

## **PROGRAM EVALUATION IN THE CONTEXT OF DEBATES IN THE FIELD: THE EVALUATION OF PR-CETP**

---

M. BRAVO

*University of Puerto Rico, Rio Piedras*  
mlbravo@caribe.net

### **Abstract**

The purpose of this paper is to present some controversies in the field of evaluation that shape the way evaluation is currently carried out. Since evaluation uses methods developed in the social sciences, some of the controversies come from debates in this field (i.e., the quantitative vs. qualitative debate), but others derive from the nature of program evaluation as applied research (i.e., external vs. internal evaluations; independent vs. collaborative evaluations; role of evaluator regarding use of evaluative results). The article examines these controversies and proposes a multi-method or mixed methods approach as the most appropriate way to carry out program evaluations. It also describes the positions adopted in the evaluation of the Puerto Rico Collaborative for Excellence in Teacher Preparation (PR-CETP) project concerning other controversies in the field. Understanding these controversies might help math and science faculty to become more knowledgeable consumers of evaluation results, as well as more informed collaborators in evaluations of programs in their fields.

### **Introduction**

The evaluation of a program is a systematic study of its characteristics and merits. Evaluative research involves the systematic application of the methods of the social sciences to study programs [1]. Consequently, the currents and controversies present in the social sciences influence the evaluation field. One of the important debates in this field is the use of qualitative (data expressed in words or narrations) or quantitative (data expressed in numbers) methods in research in these disciplines. This debate is fundamental because it is the basis of the research activity and underlies other important issues. Additional controversies arise from the nature of program evaluation as applied research that is highly influenced by the social, political, and economic context in which it is carried out [2]. A debate regarding the working conditions of evaluators—that is whether evaluation should be carried out by people external to the program or by those that work for the program (external vs. internal evaluators)—has arisen. Another one concerns whether the responsibility for designing and carrying out the evaluation should fall on an independent person/entity or, if it should be a collaborative endeavor. The use of evaluative results has also been a matter of controversy; specifically, whether evaluators should have the responsibility for promoting the use of evaluation results and what kind of use should be sought. Understanding these controversies might help math and science faculty to become more knowledgeable consumers of evaluation results, as well as more informed collaborators in

evaluations of programs in their fields. The present article examines the above mentioned controversies and proposes a multi-methods or mixed methods approach as the most appropriate way to carry out program evaluations. It also describes the positions adopted in the evaluation of the Puerto Rico Collaborative for Excellence in Teacher Preparation (PR-CETP) project concerning other controversies in the field.

### **Debate Concerning Qualitative vs. Quantitative Methods**

One of the most important controversies in the realm of research in education and the social sciences, as well as in program evaluation, is the use of qualitative or quantitative methods. The strengths and weaknesses of these types of methods have been thoroughly debated [3-4]. Those that criticize the quantitative methods denounce the dehumanizing tendencies of numeric representations, claiming that a better understanding of causal processes could be obtained through the intimate knowledge of people and the resultant qualitative observations [4-6]. They adduce that it allows an understanding of the concrete manifestations of a program that produces valid knowledge about its effects. On the other hand, those that advocate the use of quantitative methods reply that qualitative data are very expensive if they are used extensively, they are highly subject to bad interpretations, and they usually contain information that is not uniformly gathered in all cases and situations. In the case of impact evaluations especially, it has been pointed out that, to obtain results scientifically acceptable which yield relatively precise estimates of the effects of a program, quantifiable information that is systematic and uniformly gathered is required [1].

### **Epistemological Pluralism**

This controversy arises from a clash of epistemological paradigms [3]. Researchers and scholars differ about the respective merits of the two approaches due to their different views about the nature of knowledge and how knowledge is acquired [7]. Qualitative researchers argue that there is no objective social reality, that all knowledge is “constructed” by observers who are the product of the cultural, social, and political environment in which they operate. On the other hand, while quantitative researchers no longer believe that their methods yield absolute and objective truth, they adhere to the scientific model and seek to develop sophisticated techniques and statistical tools to study social phenomena.

