities that are “both the causes and the consequences of such transformation”.

‘Development’ is the path and progress of a country towards that state. Yet global, regional and even national discourses very often fail to address some of its most important components – and evaluation follows silently. Interventions that focus on individuals and ‘communities’, and on “siloes” projects and programmes may provide some building blocks; some may even catalyse development. But our inability to evaluate effectively at higher levels of aggregation – i.e. at the ‘big picture’ level – exacerbates the ‘micro-macro disconnect’ that haunts development.

We urgently need more evaluation instruments that can help improve development in a powerful way. Engaging effectively with the intersection between development and evaluation is therefore not a trivial issue. It means that we have to understand better what might catalyse, drive and influence development – and evaluate cognisant of these.

At the very least it demands that we be explicit about the assumptions, values and

frameworks that underlie and link development and evaluation, and that we pursue innovation in evaluation with attention to the consequences for development success. For example, evaluation that ignores the role of power negates its crucial role in development efforts. Focusing an evaluation on the interests of individuals at the cost of community harmony reflects an understanding of development where individual interests dominate those of the collective, which might counter a nation’s culture. Rigidly applying ‘logframes’ within a too-short funding cycle ignores the reality of evolving development contexts and slow, initially even negative change trajectories. Using people as objects for extraction and experimentation without giving them voice in evaluation dismisses the need for their voice in their own development. Isolating strands of a development intervention when evaluating impact, ignores the fact that the whole is more than the sum of the parts. Focusing on achieving and then measuring (average) impacts without also focusing vigorously on what can enhance the chance of sustained impact can give an inflated sense of success.

Development specialists, evaluation commissioners and evaluators need to work in a more concerted fashion on evaluation priorities that will support development more effectively. Perhaps we need a set of ‘*Grand Challenges*’² that can focus our attention on resolving, with a sense of urgency, the most intractable challenges plaguing evaluation for development. Identifying them will require careful thought, but may relate to *i.a.* theory and exemplars around some of the following:

Firstly, how can we best promote simulation, experimentation, real-time evaluation, fast adaptation and scaling – including understanding in depth, respecting and systemati-

¹ Based on a forum article by the author in the *American Journal of Evaluation* 34(4), pp 582–586, Dec 2013.

² The Grand Challenges originated in mathematics more than a hundred years ago. The transfer of the concept to Global Health, for example, encourages innovation with the aim of engaging creative minds across scientific disciplines to work on solutions that can lead to breakthrough advances.

cally working with 'context' and the crucial role of 'culture' in our very diverse world?

Secondly, how can we develop a much more sophisticated understanding of how to use and communicate the much-abused concepts of 'systems' and 'complexity'? Working with these concepts is necessary in all evaluation efforts, as Ben Ramalingam eloquently highlighted in his recent book – and even more so for those issues that need urgent resolution: technical challenges such as understanding change trajectories, scaling up, fixing the 'micro-macro disconnect' and predicting and neutralising potentially negative consequences; crucial development foci such as empowerment; institution strengthening and sustaining impacts; and increasingly prominent areas of work such as the use of evidence (research; evaluation), enhancing resilience and meeting global challenges.

Thirdly, how can we best evaluate impact with much more rigorous attention to how to sustain it – where this is desirable? Although the pendulum is now swinging back from the unhelpful enthrallment with RCTs, it still has to do so from 'measuring impact' and 'value for money' towards also enabling smart engagement with *managing*

for sustained development – including more systematically predicting and searching for unintended negative impacts.

Fourthly, how can we ensure the development and use of sets of practical standards for evaluation quality, 'rigour' and 'credible evidence' that transcend too narrowly defined ideas of the 'scientific method' – especially for qualitative and mixed methods designs? And what are the ethical and practical implications for the evaluation profession of ignoring that much of our 'data from the field' are incorrect, and that 'big data' are open to manipulation in a highly networked, competitive world where 'the truth' is largely hidden?

