

EVALUATION, VALUATION, NEGOTIATION: SOME REFLECTIONS TOWARDS A CULTURE OF EVALUATION

Pablo Rodríguez-Bilella
CONICET / IOCE. Argentina

Rafael Monterde-Díaz
Universidad Politécnica de Valencia. / SEE. España

Abstract: Although the evaluation of public policies is a subject of growing interest in Latin America, there are problems with the design and implementation of evaluations, as well as with the limited use of their results. In many cases, the evaluations have more to do with generating descriptions and less with assessing these activities and using those assessments to improve planning and decision making. These points are explored in a case study of the evaluation of a rural development program in Argentina, emphasizing the process of negotiation and consensus building between the evaluators and the official in charge of approving the evaluation report. The lessons learned from the experience point to the generation and consolidation of a culture of evaluation in the region.

Résumé : Bien que l'évaluation des politiques publiques fasse l'objet d'un intérêt croissant en Amérique latine, il existe des problèmes avec la conception et la mise en œuvre des évaluations, ainsi qu'avec l'utilisation limitée de leurs résultats. Dans de nombreux cas, les évaluations ont plus à voir avec la génération de descriptions qu'avec l'évaluation même de ces activités et son utilisation afin d'améliorer la planification et la prise de décision. Ces points sont abordés à partir d'une étude de cas sur l'évaluation d'un programme de développement rural en Argentine. L'accent est mis sur le processus de négociation et de consensus entre les évaluateurs et le responsable de l'approbation du rapport d'évaluation. Les leçons tirées de l'expérience préconise la génération et la consolidation d'une culture de l'évaluation dans la région.

INTRODUCTION

█ The evaluation of public policies has become a topic of growing interest in multiple contexts, particularly in Latin America. Managers of public institutions and policy makers have begun to use evaluation both to streamline public spending and to comply with accountability issues.

Problems are shown in the design and implementation of evaluations, as well as in the limited use of their results. Assessments are often used as a form of financial and administrative control, rather than to provide services to planners and administrators. In many cases, they have more to do with generating descriptions than with assessing these activities and using those assessments to improve planning and decision making. This reflects the existence of an incipient and weak evaluation culture in the region.

In order to examine some dimensions of these issues, this article presents the case of the evaluation of a rural development program in Argentina. The narrative tries to account for the process of negotiation and consensus between the evaluators and the official in charge of approving the evaluation report. On the basis of the case, reflection is provided on the value judgements in the evaluation and the negotiation layout built into the interaction. The lessons learned from this experience point to the consolidation of an evaluation culture of the region.

REPORTING ON THE EVALUATION: THE CASE OF THE AGUAS TURBIAS REGION

In the early 1990s, in the context of the implementation of structural adjustment policies, the first national program that focused on small farmers was designed in Argentina. The program provided micro-credit, technical assistance, and training to small farmers' groups. Years after its implementation, the program was merged with another one largely financed by the World Bank, with the same operational model but using grants instead of credit (subsidies designed for predefined purposes.) After 5 years of operation, and before entering a second phase, an evaluation of the program was carried out in different regions of the country where it had been implemented.

Evaluators working in pairs (one a specialist in social issues and another one in agriculture) were deployed for each region where the

program was running. The evaluation report from each region was submitted for approval to the supervisor of the program evaluation, who had hired the evaluators. The terms of reference showed an emphasis on analyzing the use of the subsidy received in the dynamics of job creation and entrepreneurship associations, the role played by technical advisors from each project, the relevance of training instances, and the interaction with other government agencies.