Until recent times, the positivistic and empiricist approach prevailed in science. This conception arises from the mechanic model based on the works of Galileo and Newton [8]. The clock was used as a model of the cosmos, and eventually was taken as the cosmos itself. The

world was conceived as a great machine whose operation could be discovered by breaking it down into their constituent parts: take the whole, divide it into parts, analyze it, and put everything together again. This conception involves a vision of the human reason as autonomous, as bequeathed to us by Descartes [9]. It establishes a separation between the knowing subject and the object of knowledge; it assumes the existence of a stable material reality that can be totally captured by the human mind.

However, the physics of Einstein and Plank opposes the mechanical conception of the world. On the microphysics level, constituent elements of that great machine are not found but, instead, relationships exist between abstract entities [8]. The essence of nature is not objects, but interconnections; matter is composed of interconnections. No intrinsic properties can be identified in these abstract entities; rather, they depend in good measure on the theoretical and methodological models used to study them. The resulting observation, consequently, is as much a function of the phenomenon itself as of the theoretical and methodological schemas used to study it [10]. The image of the universe that arises is that of a dynamic, indivisible whole, whose parts are essentially interconnected and could only be understood as patterns of a cosmic process.

This vision has important repercussions for research in the human sciences, and therefore, in program evaluation. It indicates that knowledge is not the direct result of experience, but rather is, in good measure, a human construction. It leads researchers to conclude that there is not a single theoretical perspective nor a single method that will lead us to approach human phenomena with more certainty [10]. One cannot talk about a single scientific method that is appropriate to capture reality as it is, since this is a futile objective, impossible to achieve. It is only possible to capture deformed approximations to that reality, that result from the interaction between the phenomenon and the theoretical and methodological models used to study it; hence, the importance of explicitly describing them. This conception entails an epistemological pluralism in which it is considered that the comparison and interaction of diverse systems of research permits a better understanding of the phenomena under study than is possible to achieve with a single system. A better understanding could arise from the syncretic interaction of several positions [10]. Since knowledge is not the direct result of experience, but mostly a human construction, the comparison of several constructions could lead to a deeper and more certain understanding of a studied phenomenon. Using more than one method can strengthen the validity of research results, an approach called triangulation [7].

### **Mixed Methods in Program Evaluation**

The field of research and evaluation in education, and other human sciences, has been moving toward the acceptance of this epistemological pluralism. Shadish, Cook and Leviton point out that no paradigm for the construction of knowledge is better than another since all the approaches are plagued with considerable difficulties [2]. A better theory of program evaluation is one that assumes this position, according to these authors. The use of both types of methods could strengthen the validity of results produced than by either one by itself [11]. Frechtling and Sharp have pointed out: "Experienced evaluators have found that most often the best results are achieved through the use of mixed method evaluations." [7] It is increasingly recognized that all data collection, qualitative as well as quantitative, operates within a socio-cultural context and is affected to some extent by the perceptions and beliefs of investigators and data collectors. This pluralistic position establishes that each approach has its utility and that all methods are not equally good for all purposes, so it is necessary to identify the strengths and weaknesses of the same for diverse situations [2].

The selection of methods depends on the purposes of the evaluation and the aspect of the program in focus, as well as the specific situation in which it is carried out. For example, Rossi, Freeman, and Leviton consider that qualitative methods could play a preponderant role in the evaluation of the conceptualization and design of programs, as well as in monitoring its implementation [1]. In contrast, quantitative approaches are considered to be more appropriate for the evaluation of their impact and efficiency. On the other hand, Fink considers that qualitative methods are useful for programs whose objectives are still in process of being defined and to explore the viability of the quantitative methods to be used in an evaluation [12]. Another use suggested by this author, is in cases in which there were no valid and reliable measures available to evaluate the impact of a program. In these situations, information on the processes is gathered as well as the participants' perceptions concerning the program impact. Qualitative methods could also be used to supplement more common quantitative data (e.g., surveys, standardized tests); because they are "personalized," they could add emotion and tone to purely statistical results [12]. This particular combination of methods provides concrete examples of the life of the studied people, many times in their own words, deepening and giving emphasis to the human experience suggested by the numeric discoveries [13].