Fifthly, as money is getting scarcer and alternative financing and funding models by *i.a.* the BRICS are gaining momentum in Africa, Latin America and Asia, new financing mechanisms from traditional donor countries using seductive names such as 'impact investing' and 'social impact bonds' may put vulnerable societies at risk. How can we be better equipped to help ourselves and stakeholders plan and assess benefits and risks, and quickly detect negative consequences for development efforts?

Finally, how can we ensure that meta-analysis and synthesis of results and lessons across evaluations are not only done with greater vigour and rigor, but made available and communicated in a manner that can support in theory and practice the many different worldviews of development?

It is time that evaluation thought leadership emerges more visibly from the global South and East. The scope for new non-Western theories and practices in revolutionising development evaluation is not yet clear, but developing countries have rich cultures with knowledge and wisdom spanning thousands of years that have yet to be applied to evaluation. We need specialists from outside the evaluation arena to engage vigorously in helping to resolve our challenges. Champions who have a propensity for conventional *and* new indigenous paradigms, and who can help mobilise intellectual and financial resources across disciplinary, paradigmatic, sector and geographic boundaries should come forward. The global evaluation community *has* to attract a larger number of smart, innovative thought leaders to advance the evaluation profession in the cause of global justice. Perhaps a focus on 'Grand Challenges' can help us to do so.

REFRAMING EVALUATION CAPACITY DEVELOPMENT IN AFRICA¹

Oswaldo Feinstein and Stephen Porter

Most evaluation capacity development (ECD) work focuses exclusively on technical strengthening of evaluative functions or monitoring systems within government. This article argues that ECD priorities should be shaped in relation to the political economy. Country specific governance conditions determine the extent to which evaluation is used in policy processes. Consequently, ECD should be undertaken in a manner that works towards improvement of the prevailing political economy.

The Regional Centre for Learning on Evaluation and Results for Anglophone Africa (CLEAR-AA) reached this overarching conclusion through synthesising findings from case studies in five African countries (Ghana, Ethiopia, Malawi, Rwanda and Zambia) undertaken with the support of the UK Department for International Development (DFID).² Researchers within the countries and from CLEAR-AA carried out the studies with inputs from a reference group.³ The initial conceptual framework and guidance for the study was provided by DFID.⁴

The five case studies mapped the opportunities and challenges for conducting evaluation amongst a variety of role-players. This mapping disclosed that technical capacity to manage and undertake evaluations within the public sector of the case countries is often below par and unlikely to display independence. This is a major constraint on the quality and use of evaluation. The mapping helps explain why evaluations are often commissioned and managed by development partners rather than government. Yet universities, think tanks and civil society actors are often potentially com-

Principals	Government Agents	Evaluation Agents
Executive	Central government	Universities
Legislature	Line ministries	Think Tanks
Civil Society		Evaluation associations and networks
Development Partners		Consultants

Table 1: Principals and Agents.

petitive suppliers of evaluation services. They are well informed about the local situation and better equipped to navigate the political context than foreign experts. Such technically competent and politically savvy evaluation actors offer convenient entry points for ECD.

The five case studies suggest that the application of political economy analysis to ECD should focus on:

- (i) The conditions under which demand is generated for evidence; and
- (ii) The areas in which supply can be strengthened to meet and foster this demand;

Applying a political economy lens to the five case countries two major configurations were identified: (i) *neo-patrimonial*, in which public affairs have been captured by personal or private networks and (ii) *developmental patrimonial*, in which a central ruling elite commands and promotes long-term development. Adding (iii) *liberal democracy* to the analytical framework facilitates the design of tailor made ECD strategies that fit

individual countries' characteristics with a view to improve evaluation governance.

The study was carried out through a combination of desk reviews, including an analysis of existing evaluation/evaluative research products, and direct semi-structured interviews with a selection of informants across critical stakeholder groupings. In total 77 agencies were interviewed. Evaluation capacity was assessed according to a conceptual scheme of 'principals' and 'agents'. Principals influence the demand-side of evaluation. Typically government agents commission evaluations, although they can also supply evaluations. A variety of evaluation agents constitute the supply-side.