Having read the report from the Aguas Turbias region, the supervisor made a series of comments to the evaluation team. First, he expressed surprise at the tone of the analytical section of the report, because the evaluators seemed “very angry” about the results they had seen in the region. Although the supervisor recognized some flaws in the local coordination of the program, he noted that they could be explained by the particular context of the region, which he said he knew very well. While he advised removing “everything that consists [of] opinion or inferences, leaving findings and taking comparisons out of the report,” he highlighted that all the negative points found in Aguas Turbias were also present in other regions—in many cases at a more critical level. With this opinion he did not aim to eliminate these points of the report, but suggested a “less passionate” way of stating them. Finally, he made the following clarification:

Remember that these reports, after being approved by the national coordination of the program, can be read by different people (from the World Bank, Ministry officials, other consultants), and we send them to the evaluated regions, where they can also be read by different people. There is not a culture of monitoring and evaluation as a way to improve performance. In general, especially by the authorities, monitoring and evaluation are considered as a “trial” to punish; that is why I am asking for these changes.

The evaluators took into account several of the requested recommendations and produced a second report. While they were aware of the style issues, they also tried to preserve what they considered the main findings and their assessment, as well as the recommendations.

After reading the second assessment report, the supervisor pointed out that it held an “evaluation position,” whereas the text should only show findings and make recommendations. He also questioned the evaluators for addressing particular situations in the Aguas Turbias region by judging them against the operational manual of

the program, as a means of indicating how far the practices fell short of what they should have been. He then suggested a further review of the report.

The evaluators wrote a third version that they sent to the supervisor, hoping it would be the final one. The evaluators had agreed not to accept further suggestions for amendment; if necessary, they would ask for an interview with the highest authority of the program to discuss the situation.

After reading the third version of the report, the supervisor made new comments, saying that he still saw some kind of “touchy-feely stuff” that had not appeared in the evaluations of other regions. He suspected that the evaluators had clashed with the working and communicating style of the local coordinator, and he insisted that the assessments should reflect facts without adjectives. The supervisor also showed his concern about “not hurting” those in the “front lines” (technicians, local officials) with some of the assertions, because that would not help to change the analyzed reality. He suggested that the changes proposed by him be accepted, in order to end the story.

The evaluators accepted the fourth version of the report as the final one, recognizing that their fundamental ideas were present. However, they felt that their argument had lost much of its original strength because of the removal of the operational manual’s benchmarks and the discussion of the program implementation in other regions. While several of their initial recommendations were in the final evaluation report, those considered “potentially hurtful” by the supervisor were deleted.

DISCUSSION

The key point of disagreement between the evaluators and the supervisor revolved around the “value” they gave to the program results in Aguas Turbias. What the discussion brought to the table was whether the role of the evaluators should include an assessment of the findings by making judgements about what they considered good (valuable) or less good (not valuable) about the intervention.

A semantic distinction is mandatory at this point: “to evaluate” means to value or to judge the merit of a given object (Scriven, 1994). It involves stating the value or quality of something, from some point

of view based on acceptable research procedures, and establishing whether it is adequate, appropriate, desirable, or proper for the intended purpose.

The main purpose of a program evaluation is to determine the program's quality through the formulation of a judgement about it. For Scriven (1990), the primary function of evaluation is the production of legitimate and justified judgements as a basis for relevant recommendations. A judgement is legitimated if it is formed by comparing the findings and their interpretation to one or more performance standards. It will be justified when it is linked to evidence and it is consistent with the values and standards agreed with the stakeholders.

Therefore, the valuation must be done in comparison with some kind of standard. The evaluators based their assessment of the Aguas Turbias case on the benchmarks established by the operational manual. In other words, they understood its standards to be the theory of the program. The evaluators agreed with the theory of change present in the program, which they had witnessed in other regions before, arguing that the program was relevant to the intended goals.¹ They tried to make this point clear in the successive versions of their report, by trying to weigh their judgements based on the desirable values of the program theory. Their dilemma was how to point out the negatives in the implementation of the program in Aguas Turbias, although they actually believed the design was appropriate.

Meanwhile, the supervisor did not accept that assessment of various critical and negative items, and focused his interest on finding particular facts. However, he had no problem in accepting the positive assessments regarding certain dimensions of the program. Consequently, his interest was not centred in the methodological excellence but in a political factor: avoiding strong critical judgements. A further expression of this related to the use of certain terms and the general tone of the report, which the supervisor considered needlessly emotional because they would potentially help to raise defenses of the program stakeholders as well as other outsiders. In some sense, this was a legitimate concern, as expressed in his reference to the limited evaluation culture, where the distance between improvement decisions and "punishment" decisions was narrow or simply nonexistent.