The selection of the research approach to use in an evaluation is not only an epistemological, but also a strategic matter. Evaluations are carried out usually with the purpose of accountability, to promote program improvement, or contribute to the formulation of public

policy. These activities have a strong political dimension. Upon making decisions concerning methods, therefore, one should consider the type of information that the program stakeholders and audiences would accept as valid knowledge. A position that reflects this point emerges from the expression of Chelimsky indicating that it is rarely wise to enter a burning political debate armed only with a case study [14].

### **Mixed Methods in the Evaluation of PR-CETP**

The evaluation carried out in the PR-CETP project is an example of the use of the alluded mixed methods approach. As it is described in other articles in this special volume, the Puerto Rico Collaborative for Excellence in Teacher Preparation (PR-CETP) has as its basic goal the transformation of the main programs of science and mathematics teacher preparation in Puerto Rico. This transformation is geared to achieve that future teachers provide an education of excellence to all K-12 students. Seven higher education institutions, four public and three private, make up the collaborative. The project seeks to make changes in four elements of the programs, Curricula, Professional Development, Student Support, and Institutional Policies, through four interinstitutional components that coordinate the reform. The evaluation of the project is focused on the project goals and involves the use of diverse qualitative and quantitative methods to collect information with formative and summative purposes. Next, we present some specific examples that illustrate the sole or combined use of these methods.

### **Quantitative Methods**

The quantitative methods used in the evaluation have allowed us to obtain abundant information from a great number of participants from different institutions which, since it is quantifiable, can be easily coded and analyzed in a statistical context. The strategy used to examine the achievement of one of their basic goals, the development of future teachers' conceptual understanding of subjects that they will teach, could serve to illustrate the utility of the quantitative methods to evaluate the impact of an intervention. The quantitative strategy used is a set of tests of conceptual understanding developed to evaluate the understanding of the main science and math concepts that are covered in university gateway courses: pre-calculus, general biology, general chemistry, and general physics. A test was also developed to evaluate the understanding of basic concepts about human development, concepts on which educational methods are based and, are therefore pertinent for all future teachers. These tests were constructed by a team of professors of different disciplines, based on the content and professional standard that guide the education of K-12 students. They were subjected to a rigorous process of item analysis, modifications based on these results, and testing of their psychometric properties. The

tests were used to answer the following evaluation question: To what extent are the future teachers moving toward the goal of achieving conceptual understanding of the subject matter they will teach?

These tests are being used in two main ways. First, they are administered as pre- and post-tests in courses that are reformed according to the constructivist educational philosophy that sustains the transformations of PR-CETP. When comparable groups exist, these tests are also administered in order to contrast results obtained from reformed courses versus non-reformed courses. In this way, the value added in conceptual understanding by the reformed courses could be evaluated. Second, tests are administered to graduating future teachers of the corresponding disciplines in order to know the level of understanding achieved in basic concepts of the subject that they will teach. The use of the quantitative technique employed, given its ease of administration and analysis, has allowed its systematic use in numerous groups of students. These large numbers allow comparisons and generalizations about the targeted future teachers.

### **Qualitative Methods**

Qualitative methods are used in the PR-CETP evaluation to understand the experience of participants from their own frame of reference in the context of individual institutions. It allows an understanding of the concrete manifestations of the project that produces knowledge about its implementation and effects. For this purpose, information is gathered using varied qualitative methods.

Some of these methods are focus groups and interviews. Two focus groups are annually carried out in each institution, one of students and another of professors. These have allowed capturing the perceptions that they have about the implementation and achievements of the project. In the past year, a series of two interviews per institution were also begun, one with a dean or director of a science or math program, and another with a corresponding administrator of an education program. These techniques are basically used to answer the following evaluation questions: To what extent is the project moving toward its stated goals, according to participants' perceptions? Which activities or strategies are aiding the participants to move toward the project's anticipated goals? What barriers are encountered and how are they being overcome? In order to overcome the limitation that information might not be uniformly gathered in all situations, guidelines for collecting and analyzing the resultant qualitative data have been developed. The obtained results are used to adjust the activities of the project to the needs and

preferences of participants and to the advancement of the project in the achievement of its goals and objectives.

