

EVALUATIONS
CONNECTIONS

EES

EUROPEAN
EVALUATION
SOCIETY

MAY 2012

ARTICLES – NEWS – EVENTS

President's message

María Bustelo, EES President

It is my great pleasure to present the second issue of "Connections" this year 2012. "Connections" is the Newsletter of the European Evaluation Society and it is meant to be a tool of communication among members through articles and news from our community. For this reason, I would very much like to encourage members to use this instrument to send comments or submit short articles or news for future issues. Although we are

busy now with the preparations of what promise to be an enthralling Conference at Helsinki in early October, there is always time and energy for planning new activities and debating topics of interest for our community. Remember that our society is made up from all of us, from our interests, aspirations, participation and implication! So please, do not hesitate to contact me or any of your Board members as needed.



Editorial

Claudine Voyadzis, EES Vicepresident

Welcome to the 7th edition of "Connections", the Newsletter of the European Evaluation Society. This issue presents articles that touch on a variety of evaluation challenges: in impact evaluation and more specifically in the use of randomised control trials, in evaluation of complex multi-component programmes, on how to assess "efficiency" and increase its theoretical potential, and in the last article very much in line with our next conference, on how to involve social media in evaluations.

We welcome comments on these articles (please send to: secretariat@europeanevaluation.org).

Moreover, we wish to encourage members to become more actively involved in EES activities. We have initiated our first Topical Working Group (TWG) on gender and evaluation, and we are working on setting up a TWG on evaluation in conflict stricken and

fragile countries. Should you be interested in these topics or should you wish to set up another TWG, please contact the EES Board (email addresses available on the EES website: <http://www.europeanevaluation.org/about-ees/board-members.htm>) or the EES Secretariat: secretariat@europeanevaluation.org. We also take this opportunity to remind you about our upcoming Conference in Helsinki on 1–5 October 2012: "Evaluation in the networked society: new concepts, new challenges, new solutions". The deadline for the call for proposals has now passed and we have received **430 proposals** for paper presentations, panels, symposia and roundtables.

We are thrilled with the Conference participation of **three remarkable keynote speakers** and we are honoured to welcome: Tarja Cronberg, Member of the European Parliament, former Member of the Parlia-

CONTENT

President's message	1
Editorial	1
Impact evaluations using RCTs – Reflections on the EES stance	2
The Challenges of Evaluating Complex, Multi-Component Programs	4
Making Evaluations Transparent, Participatory and Relevant in a Networked World	7
Analyzing ... what? Efficiency?	10

ment of Finland, former Minister of Labour in Finland and member of various advisory councils and think-tanks; Prof. Robert E. Stake, Professor Emeritus at the University of Illinois Urbana-Champaign and Director of the Center for Instructional Research and Curriculum Evaluation (CIRCE), and Robert Kirkpatrick, Director of the Global Pulse initiative of the United Nations Secretary-General.

As for biennial conferences in the past, EES has been searching actively for sponsors to provide bursaries to evaluators from developing and transition economies so that they can participate in the conference. This year, thanks to generous sponsors, EES will provide **bursary support to about 70 applicants**. We look forward to welcoming you in Helsinki for an exciting and professionally rewarding conference!

IMPACT EVALUATIONS USING RCTS – REFLECTIONS ON THE EES STANCE

Rahel Kahlert

We would all love to have a one-stop-shop for evaluation. Criticisms of aid effectiveness have been raised for decades, but recent developments have intensified the demand for more accountability. The 8th goal of the 2000 *United Nations Millennium Declaration* is more effective aid coupled with greater accountability, and the *Paris Declaration on Aid Effectiveness* (2005) calls for results-based and evidence-based management of development aid. In light of the greater stress placed on accountability, impact evaluation (IE) has become a welcome tool in determining the effectiveness of program interventions. However, the international community has yet to reach a consensus on where randomized control trials (RCTs) exactly fit into IEs.

Twice the European Evaluation Society (EES) has released statements on using IE for development interventions: In December 2007, EES stressed “the importance of a methodologically diverse approach to impact evaluation – specifically with respect to development aid and development interventions.” Again in April 2009, EES released the “Comments on the Draft NONIE Guidance on Impact Evaluation,” which addressed the guidelines set forth by the Network of Networks on Impact Evaluation (i.e., NON-IE). In both statements, EES raised concerns regarding the increasing preference for RCTs in evaluating program impact.

What triggered the EES response? In 2006, the Center for Global Development released their report “When Will We Ever Learn?” calling for more and more rigorous impact evaluations. These IEs would need to test the “net effect” directly attributable to a specific program. The group stated that “no responsible physician would consider prescribing medications without properly evaluating their impact or potential side effects,” and that therefore clinical trials had become “the standard and integral part of medical care.” Referencing the medical

model of experimental trials underscored the group’s preference for RCT evaluations.

The 2007 EES document was partially a response to what was widely perceived as a promotion of randomized IEs by the Center for Global Development; to various international institutions’ push for RCTs (e.g., World Bank’s Spanish Impact Evaluation Fund); and to the push by a fledgling of academic institutions to provide economist-led RCT evaluations for international development (e.g., The Jameel Poverty Action Lab at MIT). The 2009 EES comments criticized the donor-led Network of Networks on Impact Evaluation, on the grounds that its Guidance draft would stress quantitative methods, including RCTs, and would therefore lack credibility and usefulness.

In both cases, the main message sent by EES was that researchers should pursue methodologically diverse approaches for evaluating impact. Both documents criticized the notion of a “gold standard,” wherein the RCT would function as the apex in a design or framework hierarchy.