The key finding of the research is that the country's political economy drives and conditions policy processes, within which evaluation supply and demand interact. Many demand and supply-side constraints or barriers are technical, yet because the overall policy space is politicised it is essential to undertake reviews of the political economy in order to identify good entry points for

¹ Full versions of the case studies, the final report and the original ToR can be found at: <http://www.clear-aa.co.za/publications/>.

² CLEAR-AA is based at the Graduate School of Public and Development Management at the University of the Witwatersrand, Johannesburg.

³ The team members supporting this study were: Stephen Porter, Salim Latib, Oswaldo Feinstein and Anne McLennan from CLEAR-AA/Wits; and from the countries Oswald Mulenga (Zambia), Getnet Zewdu (Ethiopia), Charles Gasana (Rwanda), Hannock Kumwenda (Malawi), Samuel Adams and Charles Amoatey (Ghana). In addition, a panel of international experts provided comments.

⁴ Specifically David Rider Smith who commissioned this study.

ECD. Table 2 below, presents an abbreviated version of the study findings.

The country studies made clear that a theory of change for evaluation capacity development should emphasize the political economy context. Neopatrimonial and developmental patrimonial political logics offer distinct entry points and suggest distinct approaches for ECD. In *neo-patrimonial* states the informal forces that shape decision-making or implementation processes are opaque and hard to reform through direct intervention. However, the existence of competing interests provides alternative entry points for ECD as well as for the promotion of competing users of evaluative evidence. In *developmental patrimonial* states policy-making is centralised. This imposes strict limits on policy influence. Nor is persuasion targeted at a narrow elite always a promising option.

Given these realities some ECD critics argue that in such situations the enabling environment for evaluation is so adverse that nothing can or should be done. An alternative stance is to acknowledge the risks while considering the potential rewards of a “possibilist”

approach informed by a political economy analysis that identifies technically sound and politically savvy targets for ECD activities.

Other critics have argued that donor driven evaluation does not address issues relevant to the local context and curtails the use of aid evaluation. In developmental patrimonial countries that have strong ownership of policies and strong technocratic agencies, this is a non-issue since governments are the dominant partner and aid evaluation offers opportunities for policy dialogue. However in such cases donors may see fit to encourage involvement of non government evaluation suppliers to enrich the debate. In neopatrimonial states the usefulness of evaluations hinges on the judicious selection of evaluation suppliers and how the subsequent informal policy contestation unfolds. ECD thus conceived can strengthen the civil society and gradually open up a useful role for donors committed to equity through evaluation.

Summing up, the case studies show that in all countries there are opportunities to strengthen technical evaluation supply and to elicit demand for evaluations.

- Through ECD think tanks and universities may enhance their technical capacities to conduct evaluations in a manner that employs their political capital to support use.
- Donors can provide opportunities for learning by doing through support within sector-working groups that are country-led.
- Politically legitimate civil society organisations can promote evaluations
- Sharing evaluation experiences among sub-Saharan African countries can strengthen local and regional evaluation networks, contribute to the development of regional evaluation capacities and foster demand for evaluation by making policy makers aware of the knowledge generated by evaluation and the possibilities of using that knowledge to improve policy making.