The pressure on the evaluators at this point was to convey a clear message in certain critical items, where softening the language

would be counterproductive. This fact highlighted the lack of semantic standards for evaluation, where the language does not have only an instrumental function, and the choice of terms is not void of interpretative weight or the display of the evaluators' standpoint.

Given that the assessment must be conducted in reference to some standard, the question is who should assign value in the assessment. While the original contributions from Scriven stressed the role of the evaluator, others (such as Stake, House, Eisner, Lincoln & Guba, etc.) expanded the possibilities of value assignment to a plurality of stakeholders, usually by placing the evaluator in a facilitator role.

The Aguas Turbias case is a good example of how some clients think that evaluators should only explore the extent to which program objectives are logically linked to certain products, while value judgements are held as a prerogative of those who designed the program and, mainly, the officials or authorities that demanded the evaluation.² In other cases evaluators tend to make value judgements based on their own expertise, which is precisely one of the main reasons why they are hired. Sometimes this means they go beyond their terms of reference, because focusing only on the targets and goals of the program may mean missing the whole picture of what the program should be doing and the eventual discovery of the unintended effects of the intervention.

In the present case, accepting the final version of the report meant sacrificing the integrity of the evaluation and therefore its quality. This is shown in that the supervisor's objections to the judgements of the evaluators were limited to those he disagreed with. On the other hand, the evaluators had a very narrow margin for negotiation in order to express their view as well as assess how they thought the program was running in the region. They were not able to present a strong argument that let the stakeholders achieve certain consensus, or at least allow greater plurality in the final expression of the report. Fears of the effects of a highly critical assessment prevailed, which were expressed—from the evaluators' perspective—in a sweetened final report.

The evaluation of the program in Aguas Turbias showed an important quality flaw, especially if we understand it as a measure of the usefulness and relevance of findings and recommendations. In this sense, the lack of a clear understanding of the role of each actor in the process resulted in an assessment report that reached its planned objectives, but did not cover the key dimensions of an evaluation as

identified by Patton (1997): be useful for the improvement and the learning process, both of the program and the institutions involved.

CONCLUSIONS

The evaluation should be directed toward consistency between its purpose and the activities put into practice during the process, as well as the methodological tools employed. Starting from the values of the intervention, evaluators should provide logical and based-on-facts statements in order to make recommendations. Furthermore, the political nature of the evaluation is also evident in the choice of the central actors for value assignment.

A negotiation pre-evaluation could then be converted into an appropriate place for reviewers to show (to a feasible extent) the values and attitudes they have as both individuals and professionals. These values can be compared to those that the primary clients expect will determine the success of the intervention. This pre-assessment analysis will undoubtedly receive adjustments during the evaluation, but its early inclusion can free the field of (at least some) potentially critical points. It could also enable the evaluation to generate an atmosphere of learning and transformation, which can become key elements for the emergence and consolidation of an evaluation culture.

This culture of evaluation also involves a process of capacity development, which facilitates different stakeholders to observe and understand events, actions, objects, or situations related to the assessment. When this culture is weak or nonexistent, policy and institutional constraints are more difficult to overcome, the fight for adequate budgets demands additional efforts, the evaluator is closer to the image of a detective or a spy rather than a critical friend, and the assessment itself will have more difficulty reflecting regional and local priorities (Haarich, 2005).

The role of the evaluators is key in deciding the evaluation approach most appropriate in each case. In this sense, the fact that the evaluators make a particular interpretation of the findings and build their recommendations from there is as valid as the possibility that they facilitate a negotiation process in order to allocate the weight of the evaluative judgements (without in any way excluding reference to standards of evaluation). In between, there is a whole gradient of valid choices for addressing this reality, which contains an important part of social construction and not an immutable fact a priori (in the

Kantian sense; Pawson & Tilley, 1997). The lack of previous agreement on the role of the evaluators is where the source of disputes over the validity of the judgements most likely lies, rather than on the usual explanations about the professionalism of the evaluators, the reliability of certain tools, and so on. In the end, it is all about the proper management of power relations between actors throughout the evaluation process.