What does EES mean by “method”? The very expression “method,” derived from the Greek word for ‘way’ (*metodos*) is an ambiguous term with various meanings depending on context, though it most frequently refers to ways of data collection, data analysis, and overall function. EES seems to regard the “RCT” as an exclusive method—one either uses it or one does not. I would argue, in contrast, that the RCT is a sample-generating method, which should be combined with multiple methodological tools for data collection, analysis, and interpretation. An RCT is not by definition a quantitative method; it is rather a method that may equally employ qualitative elements. Like other approaches, an RCT attempts to answer the counterfactual question: What would have happened in the absence of the intervention? Yet, we can never know

the answer to that question unequivocally, even with an RCT. An RCT attempts to emulate the counterfactual by constructing an actual non-intervention group. Quantitative case studies use the treatment units themselves to simulate counterfactuals. Qualitative approaches may then elucidate the structures underpinning the counterfactuals. In any case, a counterfactual model is useful in an evaluation process, regardless of whether RCTs are used.

An RCT is a tool, which needs to be combined with several other tools to be meaningful. The RCT approach is distinguished from other frameworks by randomly assigning individuals or social units to an intervention group, on the one hand, and to a non-intervention group (or alternative intervention group), on the other. An effective RCT must meet several conditions:

1. There needs to be units that can undergo randomization. This excludes macropolicy interventions, such as adjustment of interest rates or changes in the tax code, where individuals or groups cannot be singled out.
2. Randomization needs to happen before initiating a program. Meeting this criterion is often impossible due to political or ethical constraints. Evaluations therefore often start in the middle of an intervention.
3. Baseline and outcome data need to be collected equally from both groups. High attrition rates often make this impossible. RCTs involving marginal groups, in particular, often suffer from high attrition.
4. Sufficient funding is needed for intensive data collection for both groups in at least two instances.

Given these conditions, RCTs are not always a feasible choice for evaluating impact.

Moreover, the notion that RCTs are the irrefutable gold standard is problematic.

As with other methods, an RCT only *approximates* the counterfactual. Overstating the RCT as the best policy tool may lead to blind-sidedness. In the 1930s, Ronald Fisher, the British agricultural statistician, had already cautioned against the non-critical use of RCTs. Statistical skill alone would not be sufficient. Rather, general intelligibility and common sense would be the priorities in designing, executing, and interpreting high-quality experiments (Fisher 1933). Fisher's contemporary and the medical statistician, Austin Bradford Hill, further criticized generalizing from an experimental sample to the general run of patients. Even today, medical trials often exclude elderly, adolescent, or comorbid patients. Thus, findings from large-scale experiments may not apply to individual patients, for whom a drug may act differently. Although Fisher and Hill wrote for the fields of agriculture and medicine many decades ago, their insights equally apply to current-day social programs. Social evaluators may point to the "success story" of RCTs in medicine, but we could equally point out its shortcomings within the medical field. The personalized-medicine movement and the focus on rare drugs, for example, seek ways to generate evidence beyond large-scale RCTs.

If RCTs are used, the following should be kept in mind:

(1) **Uncover qualitative-interpretive components in RCTs:** Every RCT is

interpretive in nature. Although RCTs measure differences between groups, they also rely heavily on qualitative-interpretive reasoning throughout the evaluation process. This includes clarifying the evaluation question, assessing prior knowledge, deciding on the nature and size of the sample, predicting potential causal effects, determining the needed baseline measures, interpreting the findings, and generating policy conclusions. All these steps require qualitative reasoning, which cannot be derived from pure cause-effect quantification. In fact, the quality of an RCT directly depends on the appropriate utilization of interpretive-qualitative skills when determining and understanding program impact. Inattention to these qualitative components in planning and executing an RCT seriously undermines the quality and the validity of RCT findings.

(2) **Clarify policy relevance of RCT findings:** No isolated experiment, however significant in itself, can suffice for demonstrating general program effectiveness. Applying RCT findings to other policy contexts, or moving from a pilot RCT to a large implementation of a program, requires establishing a theory of equivalency, a task that is necessarily qualitative in nature. This means determining the relations that need to hold between the RCT sample and the population in the policy context of interest.

(3) **Expanding evidence-base beyond RCTs:** Using multiple tools from

Rahel Kahlert



Rahel Kahlert has worked as an evaluator of higher and secondary education in several capacities at The University of Texas at Austin, involving national, state, and local-level policy questions. Her current research interest lies in methodological choices in impact evaluation. Rahel's Ph.D. research at the Lyndon B. Johnson School of Public Affairs at UT Austin covered the use of the medical RCT model in policy making, both for domestic and international programming. She received the Emmette S. Redford Award for her field research on evaluation practice. Rahel's original training was in Protestant Theology from the University of Vienna, Austria. After 13 years in Texas, Rahel plans to relocate to Europe next year. Please address correspondence to rahelck@gmail.com.

the methodological tool-box in tandem will likely increase the likelihood of arriving at robust evaluation findings.

The EES statements alerted us to take into account alternative approaches and methods for more policy-relevant evaluations. More research, however, is needed to further explicate these alternative approaches and methods.



3 – 5 October, 2012, Helsinki, Finland
(pre-conference workshops 1 – 2 October)

10th EES Biennial Conference
Evaluation in the networked society:
New concepts, New challenges, New solutions

www.ees2012.org

THE CHALLENGES OF EVALUATING COMPLEX, MULTI-COMPONENT PROGRAMS¹

By Jim Rugh and Michael Bamberger

The move towards complex, country-level development programming

In 2000 most major international development agencies agreed to collaboratively work towards the Millennium Development Goals (MDGs), reaffirming the growing recognition that in order to focus on the big picture, development should be structured around and evaluated in terms of a set of broad goals that encompassed the main areas of development. The MDGs recognized the need to develop a broad framework for assessing the overall contribution to development of the large number of projects and sector-specific programs being supported by different development agencies along with the governments of developing countries.

These kinds of interventions are commonly referred to as “complex” programs or development interventions. The term was coined to reflect the greater difficulties that development agencies experience to assess the effectiveness of their interventions in achieving their broad objectives as parts of larger, multi-agency collaborative programs. Conventional evaluation methods that may work relatively well for assessing the impacts of individual (relatively simple) projects normally cannot be applied to most broad-based development initiatives. In particular it has proved very difficult to define a counterfactual to address several related evaluative questions:

- “What would have been the situation if the intervention had not taken place?”
- “To what extent can the difference between the hypothetical no-intervention and the with-intervention situation be attributed to the effect of the intervention?”

