3rd International Conference on National Evaluation Capacities: Innovative Solutions to Challenges Linked to Independence, Credibility and Use of Evaluations

Guiding Thematic Paper on Credibility

María Bustelo
European Evaluation Society
Universidad Complutense de Madrid
mbustelo@cps.ucm.es

____

This paper aims to open a discussion on the credibility of evaluations and on how to enhance and ensure credibility. 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 but some. 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.

My standing point in writing this paper comes from my experience of training evaluators and conducting the Masters programme on Evaluation of Programmes and Public Policies at the Universidad Complutense de Madrid (Spain), which will start its 12th course in 2013-2014. It also comes from my service at the Board of the European Evaluation Society since 2009, the last two years of which I have been honoured to lead the presidency of this regional-based society. For context, I will be using Spain, which is my own country where I have been working on evaluation for the last 25 years, but also a more ample and diverse European region. And I will be referring more than once to the inspiring EES Public Hearing at the European Parliament on 'Evaluation in Democracy' held in Brussels in April 2013.[1]

1. The political nature of evaluation and the need for credibility

The political nature of evaluation has been amply recognized by the evaluation community since the late 1980s. In her excellent key note address at the American Evaluation Association's annual meeting in San Diego in 1997, which was later published as an article (Chelimsky, 1998), Eleonor Chelimsky talked about "the role of experience in formulating theories of evaluation practice". She states that evaluation and politics are "viscerally connected". For her, "evaluations have been deeply affected by the way politics works in a

[1] www.czech-in.org/ees/ees-newsletter-2013-06-june-special.pdf.

democratic, pluralistic society, and especially by the continually changing nature of the political environment". Also, because "in a world of highly sophisticated and continuous jockeying for political advantage, advocacy abounds. Not only do policy makers have their own political agendas, they are also besieged by pressure groups, vested interests and lobbyists, all with their war stories about 'success' or 'failure', and all trying, with money, power, and data, to move policies and programs in specific directions" (Chelimsky, 1998: 37-38). So we need to be aware and include that political nature in the very heart of the evaluation concept and theories. That need for credibility is particularly important in considering politics as central to evaluation practice. For Chelimsky, credibility implies impartiality and that should be something that is preserved in the evaluation process, while evaluations need to be perceived as impartial. So not only should they *be* impartial, they should be *seen* to be impartial by stakeholders. In her substantial experience in the USA's General Accounting Office, this impartiality was often preserved by correcting possible perceptions of an evaluator's bias by hiring the opposite bias in the same evaluation team, as well as by correcting this potential bias during the evaluation process, especially in the literature review, the methodological design and the report of the evaluation phases. These procedures were useful in a context such as the General Accounting Office, which is part of the USA's Legislative Branch and directly serves Congress. In other contexts, evaluation credibility might be more derived from the transparency of the process than from an absence of biases, but whatever the strategy it is crucial that an evaluation is conducted so that it is perceived by the general public and the stakeholders as credible.

Indeed, even in contexts less politicized than the General Accounting Office, the political nature of evaluation, poses challenges for the independence, credibility and use of evaluations. And that is why evaluation is about much more than dealing with methodologies and techniques that allow us to acquire good enough evidence of what has happened in a concrete project, programme or policy. Or, as I like to say, the methodology of evaluation needs to go far beyond what it is understood as social science research, including most of what is called applied social research. We need specific ways to think about and conduct evaluations that are unique to evaluation and embedded in its political nature and in its purpose – improvement, accountability and/or enlightenment.

The political nature of the evaluation means that, apart from providing systematic and rigorous evidence, it rises, should be tailored, and must respond to and from a particular context to be credible and usable. This political nature should also be acknowledged in independent evaluations. Even if it appears to be apolitical, an evaluation is invariably and inevitably political, because it is formed from the interactions of actors with different interests (some of them in apparent contradiction, but all of them legitimate), perspectives and power positions. There is interdependency among the actors, and they operate in a concrete territory and period of time. Any project, programme or policy I can think of that can be subject to evaluation can be described in such a context.

This important presence of different actors and stakeholders brings us to the next point.