The evaluation report of Aguas Turbias did not result in any change in the way the program worked in the region, as the evaluators guessed. However, when the national authorities in charge of the program were renewed a couple of years later, the report was used as an important input to lobby for the replacement of local program managers in Aguas Turbias, but not for the elimination or reduction of the budget (as the supervisor feared). In this sense, evaluation is potentially able to construct new realities in the evaluation process (Guba & Lincoln, 1989), assuming that

[t]he truth is not something “out there” waiting to be documented but is, rather, a story to be written by those performing the assessment. It is a story which is gradually constructed by individuals trying to make sense out of complex realities, and who cannot be separated from the context within which they perform the assessment. (Phillips & Edwards, 2000, p. 47)

From this point of view, the value assignment, which is definitely a key element of the evaluation process, could also be considered as a transactional and/or negotiable issue, and it would hardly fall to the evaluator to carry all the weight on it. This value assignment is not an odd element of the political and controversial nature of the evaluation process in its double interpretative/descriptive and prescriptive function, as a conditioning for the future.

NOTES

- 1 If this had not been the case, an additional factor would have added complexity. That is, if the program theory had been considered partially or completely invalid, what could they consider as a reference for value? Asking this critical question is beyond the scope of this article, but the authors hope to be able to develop it in a future paper.

- 2 In some critical cases, the professional ethics of the evaluators could force them to overcome the required absence of a “valuing” position. This is true when considering, for example, that their primary clients are pursuing their own interests and their own personal agenda, or acting in opposition to honest and transparent practices.

REFERENCES

- Guba, E. G., & Lincoln, Y. S. (1989). Judging the quality of fourth generation evaluation. In E. G. Guba & Y. S. Lincoln (Eds.), *Fourth generation evaluation* (pp. 228–251). Newbury Park, CA: Sage.
- Haarich, S. N. (2005). Diferentes sistemas de evaluación de las políticas públicas en Europa: España, Alemania y los países del Este. *Revista Española de Control Externo*, 7(20), 61–87.
- Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text*. Thousand Oaks, CA: Sage.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London, UK: Sage.
- Phillips, S., & Edwards, R. (2000). Development, impact assessment and the praise culture. *Critique of Anthropology*, 20(1), 47–66.
- Scriven, M. (1990). The evaluation of hardware and software. *Studies in Educational Evaluation*, 16, 3–40.
- Scriven, M. (1994). Product evaluation: The state of the art. *Evaluation Practice*, 15(1), 45–62.

Pablo Rodriguez-Bilella is a consulting sociologist, with more than 15 years of progressively more responsible positions in applied and academic social sciences: evaluation, planning, community participation, and program management. He is committed to practical approaches that emphasize sustainable development over charity, community participation over relief, and empowerment over dependency, yet realistic enough to understand the need for pragmatism. He is a board member of the ReLAC (the Latin-American Network of Evaluation) and IOCE (the International Organization for the Cooperation in Evaluation). He is a lecturer of Social Anthropology

at the Universidad Nacional de San Juan, and researcher at the CONICET, both in Argentina. He usually blogs about development, evaluation, and complexity at albordedelcaos.com.

Rafael Monterde-Díaz is a project management and evaluation specialist, with extensive work in consultancy, training, and evaluation capacity development in Spain and Latin America countries. He is an Associate Professor at Universidad Politècnica de València, where he teaches Project Design and Management in undergraduate courses and Program Evaluation in Development Studies at post-graduate level. He is the author of EVALUATECA (<http://evaluateca.wordpress.com>), the first blog in Spanish devoted exclusively to evaluation. He serves as a board member of the Spanish Evaluation Society for external relations with evaluation societies and networks worldwide. He holds a B.Sc./M.Sc. in Industrial Engineering and Ph.D. in Project Management and Innovation, with complementary postgraduate training in Data Analysis.