- Just what is “the intervention” being evaluated, when there are many different policies and programs being implemented by many different agencies?

Simple projects, complicated and complex development programs.

It is helpful to begin by distinguishing between *simple* projects and *complex* programs or development interventions. Due to space considerations we will not discuss *complicated* programs, which fall between the other two categories (see RWE: pp. 396-401).

“Simple” projects: The term “simple” refers to the scope, relative clarity in the definition of objectives and the organization of the project. “Simple” does not mean that the project is “easy” to implement or that there is a high probability of success. Many “simple” projects operate in poor and vulnerable communities with high levels of insecurity and conflict, and success in achieving project objectives will often be quite low, particularly higher level objectives (e.g. sustainable improvements in human conditions) that funders expect to be achieved in a relatively short time period. “Simple” projects usually have many of the following characteristics:

- a. They are frequently based on a blueprint approach implemented in a similar way in each project location and intended to produce a uniform set of products or services.
- b. The number of project components or interventions is usually relatively small.
- c. The implementation procedures are usually relatively straightforward and require a low level of technical expertise; although

often requiring a high level of cultural sensitivity and communication skills.

- d. They usually have a clearly defined target population which is often relatively small.
- e. The objectives are usually, but not always, clearly defined.
- f. Projects typically have clearly defined start and end dates.
- g. There is often a defined budget, frequently from one major source.
- h. The project logic model is often described as being relatively linear with a defined and limited set of inputs expected to produce a defined set of outputs which in turn are presumed to produce a set of outcomes or impacts.

Complex programs are more difficult to characterize as there are many different scenarios, but they will usually have some of the following features:

- a. A number of different components and often a number of distinct programs.
- b. Usually large-scale, often covering the whole country or even several countries.
- c. Often a number of different donor agencies are involved in funding and perhaps implementing different components.
- d. Increasingly the interventions are country-led.
- e. There is often no clear definition of the range of services provided, the target population or the precise program objectives.
- f. While some programs cover a particular time period, many others are not time-bound.

All of these aspects have significant implications for the types of evaluation design that can be used.

¹ This paper is a summary of chapter 16 of Bamberger, Rugh and Mabry 2012 (Second Edition) *RealWorld Evaluation: Working under budget, time, data and political constraints*. Sage Publications. We refer to this as RWE.

Attribution, contribution and substitution analysis

Conventional forms of attribution analysis (experimental and quasi-experimental designs) can only be used when project beneficiaries can be clearly identified, when they all receive the same treatment (or one of a set of defined options) and when a well matched comparison group can be identified. Unfortunately, for the reasons discussed in the previous section, it is rarely possible to apply these rigorous designs to the evaluation of complex, national level programs. Consequently, many development agencies recognize the fact that they will only be able to use **contribution analysis** (Mayne 2008) to assess the plausible contribution of their agency to the changes that may be the result of the collaborative activities of many different development agencies and/or national governments and civil society. Actually, that is not a bad approach, as long as there can be convincing evidence of what each agency's contributions were to the achievement of higher-level impact. A complementary approach is **substitution analysis** (RWE pp. 403-404) where an assessment is made of the net increase in resources resulting from a donor's contribution after adjustments have been made for any transfers of previously committed national resources to other uses.

Alternative approaches for defining the counterfactual

Given the complexity of strategic interventions, and that many are intended to cover the whole country or sector, it is usually not possible to define a conventional statistical counterfactual. Consequently, there is a demand for creative approaches that development agencies can use in real-world contexts to assess what the situation would have been if the program had not taken place: in other words, to define alternatives to the conventional counterfactual. Alternative counterfactual designs can be categorized into 5 main groups (see Fig 1). They can only be listed here but are described in RWE Chapter 16 pp. 405-417 :

- Theory driven approaches
- Quantitative approaches: experimental and quasi-experimental designs, concept mapping, statistical analysis of comparator

countries, citizen report cards and social network analysis

- Qualitative approaches: realist evaluation, expert judgment, PRA and other participatory group consultations, public expenditure tracking, case study designs and comparison with other countries
- Mixed methods designs that combine the strengths of both quantitative and qualitative methods (RWE Chapter 14)
- Rating scales: these are often based on the guidelines developed by the OECD/DAC Network for Development Evaluation (OECD/DAC 2010).

Techniques for strengthening counterfactual designs

There are a number of techniques that can be used to strengthen most of the previously described counterfactual designs:

Disaggregating complex programs into simpler components: Many country or sector programs have a number of different components which makes it very difficult to define any kind of counterfactual. A first option is to disaggregate a multi-component program into its sub components and then conduct separate evaluations of each component. Many components may not have been implemented in all districts or states, and these areas provide a potential comparison group for this particular component. The findings from these component evaluations would be combined to provide an overall assessment of program effectiveness.

Comprehensive logic model: The design of the evaluation could probably be strengthened by developing a program theory model, including a pyramid-style, multi-level results-chain analysis, to define the intended linkages between different levels and the expected outcomes at each level. The program theory model should include a clear identification of the main rival hypotheses that could explain the observed changes together with the collection of evidence to test and eliminate the alternative explanations.

Reconstructing baseline data: Often the best option for defining the counterfactual will be to estimate the baseline conditions of the project group and a comparison group before the project began. A number of strategies can be used to reconstruct baseline data (see RWE Chapter 5).