Finally, acknowledging the importance of the political economy in ECD can help to reframe technocratic issues as political and contribute to transform evaluation from a bureaucratic/development partner requirement into a tool for helping countries to “sail against the wind” towards democracy. ■

Developmental Patrimonial	Neopatrimonial	Commonalities
<p>1.1 Policy is difficult to influence through evidence unless you can access central policy making structures.</p> <p>1.2 Centralised patronage structures allow strategic resource allocations.</p> <p>1.3 Influence on implementation through evaluation is possible.</p> <p>1.4 Development partners have limited input into policy decisions.</p> <p>1.5 Openness to debate is limited, but there are strong technocratic central ministries to oversee implementation. Strong central ministries offer a focal point for ECD.</p>	<p>1.6 Actual policy change is difficult to achieve, but interest groups can use evidence to their advantage to access resources.</p> <p>1.7 There are weaker central ministries and technocratic controls, but there is open contestation between interests. Champions need to be identified at both central and in line ministries (e.g. health).</p>	<p>1.8 Loyalty is aligned to elite interests rather than performance (although the two can overlap).</p> <p>1.9 Entry points for evaluation in civil society can be identified amongst older civil society actors that have developed their political legitimacy across different actors over time.</p> <p>1.10 Lack of embedded networks of evaluators that link to policy processes.</p> <p>1.11 Some quality university expertise in economics, health and agriculture.</p>

Table 2: ECD through a Political Economy Lens.

EVOLUTION OF UNDP'S APPROACH TO NATIONAL EVALUATION CAPACITY DEVELOPMENT

Roberto La Rovere and Ana Rosa Soares

Evaluating the performance of public policy is fundamental to good governance, government accountability and development effectiveness. Accordingly the Evaluation Policy of the United Nations Development Programme (UNDP) emphasizes the priority of national evaluation capacity development. In this context, UNDP programme units promote and coordinate South-South and trilateral cooperation in support of capacity-building for evaluation at country level by strengthening communities of practice in evaluation and maintaining regional rosters of evaluation experts and institutes.

Since 2009, International Conferences on National Evaluation Capacities (NEC) have been privileged instruments through which the Independent Evaluation Office (IEO) of UNDP has contributed to improved enabling environments for national evaluation capacity development. The latest such event (www.nec2013.org/) was held in 2013 in Brazil (papers and proceedings, also of the 2009 and 2011 NEC conferences, can be found at: <http://web.undp.org/evaluation/nec/proceedings.shtml>). It sought "solutions to challenges related to independence, credibility and use of evaluation". The Conference brought together representatives of national institutions responsible for commissioning, conducting and using evaluations of public policies, projects and programmes, as well as civil society representatives, eminent evaluation practitioners and academics Voluntary Organizations of Professional Evaluation (VOPEs) leaders and senior UN/development agencies officials.

Building upon prior national scale events these biannual conferences have provided forums for debate about the issues that confront evaluators in diverse country contexts. They have enabled participants to share innovative experiences and to identify South-South solutions. For example, the 2013 event used a participatory approach that empowered participants, allowed in-depth discussions of evaluation issues, showcasing of best practices and culminated in a list of 18 NEC commitments: see *Actions and Com-*

mitments. Institutions and representatives from 52 countries formally endorsed these commitments thus highlighting their interest in collaborating and committing to national evaluation capacity development.

Emphasis was placed on interconnected aspects of establishing national M&E systems; follow up mechanisms for dialogue and enhanced network cooperation based on concrete agreements. This is expected to improve international standards for the evaluation of public programmes, policies and initiatives.

With live webcasts of key sessions, viewers from all over the world were able to join the proceedings. For each sub-theme parallel solution forums were organized. The solutions were then presented to plenary sessions to encourage broad based commitments to South-South cooperation on evaluation. This process highlighted several innovative initiatives designed to promote critical partnerships between governments, parliamentarians and VOPEs with a view to strengthen the demand for (and effective use of) evaluations. Four key evaluation capacity building elements were identified: 1) promoting evaluation use through national and global advocacy initiatives, 2) defining and strengthening evaluation processes and methods, 3) engaging existing and new stakeholders; and 4) exploring different institutional solutions for evaluation governance and management.

Looking ahead, it is expected that participants will keep engaging in knowledge exchange and cooperation to help implement their commitments. UNDP will monitor these efforts, track the results of cooperation agreements and help in follow up. It will also link interested parties to potential partners and UNDP programmatic units able to support south-south cooperation initiatives.