2. Who participates and has a say in the evaluation? A key issue for credibility

Credibility is also inevitably linked to the participation of stakeholders in the evaluation processes. A participatory approach to evaluation, apart from the benefits of inclusiveness, promotes stakeholders' ownership of evaluation processes and results. Consequently, it also raises perceptions of the credibility of those processes and results by those who have participated and feel they have a say in the evaluation.[2] Of course, this is easier said than done and real stakeholder participation is a process that requires time and political sensitivity by commissioners and evaluation teams. Getting to know and understand the context and the programme to be evaluated, identifying key stakeholders and their information needs, including those needs in the evaluation questions and contrasting or even negotiating them with stakeholders, gathering systematic information and evidence from stakeholders and other sources, and contrasting and/or elaborating collaborative conclusions and recommendations, are some of the ways that stakeholders can be involved. That will help the evaluation to be perceived as more credible, because the process has been transparent and stakeholders have had opportunities to contribute to several parts of the process.

Often, the context in which an evaluation takes place, or constraints of time and resources and of terms of reference, do not allow real stakeholder participation in *all* the different parts of the process. But we should be aware that to enhance credibility it is necessary to methodologically design an evaluation process in which stakeholder participation in some parts of the evaluation process is possible.

So far, we have discussed stakeholder participation as a methodological requirement for enhancing ownership and hence evaluation credibility. But there are other issues regarding the rationale for stakeholders' participation that go beyond this rather utilitarian and pragmatic perspective, that are also crucial for evaluation credibility. Many evaluation theorists remind evaluators about the centrality of stakeholder participation. For example, Monnier talks about the importance of the 'social utility' of evaluation. That social utility can only be attained with the participation of all stakeholders involved in the programme or public policy to be evaluated. The evaluation, he says, should not only have institutional, technical or scientific legitimization, it should have, above all, political legitimization, which is given by the stakeholder participation. However, going through the motions of seeking participation without a genuine intention to maximize participation could lead to favouring those that hold the levers of power. The deliberative democratic evaluation model (House & Howe, 2000), for example, proposes procedures that ensure that the disadvantaged are respectfully included in the evaluation process. For House, the deliberative democratic evaluation "aspires to arrive at unbiased conclusions by considering all relevant interests, values, and perspectives; by engaging in extended dialogue with major stakeholders; and by promoting extensive deliberation about the study's conclusions—in addition to employing traditional evaluation methodologies" (House, 2005). This recognition of structural and systemic inequalities, which produce, by default, biases or preferences for the favoured or powerful, is also a key reason

[2] The norm 10.1 of the "Norms for Evaluation in the UN System", published by UNEG in 2005, reads: Transparency and consultation with the major stakeholders are essential features in all stages of the evaluation process. This improves the credibility and quality of the evaluation. It can facilitate consensus building and ownership of the findings, conclusions and recommendations" see www.uneval.org/documentdownload?doc_id=21&file_id=562

for the need for an equity-based, gender-sensitive and human rights perspective in evaluation. All these approaches remind us that there is a need to make deliberative efforts to level the playing field by allowing the weakest to have a voice in the evaluation process.

Stakeholder participation is neither simple nor straightforward. New forms of policy-making and new instruments are necessary to deal with new and complex public problems and policy challenges locally as well as nationally, regionally and globally. Such complexities increase the scope for democratic participation and the number of stakeholder interactions. Governments have moved from a centralized and hierarchical, top-down form of government to a form of governance that involves multiple decentralized and contested types of public and private actors. To be credible in this changing environment, evaluation practice and theory cannot ignore new realities and offer simple evaluative designs for complex situations and interventions.

Another important perspective with regard to credibility and stakeholder participation is the citizenship's perspective. In democracy, as Tarja Cronberg, European Parliamentarian and host of the EES 'Evaluation in Democracy at the European Parliament' event, says, evaluation can be a tool for the empowerment of citizens. It can be also a way of promoting social learning, identifying priority policy interventions, and reducing the democratic deficit, in this case, in the European Union. At the same time, citizens must be able to trust institutions for evaluation to play a role. But if an evaluation is planned, implemented and used properly from a citizen's perspective, it should also help to recover and build institutional trust.