Michael J. Bamberger

Independent Development
Evaluation Consultant



Michael Bamberger has a Ph.D. in sociology from the London School of Economics. He has 45 years of experience in development evaluation, including a decade working with NGOs in Latin America on housing and urban development, almost 25 years working on evaluation and gender and development issues with the World Bank in most of the social and economic sectors in Latin America, Africa, Asia and the Middle-East. He also has a decade as an independent evaluation consultant working with 10 different UN agencies, the World Bank and the regional development banks, bilateral agencies, developing country governments and NGOs helping design and implement development evaluations, organizing evaluations workshops and producing handbooks and guidelines. He has published four books on development evaluation, numerous handbooks on evaluation methodology, and articles in leading evaluation journals. He has been active for 20 years with the American Evaluation Association, serving on the Board and as chair of the International Committee. He has served on the Editorial Advisory Boards of the American Journal of Evaluation, New Directions for Evaluation, the Journal for Mixed Method Research and the Journal of Development Effectiveness, and is a regular reviewer for numerous professional evaluation journals. He is also a regular peer reviewer of evaluations for a number of bilateral and multilateral development agencies. He is on the faculty of the International Program for Development Evaluation Training (IPDET) where he lectures on conducting evaluations under budget, time and data constraints; the gender dimensions of impact evaluation; mixed method evaluation and the evaluation of complex programs. He has also been teaching for over a decade at the Foundation for Advanced Studies for International Development in Tokyo.

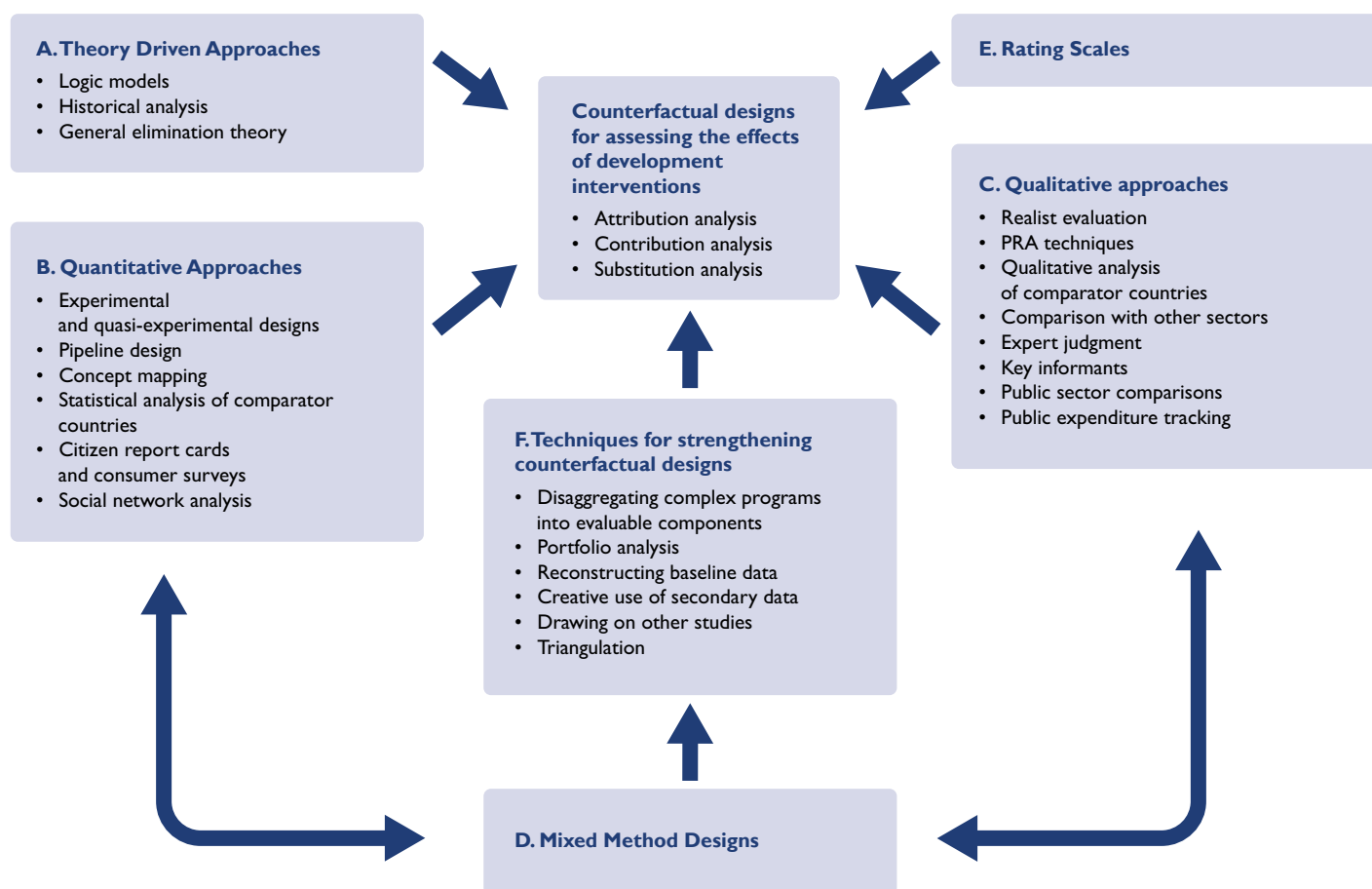


Figure 1. Statistical and alternative counterfactual designs

Creative use of secondary data: There are a wide range of potentially useful secondary data sources that are often overlooked. However, it is always important to assess these data sources to ensure the information is of good quality and relevant for the purposes of the evaluation.

Taking advantage of ongoing or planned studies: Sometimes the evaluation can take advantage of other studies that are being conducted or planned by government agencies, other donors, UN agencies, universities, NGOs or others. It may be possible to reach agreement with these agencies to include a few additional questions or to add a special module that will be administered to a sub-sample of the original sample. If the planned survey covers the universe from which the project population is drawn it may be possible to use

this as the comparison group to construct the counterfactual.

References

Bamberger, M; Rugh, J and Mabry, L. (Second Edition) 2012. *RealWorld Evaluation: Working under budget, time, data and political constraints*. Sage Publications. Funnell, S. & Rogers, P. 2011.

Purposeful Program Theory: Effective Use of theories of change and logic models. San Francisco. Jossey-Bass.