### **Combination of Methods**

Another goal of the project is the institutionalization of the practices and transformations carried out in the teacher preparation programs. One of the practices to institutionalize is the collaboration between the faculty of sciences/mathematics and the faculty of education in the preparation of future science and math teachers. To study the achievement of this goal, we have combined qualitative and quantitative methods. On one hand, a survey is carried out annually in which the faculty is requested to judge the level of the existent collaboration on a scale that measures ten levels of collaboration, from “none,” all the way up to “considerable” collaboration. By means of open questions, they are also asked to describe the collaboration in their institution, and to provide suggestions on how to improve it. In the professors’ focus groups and in the deans/directors’ interviews, they are also asked about their perception of existent collaboration, and the role that PR-CETP has played to promote it. This combination of methods provides concrete examples of the situations studied in the participants’ own words, deepening and giving emphasis to the human experience suggested by the numeric results. Upon examining results coming from diverse sources and methods, by means of a process of triangulation, many instances in which the qualitative and quantitative results converge or supplement each other have been identified. Even in cases in which they diverge, both types of information have allowed a better understanding of the studied matter.

### **Debates that Arise from the Nature of Program Evaluation**

Besides the epistemological debate, other debates have emerged in the evaluation field that are more directly related to the basic nature of program evaluation. Evaluation is an applied research that is carried out in a social, political, and economic context.

### **Evaluators’ Working Arrangements**

There exist several controversies in the field that focus on the people that carry out evaluations. One concerns their working arrangements, what has been called the insider-outsider debate [1]. That is, if the evaluators are external to the evaluated program or if they work in the program and, therefore, are internal evaluators [15]. One position is that evaluators could work better when their positions are as secure and independent as possible of the influences of the staff and management of the project. It is adduced that external evaluators could exercise more objectivity, have less conflict of interest with the internal stakeholders of the program, and could

be more attuned to the needs of external stakeholders, especially those providing its funding. The contrary position posits that the frequent contact with the personnel of the program and those that make decisions improves the evaluators' work. They can obtain a better knowledge of the objectives and activities of the program than is possible to achieve by external evaluators. They also inspire more trust so they could get more truthful information, and could be in better harmony with the necessities of internal stakeholders.

The most prevalent position at the present time in the field maintains that there are few reasons to categorically prefer external or internal evaluations [1]. The existent evidence is far from pointing out with clarity if the internal or external evaluations are of better technical quality. Moreover, the technical quality is not the only issue to consider, since the utility of the evaluation could be of similar importance. What is crucial is that evaluators have a clear understanding of their role in specific situations.

The evaluation of the PR-CETP is carried out by a committee called the "Evaluation Component" headed by a central level coordinator, an assistant, and evaluators from each one of the participant institutions. The basic design of the evaluation was developed when the proposal of the project was submitted to the funding agency. However, it has been collaboratively refined and implemented by this committee. The carried out evaluation combines elements of both internal and external evaluations, although the former are more prominent. On one hand, the institutional evaluators are professors at the participant institutions (internal element), but they were recruited expressly to carry out evaluation functions in the project (external element). That is, although they generally are members of the faculty of each participant institution, they began working in the project when they were recruited to be evaluators. Additionally, the evaluation coordinator has an office in the facilities of the central level of the project and answers to the coordinating committee at this level (internal element). However, she was also specifically hired to carry out the evaluation (external element). Another external element of the evaluation is that one of the members of the National Visiting Committee of the funding agency has served as external consultant of the evaluation committee; strengthening the qualitative component of the evaluation has been one of her contributions.

The frequent contact with the personnel of the project and those that make decisions have improved the evaluators' influence in the planning and implementation of the project's work; thus, the use of the evaluation results has been enhanced because evaluators are in good harmony with the necessities of internal stakeholders. We also consider that the technical quality of the



evaluation has not suffered much because of measures taken to systematize evaluation procedures in the different institutions. Monthly meetings are carried out to plan and coordinate activities; and, an Evaluation Manual with detailed guidelines to systematize all the data collection and analysis processes has been developed. Maintaining technical quality of the evaluation enhances accountability to external stakeholders. Moreover, one of the tasks of the evaluators' coordination has been to promote evaluators' understanding of their role in specific situations.