3. Credibility not only depends on the quality and independence of the evaluators, but also on the institutions and the systems where the evaluations are conceived, planned and managed

Evaluators do not play a lone role in evaluation. They are hired by commissioners who establish terms of reference, and clients who normally have a say on how the evaluation is conceived and performed. Commissioners and clients do no operate not alone either. They play a role in organizations, which have a purpose, a way of working and sometimes strategic aims in their evaluation work. Evaluations may be conducted using established evaluation systems or procedures, for example, or ways of developing terms of reference, of finding and hiring evaluators. Thus, credibility not only depends on the quality and independence of the evaluators, but also on the institutions and the systems where the evaluations are conceived, planned and managed.

However, the evaluation community has a tendency to think – sometimes exclusively – from the evaluator's point of view. A typical example is the set of standards, principles and guidelines for ethical conduct, which some evaluation societies have adopted in the last two decades. Most, including the Joint Committee on Standards for Educational Evaluation, which established the first evaluation code, seem only to think from the evaluator's point of view. They recommend what the evaluators should do to conduct a good, useful, correct and ethical evaluation. Even the Norms for Evaluation in the United Nations System, which were established by UNEG in 2005 and have a United Nations system perspective in most sections (definition, responsibility of evaluation, policy, intentionality, impartiality, independence, evaluability, quality, competencies of the evaluation, transparency and consultation, evaluation ethics, follow-up, and contribution to knowledge building), entrusts the two norms in the section on quality of evaluation, and the five in evaluation ethics, in the evaluators' hands. The UK Evaluation Society's Guidelines for Good Practice in Evaluation[3] are exceptional in this regard, as they are also for commissioners, self-evaluation and for evaluation participants, as well as for evaluators.

In sum, we have thought more about what evaluators can do to enhance evaluation credibility, but less on how institutions and commissioners should promote credibility. What should institutions or governments do? I would suggest that they have a clear policy on evaluation, on its purpose, and who should benefit. For example, an enabling environment for stakeholder participation is more frequently facilitated – or obstructed – by organizations, clients and commissioners than by evaluators.

The instruments and mechanisms for implementing a clear evaluation policy, which ultimately enhances credibility, may vary greatly and will depend on the characteristics of the political system and general context and culture of each country, among other things. For some countries in Europe, the creation of a specific institution for evaluating public policies for all sectors might work (Spain). Elsewhere, the goal can be better pursued through advances in each policy sector (France). Some might find that passing concrete legislation that requires evaluation is particularly useful, while others realize that this mechanism might have unintended effects, such as excessive bureaucratization or simply non-implementation (Italy). Some evaluation systems reside in the Legislative Branch (Switzerland), while the Executive is responsible in some countries (Sweden). Each political and administrative system, each political situation, may require a different solution for promoting credibility.

Spain offers an illustrative example. The *Agencia Estatal de Evaluación de Políticas Públicas y Calidad de los Servicios- AEVAL* (National Agency for Evaluation of Public Policies and Quality of Services) was established in 2006, after a very well-considered project in which different expert commissions participated. In the early part of the project in 2004, experts called for an agency depending on the Legislative Branch for the agency to have better conditions for independence, and hence, credibility. That call was made with an understanding of the difficulties that its implementation would have in Spain's non-presidential political system, which is formed of closed electoral lists and demands to strong party discipline in the parliamentary system. Finally, because of the formal difficulties that call would require, the recommendation was to start the Agency at the Executive level and eventually changing it to depend on the Senate when a foreseen constitutional reform of the Senate would come. What we did not know at the time is that Spain was about to enter a period of major economic crisis, which was preceded by a political and institutional crisis that started with tremendous – even hysterical – confrontation between the two main parties, reflected especially in Parliament. Having AEVAL depend on the Legislative in this context would have had unintended consequences on credibility from administrators and the citizenship. From my perspective,

[3] See https://www.evaluation.org.uk/assets/UKES%20Guidelines%20for%20Good%20Practice%20January%202013.pdf

AEVAL had a greater chance of enhancing its credibility and independence from the Executive than if it had been in the Legislative.

Unfortunately, AEVAL was born at the onset of a deep economic crisis. Dramatic budget cuts and its gradual reduction as a political priority meant that it did not develop as had been planned. Regarding its credibility, AEVAL started to serve the Public Administration at the beginning of a major political crisis, which led to a substantial and ever deeper political disaffection by the Spanish public, accompanied by wide political and institutional distrust (the general dissatisfaction at the functioning of democracy has continued to increase from 42 percent in 2007 to 62 percent in 2011).[4] In my opinion, having reached this point, the only way that institutional and political trust will be recovered in Spain will be by giving real voice to the citizenship. And evaluation could play an important role in that recovery.