Kane, M. and W. Trochim. 2007. *Concept Mapping for Planning and Evaluation*. Thousand Oaks, CA. Sage

Mayne, J. 2008. *Contribution analysis: An approach to exploring cause and effect*. ILAC/GCIAR.

Morra, L. and Rist, R. 2009. *The Road to Results: Designing and Conducting Effective Development Evaluations* Washington, DC: World Bank.

Organization for Economic Cooperation and Development (OECD). 2010. *Evaluating Development Cooperation: Summary of Key Norms and Standards*. Second Edition. Paris.

Patton, M.Q. 2010. *Developmental Evaluation: Applying complexity Concepts to Enhance Innovation and Use*. New York. Guilford.

Pawson, R 2006. *Evidence-Based Policy: A Realist Perspective*. London. Sage

Jim Rugh

Jim has been professionally involved for 48 years in rural community development in Africa, Asia, Appalachia and other parts of the world. For the past 32 years he has specialized in international program evaluation. He served as head of Design, Monitoring and Evaluation for Accountability and Learning for CARE International for 12 years, responsible for promoting strategies for enhanced evaluation capacity throughout that world-wide organization. He has also evaluated and provided advice for strengthening the M&E systems of many other international agencies.



He is recognized as a leader in the international evaluation profession. From 2008–2011 he served as the AEA (American Evaluation Association) Representative

to the IOCE (International Organization for Cooperation in Evaluation), the global umbrella of national and regional professional evaluation associations, where, as Vice President, he was an active member of the Executive Committee. He has now been asked by UNICEF and IOCE to coordinate the EvalPartners Initiative that aims to strengthen evaluation capacities of and through Voluntary Organizations of Professional Evaluators (VOPEs) around the world.

Jim co-authored the popular and practical *RealWorld Evaluation book* (1st edition 2006; 2nd edition published by Sage 2012, see www.RealWorldEvaluation.org) and has led numerous workshops on that topic for many organizations and networks in many countries. For 25 years he has been actively involved in AEA's International and Cross-Cultural Evaluation Topical Interest Group, and was the founder of

InterAction's Evaluation Interest Group. In recognition of his contributions to the evaluation profession he was awarded the 2010 Alva and Gunnar Myrdal Practice Award by AEA.

In addition to his global perspectives on the evaluation profession, and his M&E expertise at the HQ level of International NGOs, Jim brings experience in community development and in evaluating and facilitating self-evaluation by participants in such programs. He is committed to helping the poor and marginalized work on self-empowerment and development, and to encouraging appropriate and effective assistance offered to them. He brings a perspective of the "big picture," including familiarity with a wide variety of community groups and assistance agencies in many countries, plus an eye to detail and a respect for inclusiveness and the participatory process.

MAKING EVALUATIONS TRANSPARENT, PARTICIPATORY AND RELEVANT IN A NETWORKED WORLD: USE OF SOCIAL MEDIA IN DEVELOPMENT EVALUATION

By Alex McKenzie and Bahar Salimova, Independent Evaluation Group, World Bank

The world's increased connectivity has fostered new ways for people to reach out and communicate. With the availability of new channels (both online and mobile) people have an opportunity to share information and interact with each other instantaneously. Such connectivity also opens up opportunities for engaging and learning from one another as well as possibilities for sharing and enriching knowledge at a larger scale.

This note discusses how the Independent Evaluation Group (IEG) of the World Bank Group uses social media to engage with stakeholders at large, to streamline user feedback into evaluations, and to strategi-

cally strengthen transparency and participation in its evaluation processes.

Engaging with broader groups

The prevailing lenses for looking at the use of evaluative knowledge have been based on knowledge diffusion and knowledge translation processes. These are typically done through knowledge stock-taking such as synthesizing of existing resources and compiling lessons learned, and through ensuring knowledge flow by organizing learning events, online discussions, etc. IEG expanded this approach to include the concept of building relationships with stakeholders as another

factor for effective knowledge transfer and learning. More specifically, IEG pursues relationships with stakeholders through social media where interactions happen in various formats and dimensions. Hon & Grunig¹ elaborate on the gamut of relationships that organizations can build with their key constituents ranging from establishing trust and going all the way to having "communal" relationship where both sides provide benefits to each other.

As the number of social media channels and information outlets grow on a daily basis, organizations must deal with the imminent question of how to find and establish effec-

¹ Hon, L. and Grunig J. Guidelines for Measuring Relationships in Public Relations, 1999.

Bahar Salimova



Bahar Salimova is an Information Officer with the World Bank's Independent Evaluation Group (IEG). At IEG, Bahar is leading social media work and working on knowledge sharing practices. Prior to joining IEG, Bahar worked with the United Nations Development Programme (UNDP) as a Knowledge Management Specialist. While at UNDP, she helped to establish the International Knowledge Network of Women in Politics (iKNOW Politics) conceptualizing project's website and resources, creating a community of practice, authoring knowledge resources, and designing learning practices and strategies.

(bsalimova@worldbank.org)

Alex McKenzie



Alex McKenzie leads the Knowledge and Information Management Team at the World Bank's Independent Evaluation Group (IEG). He is responsible for planning and delivering solutions that blend public relations, institutional learning, and online media. He has prior experience in design and implementation of process automation systems, web architecture, electronic media, and strategic planning. Before joining the bank in 1994, Alex worked with regional information services firms and with IBM Peru. He has a MBA and a degree in Electrical Engineering.

(amckenzie@worldbank.org)

tive relationships with individuals consuming information through social media. An even more important question is how to reach the right audiences and establish relationships with those who will eventually become users of our information, product or message and champion it among others. According to Powell, Groves and Dimos, people participating in social media can be roughly categorized as influencers, consumers, and general individuals². Figure 1 shows the relationship between these categories. Powell et al., explain the interplay between these categories through the 90-9-1 rule, where 90 percent of the online users only consume shared content (*individuals*, elsewhere referred to as *lurkers*³), 9 percent prioritizes their engagement over other things only when the conversation strikes them as important or relevant, and 1 percent *influencers* participate intensely and contribute most of the content. In general, organizations try to find the 1 percent of active, influential social media users interested in their work, brand, or message who can be their "multipliers." An even a harder task is to identify, build and retain relationships with the 9 percent of possible users and provide them with the relevant information and message that they would react to.