Examples of actions in progress, or planned to start soon, are:

1. Results based monitoring and evaluation framework under development by Interna-

tional Policy Center for Inclusive Growth to track governments and partners' efforts and results of NEC commitments.

2. Platforms for exchange of knowledge on NEC development under adaptation for upcoming dialogues linked to regional evaluation events in 2014 (NEC CoP: www.unteamworks.org/NEC).
3. NEC oriented events under the aegis of RedLACME, Peru (November 2013), AfrEA, Cameroon (March 2014); UNEG Bangkok; South Asia Regional Consultation on National Evaluation Policies; Sri Lanka (April 2014); Improving Statistical Data and Strengthening National Statistical Institutions by Brazilian Ministry of Social Development/EO. Rio de Janeiro, Brazil, August 2014; and at International Program Evaluation Network (IPEN) operating in the Newly Independent States in Kirgizstan (September 2014)
4. Publication by Parliamentarians Forum for Development Evaluation: "Mapping Status of National Evaluation Policies". December 2013
5. UNDP programmatic units to engage with interested parties and support implementation of commitments.

All in all the NEC Conference created an inclusive and collaborative process that engaged NEC Commitments' signatories and grassroots organizations in sustaining enthusiasm and momentum. Specifically partners, NEC participants and other stakeholders will engage in opportunities to exchange knowledge and explore how national governments and partners will implement the 18 NEC commitments through south-south cooperation. In 2015, the Independent Evaluation Office of UNDP will take stock of implementation progress on the NEC commitments. The outcome of these diverse initiatives will be presented to the 4th International Conference on National Evaluation Capacities to take place in Asia in the last quarter of 2015. To lay the groundwork for this synthesis the EES 2014 Biennial Conference to be held in Dublin in October 2014 will feature a session about the NEC development vision and the main challenges ahead.

ALBERT HIRSCHMAN AS EVALUATION THINKER

Robert Picciotto

Albert O. Hirschman, one of the most influential and original thinker of his generation, died in December 2012. He left behind a prodigious and elegant oeuvre that is hard to categorize since it spans disciplines, crosses borders, mixes mental constructs, rejects dogmas and reaches out beyond the academy to decision makers and the wider public. Who then is Albert Hirschman? He has been labelled one of the most distinguished social scientist of the past half century, an original development thinker, a master essayist and a worldly philosopher. He is all of these things but as this short article demonstrates he is also an evaluation pioneer.

Let me disclose at the outset that Hirschman disagreed with this proposition. He rejected out of hand my earlier efforts to recruit him as a member of our fledgling profession. He considered the notion far-fetched if not impertinent when he commented on a draft essay that extolled the evaluative dimensions of his writings. But I stood by my assessment which was included in a book designed not only to celebrate but also to contest and challenge Hirschman and his ideas¹.

A year earlier I had discovered that Hirschman was equally hostile to my notion of a distinctive 'Hirschman doctrine'. When as a newly appointed Director General of the World Bank Independent Evaluation Group I asked him for names of individuals who might help me promote his distinctive approach to the assessment of development interventions he wrote back: "*Unfortunately (or, I rather tend to think, fortunately) there is no Hirschman school of economic development and I cannot point to a large pool of disciples where one might fish out someone to work with you along these lines*"²

Hirschman's resistance to being pigeon holed and called an evaluator is not surprising: evaluation is widely considered as a discipline in its own right and Hirschman, a free spirit, made a virtue of trespassing across disciplinary boundaries. He used political science to inform economic thinking and experimented with economic concepts when faced with political dilemmas. He stoutly resisted the parsimony of the dismal discipline and injected in it "a bias for hope". He also reached out to psychology and sociology to help explain the unforeseen consequences of policy interventions.