4. Evaluation capacity building for both evaluators and commissioners

It is clear that evaluation credibility depends to a large extent on the expertise of the evaluators – and on that of commissioners. That is why specific education on evaluation, good training courses and internships are so important for sound evaluation capacity building.

However, there are few specific graduate programmes in evaluation, whether in Europe[5], in North America or in other regions of the world. After many years dedicated to evaluator's training, I am convinced that taking one course in evaluation in a general research methods, management, public policy or development Masters or PhD, is frequently not enough to consider someone prepared to be good evaluator or commissioner. Neither are 20 or 40 hours professional training courses. A very important part of any good evaluation training should be in 'learning by doing', which is why it is crucial to provide internship and practicum experiences to new evaluators. But previous to this 'learning by doing', there is also a vast body of knowledge and evaluation theory, practices and dispositions that are part of an evaluation curriculum, and that can and should be taught.

In relation to capacity building, it is necessary to mention the efforts the evaluation community has made towards its professionalization. Those efforts include the elaboration of codes of practice. These are sets of norms, guides, and standards adopted by evaluation societies mainly during the 1990s and the early 2000s (see Bustelo, 2006). These codes started to establish some boundaries of what an evaluator should do as part of a good, useful and ethically correct evaluation (with the exception of the UK Evaluation Society guidelines mentioned above, which added guidelines for commissioners and others). Recent years have seen those codes develop into a focus on competencies or capabilities, that is, the capabilities

[4] Public trust in institutions dropped enormously between 2007 and 2011. Although we do not yet have the updated data, it has probably continued to drop in 2013, given the latest corruption cases and, in my opinion, the bad management of those cases by the Spanish Government. As examples, here are the average means in trust for 2007 and 2011 for different institutions: The King, from 7.20 to 5.79; The Constitutional Court form 6.08 to 5.00; the Congress of Deputies from 5.54 to 4.87; the Senate from 5.45 to 4.07; the Spanish Government from 5.46 to 3.93; the EU from 6.67 to 5.79; political parties from 4.22 to 3.38; and the unions form 4.58 to 3.26. Data from CIS (*Centro de Investigaciones Sociológicas* and from prof. Francisco Llera's presentation at the 6th Harvard Summer Seminar on Sociological and Political Research, August 2013.

5 See www.europeanevaluation.org/images/file/Study_programmes/16_profiles_November%202012.pdf

an evaluator should have and develop. The European Evaluation Society (EES) developed a capabilities framework and validated it through a survey to its membership.[6] This same stream has recently turned into a discussion about possible designation or accreditation of evaluators. Canada has already implemented an accreditation scheme and the International Development Evaluation Association (IDEAS) has initiated its own review. While the issues involved are controversial and sensitive, the European Evaluation Society believes that the time is right to engage in a public debate about the pros and cons of a potential peer review system geared to professional development and designation. Thanks to Evalpartners, which has provided funding for the European Evaluation Society and the UK Evaluation Society to sponsor a joint workshop designed to debate the potential for a Voluntary Evaluator Peer Review (VEPR) system, a discussion has been launched in the European Evaluation Society membership. The VEPR proposal has been designed under the aegis of the European Evaluation Society's Professionalization Thematic Working Group (TWG), and the Society's Board unanimously agreed that the TWG proposal deserves exposure to a broad evaluation audience as the first step in a proactive consultative process.

Why am I linking these professionalization efforts to the necessary evaluation capacity building as a way of enhancing credibility? Because these efforts have been, and will continue to be, an invaluable aid for building evaluation capacity by identifying what evaluators should know, the capabilities they should be trained in and, eventually, be tested against for designation or accreditation. Evaluation credibility should also be promoted with a system of 'accredited' evaluators but also commissioners. An evaluation's quality and credibility do not depend only on evaluators, so these capabilities should also be considered for commissioners and evaluation units elsewhere. Credibility also depends on political systems and cultures, institutions and contexts. Any framework or system that helps to define what a credible evaluation should be like must never be imposed, should be object of periodic review and of negotiation in the community and should generally be flexible enough to be useful in different contexts.