IEG became active in social media almost two years ago. The initial goal was to share IEG's knowledge and resources with broader communities and to engage stakeholders in discussions on particular issues, complementing other more traditional media and web outreach. To "speak in the language of social media", IEG started producing succinct and easily understandable content based on evaluation reports, available through social channels, so that a broader audience, particularly those who are not evaluation experts, could understand IEG's key messages. This new content varied in its format: from video interviews, podcasts, interactive maps, case studies and synthesis notes. IEG also engaged its senior management in blogging as a form to share knowledge and define its position on certain issues. Through blogging, IEG was

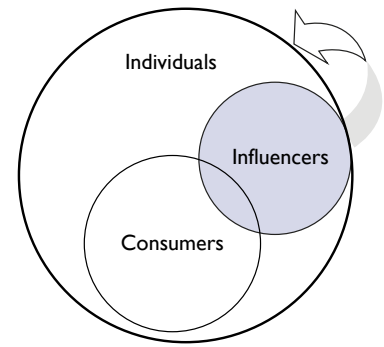


Figure 1: Source: ROI of Social Media: How to Improve the Return on Your Social Marketing.

able to reflect on its accumulated knowledge base during critical moments when instantaneous reaction was required such as during natural disasters in Haiti, Pakistan, Brazil and other countries in 2010, during the latest financial and food crises, and other international development issues.

From Communicating to Providing Learning Opportunities and Creating Participatory and Transparent Evaluations

As IEG's presence in social media solidified and brought initial results, it expanded into making the evaluative process more transparent and participatory. IEG started using social media to hold outreach campaigns for its ongoing evaluations by asking users to share their knowledge on key evaluative questions. IEG's most recent social media outreach campaigns were on youth employment, sustainable forest management and impact of development projects in Afghanistan. In the case of the youth employment campaign, IEG used its existing Facebook and Twitter accounts to reach out to target groups with key messages and questions that were integral to the upcoming evaluation of World Bank Group's Support to Youth and Employment. The questions were asked se-

2 Powell, G., Groves, G., and Dimos, J. ROI of Social Media: Hot to Improve the Return on Your Social Marketing, 2011.

3 Nielsen, Jacob. Participation Inequality: Encouraging More Users to Contribute. 2006. http://www.useit.com/alertbox/participation_inequality.html.

4 Lewis, Pea, and Rosen. Co-creation of meaning using mobile social media. P. 8. 2010.

quentially and in multiple formats, including open-ended questions and polls. Qualitative data gathered from social media enriched the evidence base, and was analyzed, triangulated with other sources of data, and incorporated into the report.

Contemporary studies suggest that learning occurs not only in formal classroom environments, but also in informal circles and interactions. Social media in that regard plays a key role as it allows participants to learn from one another and to learn by molding and shaping content in various formats. The power of social media for learning lies not so much in its ability to offer individual expression anytime anywhere, but in its potential to foster collaboration, on a scale and in tighter cycle times than ever seen before.⁴ One of the interesting elements of engagement design that emerged in IEG's learning and knowledge sharing cycles, is combining online and offline activities. For example, IEG successfully expanded discussions initiated during face-to-face workshops, to a broader online audience in order to further enrich the experience. It engaged with the gender and development community for its 2010 evaluation of World Bank's Support to Gender and Development (1998–2008). Part of the engagement plan for the study was to utilize social media to build relationships among participants attending workshops in Viet Nam and South Africa, and to foster and sustain dialogue afterwards. IEG set up an online social collaboration platform to request input from users on the recommendations and knowledge generated during the workshops. IEG also complemented this effort with video conferencing sessions that connected 12 countries and over 100 people to discuss the findings of the evaluation study, and recommendations received at the workshop. Each video conference included key stakeholders: responsible government officials, representatives of nonprofit organizations, World Bank Group staff, and academia. Holding several rounds of consultations and allowing workshop and video conference participants and a broader audience to continuously share and comment on generated knowledge, helped IEG to enrich

the experience and create a broader knowledge base useful for future studies. Additionally, it established a feedback loop, which is an often overlooked aspect in the evaluative process. Describing the "Learning Spiral" in a governmental learning environment, Blindenbacher and Nashat describe continuous feedback-seeking as action learning, where new knowledge is continuously validated and updated, and as a consequence becomes a potential new state-of-the-art knowledge for other learning systems⁵. Or more simply said, through an iterative process IEG's knowledge and content serves as a catalyst for broader discussions with other stakeholders and organizations, who can adapt the enriched knowledge to make it more relevant, and to drive action and positive change.

Challenges and Way Forward

As many other organizations, one of the challenges that IEG is facing on social media is to deepen its relationship with target audiences. It is not enough to just publish reports or post content online, and expect it to have an impact. For the information to be acted on, it is important to have a relationship with your audiences, particularly the ones who we consider key stakeholders. Having thousands of Facebook fans (or any other outlet) may sound impressive, but it does not necessarily ensure meaningful and mutually satisfactory engagement. IEG wishes to see itself both as a supplier of information and knowledge, and as a consumer. A two-way process can foster stronger relationships, expand feedback and understanding of issues, and enhance IEG's knowledge base.

Another challenge is transforming complex evaluative language into a simpler and more concise form, not only for the purposes of social media outreach but also for studies to be more easily understood. Many believe that it is important to write in a complex and long prose to explain complicated issues that development evaluations look at. However, for information to be widely understood and consumed outside specialized circles, writing clearly is a must.

IEG's strategy to use social media to gather feedback and qualitative data for its evaluations is a relatively new endeavor. These efforts have been largely positive, bringing many lessons learned along the way. But it will require further involvement from other evaluation entities and professionals to understand the best way to employ social media for evaluative knowledge and to use the data it provides.