It is in fact through his masterful interplay of the disciplines that Hirschman was able to weave seminal concepts that have illuminated our understanding of social phenomena – the hiding hand, the tunnel effect, the backward and forward linkages of industrialization, the exit, voice and loyalty trilogy, the rhetoric of reaction, etc. All of these constructs are now widely used by development economists, political scientists as well as sociologists and they all spring from evaluative processes. The fact of the matter is that Hirschman's technique closely resembles that adopted by evaluators who view evaluation theory as closely linked to policy research. Closing the gap between evaluation and the vast and varied domain of the social sciences is still a work in progress³.

To be sure, evaluators use program theory and results chains to fulfil their mandate defined by Michael Scriven as "*the determination of the merit and worth of programs in terms of how effectively and efficiently they are serving those affected, particularly those receiving, or who should be receiving, the services provided and those who pay for the programs*"⁴.

But the goals of such programs are set by decision makers and the assumptions and parameters embedded in logic models used by evaluators are those implicitly or explicitly defined by stakeholders. The analytical job of the evaluator is to establish the rationale of individual program theories, ascertain their validity and draw the surprising implications that often arise in the systematic confrontation between theory and practice.

This is precisely what Hirschman did throughout the odyssey of his extraordinary life. For example he subjected the certainties of free trade advocates and the rigid doctrines of balanced growth thinkers to judicious and subtle criticism grounded in his own direct observations and real world experience. While he was trained as an economist and was fully able to muster equations and data analyses to illustrate or demolish an argument Hirschman privileged down to earth evaluation methods.

His work illustrates the enormous potential of competent qualitative inquiry. Thus I still maintain that the evaluation community has a claim on Hirschman. While evaluation is a discipline it is also a trans-discipline and good evaluation draws on all the social sciences relevant to the topic at hand. Indeed Hirschman's predilection for inter-disciplinary work is a hallmark of good evaluation practice.

The other distinctive characteristic that Hirschman shares with evaluators is his commitment to independent inquiry. Throughout his long career as a policy adviser he jostled with and evaded capture by the establishment (as well as the politically correct counter-establishment). He was unimpressed by power

1 L. Rodwin and D.A. Schon, 1994, *Rethinking the Development Experience: Essays provoked by the work of Albert O. Hirschman*, Brookings Institution and the Lincoln Institute, Washington DC and Boston, Mass., pp. 210–230.

2 J. Adelman, 2013, *Worldly Philosopher: The Odyssey of Albert O. Hirschman*, Princeton University Press, Princeton and Oxford, p. 366.

3 J. Vaessen and F. L. Leeuw, Editors, 2009, *Mind the Gap: Perspectives on Policy Evaluation and the Social Sciences*, Comparative Policy Evaluation Series, Volume 16, Transaction Publishers, New Brunswick, NJ.

and status and debunked people who took themselves too seriously. He fiercely protected his independence of mind and appearance. Such attitudes and dispositions are precisely those that evaluation excellence requires.

Equally, as a policy thinker, Hirschman subjected prevailing doctrines to subtle and withering critique. But he also engaged in systematic self subversion of his own mental models if only to crush in the bud any notion that he stood for any particular school of thought or ideology. This is in line with one of the most critical capability of *bona fide* evaluators – a readiness to recognize and challenge one’s own biases and to go without hesitation where the evidence points to.

In terms of his research methods, just as good evaluation practice mandates, Hirschman observed events dispassionately, gathered evaluative information from a variety of sources and examined social interventions

without preconceptions of any kind. Unlike economists who are prone to search for evidence that fits their models or confirm their grand theories he looked at the world without pre-determined hypotheses. Similarly good evaluators search truth from facts and keep an alert eye on singularities. They appreciate that context matters enormously in the determination of outcomes and they evince deep scepticism regarding intellectual schemes that seek to explain everything.