From my perspective, an indicator of a degree of consolidation in the evaluation function is the fact that evaluation commissioners and managers recognize that they also need to learn evaluation and look for some sort of evaluation training. One of the best features of an evaluation course is to be seated together exchanging and learning at least these two different roles and perspectives –evaluators and evaluation managers and commissioners- from each other. However, more thought has to be given to the possibility of training some different capabilities and skills for different roles in evaluation. That is one of the reasons why I think we should look at the Evaluation Capacity Building picture from different angles and perspectives. As an illustrative example, I have seen commissioners so enthusiastic in their evaluation work that they have misinterpreted their role and have told the evaluation team exactly how they should conduct the evaluation. But the job of a good commissioner is not about elaborating Terms of Reference (as I have unfortunately seen more than once) where too much time is spent on explaining the concrete methodology and methods to be used. We all know evaluation teams need a good explicit framework and context, but they also require enough freedom to be able to do a good job.

6 www.europeanevaluation.org/files/EES_Capabilities_Survey_Evaluation.pdf

5. Credibility and methodology: What counts as credible evidence and for whom?

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 get through third persons or secondary sources. Who have we got information from? Have we gathered information from the whole spectrum of stakeholders, or just from some of them? There is also the issue of honesty and impartiality, because credibility requires that evaluations should report successes as well as failures.

Credibility is inevitably related to the methodological perspective or how empirical evidence is gathered and analysed. 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. Moreover, 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, have fiercely reappeared when some international organizations, networks and federal departments in the USA have identified the randomized controlled trial (RCT) as the '"gold standard" design for generating 'scientific' evidence of evaluated programmes or policies. This created much discomfit in the evaluation community during the last decade, and generated responses from several evaluation societies, including from American Evaluation Association (AEA) in 2003 and the European Evaluation Society in 2007.[7] In 2006, Claremont University organized a symposium on "What Counts as Credible Evidence in Applied Research and Evaluation Practice?", in which known evaluation academics, both in the experimental-quantitative and the non-positivistic-constructivist and qualitative approaches were invited to participate. That symposium led to the production of an edited volume with the same title as the symposium (Donaldson, Christie & Mark, 2009), which was organized around the social inquiry paradigms as a frame for the debate on 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, and not exclusively related to a concrete epistemological stance, such as experimentalism.

[7] See EES statement on methodological diversity at www.europeanevaluation.org/library.htm

From my perspective, there can be no other way, because the discipline of evaluation has evolved from joint perspectives and multiple methods and approaches, allowing debate among evaluators who come from very different traditions of study. Due to its practical and applied nature, and the need for credible evidence from different perspectives to answer varied questions in different contexts, evaluation has been one of the first fields in which quantitative and qualitative researches and evaluators have exchanged views, networks and talked to each other. Moreover, the evaluation community was quickly ready to embrace the Mixed Methods approach. In this "era of paradigm pluralism" (Greene, 2013: 111) and necessary understanding among different perspectives about social inquiry, this Mixed Methods approach is broadly accepted in the evaluation community. As Donna Mertens and Sharlene Hesse-Biber say in the Editors' note in their recent volume, *Mixed methods and credibility of evidence in* evaluation*,* in New Directions for Evaluation:

> *An old Italian proverb reads, "What's old is new, what's new is old" (Melfi, 2011). This quote characterizes the story of mixed methods in the evaluation community in that mixed methods have been used by evaluators for many years. Many evaluators intuitively came to the conclusion that evaluations on complex social programs could be enhanced by the use of multiple methods; hence the combination of both quantitative and qualitative data in the same study is nothing new. Attention to mixed methods in evaluation was apparent in the New Directions for Evaluation (NDE) edited by Jennifer Greene and Valerie Caracelli in 1997 (Greene & Caracelli, 1997). Since that time, attention to mixed methods has increased exponentially, as evidenced by the launch of the Journal of Mixed Methods Research in 2007, which had an initial impact factor of 2.219 and ranked fifth out of 83 journals in the social sciences, interdisciplinary category, according to the 2010 Journal Citation Reports by Thomson Reuters (2011). The American Evaluation Association (AEA) Topical Interest Group (TIG): Mixed Methods in Evaluation was founded in 2010 and quickly became one of the largest of AEA's TIGs. And, the Sage Handbook of Mixed Methods in Social and Behavioral Research (Tashakkori & Teddlie, 2010) is in its second edition.* (Mertens & Hesse-Biber, 2013: 1)