References

1. Lewis, Pea, and Rosen. *Co-creation of meaning using mobile social media. Social Science Information*. Vol. 49 – no 3. P. 8. 2010. http://www.stanford.edu/~roypea/RoyPDF%20folder/A169_Lewis-Pea-Rosen_SSI_2010.pdf. Downloaded on October 15, 2011.
2. Paine, K. *Measure What Matters: Online Tools for Understanding Customer, Social Media, Engagement and Key Relationships*. p. 75. 2011.
3. Hon, L., and Grunig J., *Guidelines for Measuring Relationships in Public Relations, 1999*
4. Powell, G., Groves, G. and Dimos, J. *ROI of Social Media: Hot to Improve the Return on Your Social Marketing*. 2011
5. Nielsen, Jacob. *Participation Inequality: Encouraging More Users to Contribute*. 2006. http://www.useit.com/alertbox/participation_inequality.html
6. Braffman, O. and Beckstrom, R. *The Starfish and the Spider: The Unstoppable Power of Leaderless Organizations*. 2008.
7. Blindenbacher, R., and Nashat, B. *The Black Box of Governmental Learning: The Learning Spiral*. p. 159. 2010.
8. Performance Juxtaposition Site. *Instructional Design – Social Learning and Social Media*. 2010. http://nwlink.com/~donclark/hrd/media/social_learning.html. Downloaded on October 16, 2010.

5 Blindenbacher, R., and Nashat, B. *The Black Box of Governmental Learning: The Learning Spiral*. p. 159. 2010.

ANALYZING ... WHAT? EFFICIENCY?

WHAT WE CAN DO ABOUT CLOSING THE GAP BETWEEN WHAT IS **EXPECTED** AND WHAT IS **DELIVERED** WHEN EFFICIENCY OF DEVELOPMENT INTERVENTIONS IS ANALYZED.

Munich, May 7, 2012 by Markus Palenberg and Michaela Zintl

What is efficiency? Efficiency is a stretched term, comprising a variety of concepts that differ along different dimensions (also illustrated in Figure 1):

- Efficiency can either be defined by *transformation* of inputs into results (such as benefit-cost ratios or unit costs) or by *optimization* of net quantities (such as net benefits and utility).
- *Production efficiency* is restricted to output-level results (as in unit costs) while *allocation efficiency*¹ includes outcome level effects (as in benefit-cost ratios).

While general measures to improve efficiency usually address both inputs and results (or costs and benefits), two special cases address only one of both: *yield maximization* aims to increase results with a fixed amount of inputs and *cost minimization* searches to reduce the amount of inputs required to produce certain results.

Last but not least, when assessing methods for efficiency analysis, it proves useful to differentiate between three *efficiency analysis levels*:

- Level 2 analysis, the most potent, is capable of assessing the allocation efficiency of

an aid intervention so that it can be compared with alternatives or benchmarks.

- Level 1 analysis is capable of identifying the potential for efficiency improvements within aid interventions. Level 1 analysis can be a by product of a level 2 analysis.
- Finally, level 0 analysis is entirely descriptive and can usually not produce well-founded recommendations.

Level 1 and 2 analyses have complementary functions: while level 2 analyses inform the choice between different intervention alternatives, level 1 analyses primarily help to improve individual intervention.

Why is efficiency important? Efficiency is a powerful concept for rational decision-making. At least in theory, welfare can be maximized based on efficiency information alone and efficiency would therefore represent the most important criterion in appraisals and evaluations.

In practice, however, challenges such as the limited scope of efficiency analysis, model simplifications and calculation approximations used reduce this theoretical potential. Because of these limitations in practical applications, efficiency is often not

the principal decision-making criterion but rather informs decision-making by providing key information.

Therefore, the power of efficiency analysis depends on the amount, the relevance and the realism of efficiency-related information. However, even without *any* accurate efficiency-related information at all, the *concept* of efficiency remains important for informing a welfare-maximizing approach to development.

The gap between expectation and delivery of efficiency analysis.

Evaluation guidelines and national budget codes raise expectations and define obligations regarding efficiency analysis. Many evaluation manuals and guidelines list efficiency among the standard evaluation criteria as, for example, the OECD DAC Criteria for Evaluating Development Assistance. The German national budget code prescribes adequate efficiency analysis for *all* measures – not restricted to development – with financial impact on the national budget. A United States White House executive order mandates federal US agencies to conduct Cost-Benefit Analysis (CBA) for significant regulatory actions.

This high importance of efficiency analysis stands in sharp contrast to the frequency and quality with which it is applied in appraisals and evaluations of aid interventions. Several studies have shown that efficiency is often analyzed with low frequency and quality – both with respect to absolute standards as well as relative to the frequency and quality with which other evaluation criteria are analyzed.

Efficiency described as ...	Production Efficiency (input/output)	Allocation Efficiency (input/outcome)
... ratio	<ul style="list-style-type: none"> • Unit cost • Cost per person reached 	<ul style="list-style-type: none"> • Benefit-Cost ratio • Cost-Effectiveness ratio
... net quantity	<ul style="list-style-type: none"> • Financial profit • Net present value 	<ul style="list-style-type: none"> • Net present benefit • Aggregated utility

Figure 1. Examples for efficiency measures along different dimensions.

¹ We use the term „allocation efficiency“ in a more narrow sense, and on the level of single interventions, than the related term “allocative efficiency” that describes resource allocation in markets.

A study conducted by the World Bank Independent Evaluation Group, for example, investigated the use of Cost-Benefit Analysis (CBA) in World Bank project appraisals and found that the frequency with which CBA had been applied in project appraisals had fallen from about 70 per cent of all projects in the 1970s to about 30 per cent in the early 2000s, and in recent years. This stands in contrast to World Bank policy that mandates economic analysis, mostly CBA, for all investment projects. In terms of quality, the share of projects with acceptable or good economic analysis in appraisal documents has declined from 70 per cent in a 1990 assessment to 54 per cent in a similar assessment conducted in 2008.