Engagement with the world rather than retreat in the ivory tower of the academy is another reason why Hirschman is the evaluator “*malgré lui*”. He was committed to the freewheeling exploration of complex situations and relished engagement with decision makers. He embraced empirical field work, patient listening and thoughtful observation. Most of all he enjoyed travelling in the zones of turmoil and transition of the developing world and turned all his assignments into journeys

of discovery. His observational and evaluative skills are in evidence throughout his writings, especially the irreverent evaluation classic that he authored following a worldwide tour that he had carefully planned to probe the diverse dilemmas faced in the design and implementation of development projects⁵.

This thin volume remains unrivalled as a source of knowledge and inspiration for the foot soldiers of the evaluation profession. Like them Hirschman avoided sweeping policy generalizations and preferred to scrutinize the unique constellation of factors that make up individual social interventions. Yet by recollecting in tranquillity about his observations he was uniquely able in his own words to “*snatch systematic insight from casual hindsight*”. Illuminating the workings of social change in the way that he did is evaluation at its very best. So yes, over his objections, it is high time for Albert O. Hirschman to be inducted in the Evaluation Hall of Fame.⁶

4 M. Scriven, 1993, *Hard Won Lessons in Program Evaluation*, New Directions for Program Evaluation, no. 58, p.9.

5 A. O. Hirschman, 1995, *Development Projects Observed: with a new preface by the author*, Brookings Institution, Washington DC.

6 A panel chaired by Kim Forss (that Nicoletta Stame, Luca Meldolesi, Osvaldo Feinstein and I will join) is expected to delve deeper into Hirschman’s unwitting contributions to evaluation at the forthcoming 2014 EES Evaluation Conference in Dublin.

THE AUTHORS

Maria Bustelo, PhD

in Political Science, is Associate Professor in the Department of Political Science and Public Administration at the Complutense University, Madrid (UCM), Spain. She has worked as an independent consultant for Spanish public administrations and agencies, the European Commission, and NGOs in the field of programme evaluation, and in the areas of community development, health promotion, drug dependency prevention, and gender policies. She is also the director of the Master on Evaluation of Programmes and Public Policies (UCM), and member of the Board of Director Committee of the National Agency for the Evaluation of Public Policies and the Quality of Public Services. She leads the European QUING project team at the UCM (QUING: Quality in Gender+ Equality Policies -2006-



2011- www.quing.eu; a research project supported by the VI European Framework Programme). She has a number of publications on analysis and evaluation of public policies and gender equality policies.

Julia Espinosa

Julia Espinosa, post-doctoral researcher at Complutense University of Madrid, has worked as a researcher, an evaluator and an expert in gender equality and development in Europe and Latin America for the last ten years. Additionally, she has been teaching Evaluation, International Development Studies and Gender Equality at a postgraduate level since 2002.

Currently, she coordinates the ‘Gender and Evaluation’ Thematic Working Group of the EES and she is a member of the Associate



Experts Network – UNDP Regional Service Centre for Latin America and the Caribbean as well as the RELAC (Network for Monitoring, Evaluation and Systematization in Latin America and the Caribbean).

Her most recent publication in English is the paper “Moving towards gender-sensitive evaluation? Practices and challenges in international-development evaluation”, *Evaluation*, volume 19 (2013).

William Faulkner

William N. Faulkner is an evaluation consultant with a background in systems thinking and complexity theory. Until recently, William worked as an evaluation coordinator at Plan Políticas Públicas, an M&E consultancy based out of São Paulo, managing evaluation projects for a range of clients from local NGOs



to multinational corporations to municipal governments. In April 2014, William moved into the Integrated Monitoring, Evaluation, and Planning team of the Collaborative Crops Research Program (CCRP).

Oswaldo Nestor Feinstein

Professor at the Madrid Complutense University's Master in Evaluation. Senior evaluation adviser to the International Fund for Agricultural Development (IFAD). Editor of the Transaction Series on Evaluation and Development. Former lecturer at the European Institute for Public Administration and the Spanish Institute for Fiscal Studies, senior evaluation consultant with the ILO, World Bank, UNDP, Global Environment Fund, the United Nations Evaluation Group, UNESCO and African Development Bank.. Former adviser to the Spanish National Evaluation Agency. Former adviser and manager at the Operations Evaluation Department of the World Bank and senior evaluator at IFAD.