This important piece of work aims to ask "about the connection between the call for evidence-based programmes and the potential contribution of mixed methods to the creation of credible evidence. The purpose of this issue is to examine the contributions of mixed methods evaluation and its emerging philosophies, theories, and practices (…) as well as opening up the possibility of enhancing credibility with evaluations that start from several paradigmatic stances, such as postpositivism, pragmatism, constructivism, and transformativism" (Mertens & Hesse-Biber, 2013: 3).

The Mixed Methods approach claims that this mixed usage should not only function at the methods level, but also at the methodology and epistemology level. As Mertens and Hesse-Biber say, "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 volume puts forward the need to be conscious and explicit about such theoretical perspectives and assumptions. Jennifer Greene argues that mixed methods evaluators should be "explicit about the paradigmatic assumptions that frame and guide 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 positivism stances, so they should be explored and defined by other paradigmatic perspectives. Positivism has been the dominant paradigm for many years, but 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. So that paradigmatic and methodological transparency is needed for credibility. In my opinion, we should insist that this transparency is exercised not only by Mixed Methods evaluators, but by all evaluators. For the sake of credibility, I believe that 'classical' and dominant understandings, such as experimentalism, should not be taken for granted, and the paradigmatic and methodological choices that drive those perspectives should be explained and made explicit. This would be a real acknowledgment that there are other modes of inquiry that are not hierarchically inferior. In this way, methodological pluralism would become real – no longer will one perspective be the 'norm', while other 'alternative' choices have to be justified.

6. Communication and reporting: another key issue for evaluation credibility

As a final important point in this discussion on evaluation credibility, I would like to mention communication and reporting. Although the lack of time and space in this guiding paper does not allow me to develop further this point, I would like to mention at least three aspects or questions for discussion in the online community of practice. These are the issue of transparency and what should be made explicit in the report; a need for a fluid communication with stakeholders during the process; and the accessibility of evaluation reports.

If an evaluation is to be credible, it must be transparent and state the political and institutional contexts in which it took place. It should also consider – and include in the evaluation report – paradigmatic, methodological and values-related issues.

Fluid communication with commissioners and stakeholders during the evaluation process is known to promote better use of evaluation results (see, for example, Torres, Preskilll & Piontek, 1996). But fluid communication is also important in enhancing the credibility of the evaluation process and the evidence resulting from it. If commissioners and stakeholders have regular and transparent feedback during the evaluation process, this results in a higher probability of perceived credibility by them.

A key issue for transparency and credibility is related to the accessibility of evaluation reports. The evaluation process may have been well designed, the evaluation may have been conducted honestly and transparently, but if the final report is not accessible to the public it will be almost impossible to be perceived as credible by those without access to it. The ease of access to information through the internet has vastly improved openness, but it also challenges the cases in which there is no immediate access to evaluation reports.

References

Bustelo, M. 2006. "The Potential Role of Standards and Guidelines in the development of an Evaluation Culture in Spain" *Evaluation*, Vol. 12(4) pp.436-452.

Chelimsky, E. 1998. “The role of experience in formulating theories of evaluation practices”. *American Journal of Evaluation*, 19(1): 35-55

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.

House, E. 2005. “Deliberative Democratic Evaluation”. In S. Mathison (Ed), *Encyclopedia of evaluation.* Thousand Oaks, CA: Sage Publications.

House, E. & Howe, K. 2000. “Deliberative Democratic Evaluation in Practice”. In George F. Madaus, Michael S. Scriven, Daniel L. Stufflebeam (eds.), *Evaluation models : viewpoints on educational and human services evaluation*. Boston : Kluwer-Nijhoff.

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.

Monnier, E. 1995. *Evaluación de la acción de los poderes públicos*. Madrid, Instituto de Estudios Fiscales. Ministerio de Economía y Hacienda.

Torres, R., H. Preskill, and M. Piontek. 1996. *Evaluation Strategies for Communicating and Reporting*. Thousand Oaks: Sage.