What can be done to close the gap?

On the basis of our research, we draw four general conclusions. The first two illustrate how efficiency analyses can be applied more widely and with higher quality:

- First, the application potential of efficiency analysis methodology is not exhausted, both in terms of frequency and quality. Several existing efficiency-analysis methods are little known and/or ill documented but can complement more established methods or deliver results when the latter cannot be applied. Examples are Cost-Utility Analyses, Methods for Multiple-Attribute Decision-Making, and more pragmatic methods such as the “Follow the Money” approach.

In addition, for some methods, the evaluation design needs to be changed from vertical assessments that evaluate several criteria for a single intervention to horizontal assessments that focus on the efficiency criterion across several comparable interventions.

- Second, some efficiency analysis methods presented are considerably less developed than others. A number of simple and common-sense based approaches may benefit from professionalization and standardization as, for example, comparative stakeholder ratings. In contrast, highly sophisticated approaches may benefit from rendering them more practicable and adaptation to different sectors. For example, Cost-Utility Analysis is virtually absent in efficiency analyses of aid interventions outside of the health sector.

Yet, even if frequency and quality of efficiency analysis are increased in these ways, expectations will not be entirely fulfilled. We therefore recommend also clarifying and specifying expectations in two ways:

- Third, expectations regarding efficiency analysis need to be adapted to what current and near-future methodology can realistically accomplish. This does not necessarily imply lowering expectations but rather clearly specifying the purpose for conducting efficiency analysis. The analysis levels introduced earlier allow for such

Markus Palenberg

*Managing Director,
Institute for Development
Strategy*



Markus manages the Institute for Development Strategy (IfDS), an independent research institute in Munich, Germany (www.devstrat.org). IfDS researches evaluation methodology, conducts theory-based evaluations of complex interventions, and advises programs on strategy, governance and M&E. Markus is Management Team Member at Harvest-Plus (www.HarvestPlus.org), member of the Executive Board of the Generation Challenge Program (www.GenerationCP.org), member of the Technical Advisory Panel on Global and Regional Partnership Programs of the World Bank Independent Evaluation Group (ieg.worldbankgroup.org/content/ieg/en/home/initiatives/global.html) and fellow of the Global Public Policy Institute (www.GPPI.net). Before founding IfDS, Markus was managing director of GPPI’s consulting practice, worked as corporate manager in one of Germany’s largest online marketplaces, as strategy consultant with McKinsey&Company, Inc., and as postdoctoral researcher at the Massachusetts Institute of Technology. Markus holds a Doctorate in Theoretical Physics.

Degree to which method is known	Level 2 methods	Level 1 methods	Descriptive methods
Well-known methods	Cost-Benefit Analysis (CBA)	Benchmarking of unit costs	Expert judgement
Somewhat less well-known methods	Cost-Effectiveness Analysis (CEA)	Follow the Money	
		Financial analysis	
		Stakeholder-driven approaches	
Methods unknown to a substantial fraction of evaluation experts	Multi-Attribute Decision-Making (MADM): Intuitive scoring models	Comparative ratings by stakeholders: Comparative rating of effectiveness and cost analysis	Specific evaluation questions on efficiency
	Multi-Attribute Decision-Making (MADM): Scientific decision analysis		
	Effects Method	Comparative ratings by stakeholders: Comparative rating of efficiency	
	Cost-Utility Analysis (CUA)		

Michaela Zintl

Head of the Evaluation
Division, Federal Ministry
for Economic Cooperation
and Development



The evaluation division of the German Federal Ministry for Economic Cooperation and Development (BMZ) is in charge of strategic evaluations of German development coop-

eration as well as of norms and standards of evaluations conducted by implementing partners. Summaries of these evaluations are published on the Ministry's website. Contributing to international methodological discussions – in particular their adaptation to the German context – has been an additional focus of work over the past years. The unit is a member of the OECD DAC Network of Development Evaluation and has been one of the founding members of the Network of

Networks for Impact Evaluation (NONIE), organizing the annual conference 2010 in Bonn.

Michaela has previously worked in various capacities in the ministry and with implementing agencies and served as an executive director – representing a diverse constituency – of the Inter-American Development Bank (IDB). She is an economist by training, studying in Austria, Germany and the US.

a specification. For projects and simple programs, we estimate that some level 2 and level 1 analyses should always be possible. This implies that the efficiency of several alternatives can be compared to each other and that efficiency improvement potential within specific alternatives can be identified. For more complex and/or aggregated aid modalities, we consider that efficiency assessment is usually limited to level 1 analysis. This implies that for these types of aid, the expectation of selecting the most efficient option by means of efficiency analysis alone, as for example in aid modality comparisons, needs to be reconsidered. For these types of aid, efficiency analysis is realistically restricted

to identifying operational improvement potentials.

- Fourth, efficiency analysis should not be conducted whenever it is analytically possible. Instead, we recommend choosing carefully when to apply it. Efficiency analysis itself also produces costs and benefits. Depending on circumstances, the benefits of efficiency analysis may not justify its costs as for example in expert judgments with low credibility, level 2 analyses without influence on the selection of interventions or efficiency analyses of interventions that are already known to have either very high or very low efficiency.

Available methodology. Overall, 15 distinct analysis methods have been identified and are described in the study report. The following table provides an overview of these methods, ordered according to the analysis level and the degree to which methods were known by the experts interviewed for this study.

This initial collection of methods is currently being updated and some methods will be further investigated in the context of a OECD DAC Evalnet Working Group.

Interested? Go to www.AidEfficiency.org for the full report and references.



Evaluation in the networked society:
New concepts, New challenges, New solutions.

Notice of the 2012 EES Annual General Meeting (AGM)

2012 EES Annual General Meeting
will be held during the 10th EES Biennial Conference
3–5 October, Helsinki, Finland.

All EES members are cordially invited to participate.
More information and agenda will follow soon.