Kate McKegg

I am a New Zealander descended from Irish, Scottish and English ancestors. I live in Hamilton, New Zealand with my husband Richard and youngest son Patrick.



I have worked in the fields of evaluation, evaluation capacity building, research, policy and public sector management since the late 1980s. In 2002 I started The Knowledge Institute Ltd (www.knowledgeinstitute.co.nz), after a decade of working in policy, research and evaluation in the government sector. I have specialist skills in developmental evaluation, strategy, policy, and programme evaluation, evaluation capacity building, research, evaluation training and facilitation.

I am a founding member, and the current Convener of the Aotearoa New Zealand Evaluation Association (ANZEA). I have a deep and enduring commitment to the development of the evaluation field and profession because I firmly believe that evaluation, and evaluative thinking and practice can contribute to the betterment of our society.

Prof Zenda Ofir

is a South African international evaluation specialist with a PhD in Chemistry. She has conducted evaluation and knowledge management assignments in more than 30 coun-



tries, primarily in Africa and Asia, and provides strategic evaluation advice to a variety of international organisations across the world. She is a former President of AfrEA, former Vice-President of IOCE and former AEA Board member. She is a fellow of the Stellenbosch Institute for Advanced Study and a Honorary Professor at the School for Public Leadership at the University of Stellenbosch. She is currently based near Geneva, Switzerland.

Robert ('Bob') Picciotto,

(UK) Professor, Kings College (London) was Director General of the World Bank's Independent Evaluation Group from 1992 to 2002. He previously served as Vice President, Corporate Planning and Budgeting and Director, Projects in three of the World Bank's Regions. He currently sits on the United Kingdom Evaluation Society Council and the European Evaluation Society's board. He serves as senior evaluation adviser to the International Fund for Agricultural Development and the Global Environment Fund. He is also a member of the International Advisory Committee on Development Impact which reports to the Secretary of State for International Development of the United Kingdom.



Stephen Porter

Stephen Porter is currently a results and evaluation advisor for the UK Department for International Development. Formerly he was the Director of the Centre for Learning on Evaluation and Results for Anglophone Africa (CLEAR-AA) at the University of the Witwatersrand, Johannesburg. His has worked for over ten years in Africa on development issues with academia, government and civil society. The majority of this work has been related to the monitoring and evaluation of development. He has published a number of article, book chapters and reports related to evaluation capacity development.



Roberto La Rovere, PhD,

Development Economist, Evaluation Specialist at the UNDP IEO, managed the NEC 2011 and 2013 proceedings. Coordinator for Arab States Region of UNDP, brings 15 years of impact assessment and evaluation experience, also from research for development in the Middle East, Central and Latin America and Africa.



Ana Rosa Soares,

Evaluation Specialist, Independent Evaluation Office (IEO) of the United Nations Development Programme (UNDP). Ms. Soares, as coordinator for Latin America and the Caribbean Region and National Evaluation Capacity Development, has conducted thematic, regional and country level evaluation. She has a MA in Public Administration and International development.



Claudine Voyadzis

has spent some 20 years working in the evaluation field. Her work consisted mainly of appraisal and monitoring of projects and programmes, impact studies and ex post evaluations in West and East Europe and in Africa. She earned her Ph.D. in the United States in Demography and Sociology and has a dual experience of Anglo-Saxon and European approaches to evaluation. After working in the Independent Evaluation Group of the World Bank in Washington, she joined the Council of Europe Development Bank in Paris as technical adviser for project appraisal and monitoring. Subsequently she was appointed to set up the Ex Post Evaluation Department and remained for six years as Director of the Department. Claudine Voyadzis works now as an independent consultant, and leads evaluations for multilateral organisations and consulting firms.