### **Responsibility for the Evaluation**

Another controversy in the field involves whether an independent person or entity should do the evaluation or, on the contrary, if it should be done collaboratively. On the one hand, it has been pointed out that the evaluations designed and supervised in an independent fashion have the advantage that more control can be exerted on the evaluation activities and thus research processes could be more systematically carried out. On the other hand, it is adduced that collaborative evaluations can do more with less resources and, since this involves a larger number of people in the evaluation, a greater commitment to the use of the results can be achieved [6].

The evaluation of PR-CETP is collaborative. As previously stated, it is carried out by a team of evaluators with representation from all participant institutions. Since the project involves seven different institutions that vary in size, complexity, and type of teacher preparation programs, representation from each institution enhances the pertinence of the evaluation for each institution. This is one of the main lessons learned from using this arrangement. Moreover, since evaluators are professors from each campus, they have the standing and contacts to facilitate data collection and, their relationship with the project institutional coordinators enhances the use of results, as previously stated. In order to overcome a limitation of collaborative evaluations, evaluation processes in the different institutions have been systematized. Monthly meetings are carried out to plan and coordinate activities; an Evaluators' Manual with detailed guidelines to systematize all the data collection and analysis processes has been developed. Although this systematization has enhanced the evaluation's technical quality, a limitation has also been identified. Since the collaborative arrangement implies less control on research processes and personnel at the central level, it has caused data gathering procedures to be slower than they probably would be by using a more direct supervisory arrangement.

The evaluation of PR-CETP has other collaborative elements. People who are directly involved in carrying out project activities have also participated in the planning and implementation of the evaluation. For example, institutional coordinators have been actively

involved in carrying out “flashlight projects” with evaluators. These are small action research projects that focus on issues of special interest to the institutions. Topics such as, effectiveness of specific reformed courses and identification of future teachers who are not classified as education majors, have been researched. Another instance of collaborative evaluation is the development of portfolios by professors who are piloting reformed courses in the different institutions. In these portfolios, they document the changes carried out in their practices of teaching and assessment, and evaluate the obtained results. They also reflect on their own practice and the way they could use the obtained results to improve their practice. The approach used is that of reflexive practitioners to carry out action research in their classrooms [16]. To implement this strategy, a guide for the development of the portfolios was developed. Meetings with the professors before and after developing them were carried out. During the latter, they evaluated their work, as presented in the portfolios, using a rubric that had been previously developed by project staff and evaluators. Throughout this process, the development of a community of apprentices was promoted in which all people collaborate toward the achievement of a common goal. The latest evaluation of portfolios indicated that, although changes in the aimed direction have occurred, many transformed courses still do not reflect the PR-CETP constructivist philosophy to the degree that we would expect. These results were recently presented to institutional coordinators, and institutional plans are going to be developed to overcome this limitation. In a recent planning meeting, a more comprehensive collaborative professional development model that integrates a cycle of in-service training, reflective follow-up, and formative evaluation carried out in collaborative groups was discussed. The main aim is to strengthen the development of the community of apprentices.

### **Utilization of Evaluation Results**

Another debate in the evaluation field concerns the use of results produced by evaluations [2]. One of the important elements of this debate focuses on the responsibility that evaluators are supposed to have for the use of evaluation results [17]. On the one hand, it has been argued that stakeholders and other potential users are solely responsible for the use of evaluation results. The role of the evaluators is to carry out methodologically sound evaluations. This was the position of evaluation theorists belonging to the first stage of the development of program evaluation [2]. Theorists such as Striven and Campbell represent this position.

In later stages of program evaluation development, it was pointed out that the worth of the evaluations must be judged by their utility [1]. Program evaluation loses sense if the findings are not used. Therefore, promotion of use is now considered one of the main responsibilities of

evaluators. This position posits that evaluators have to balance technical rigor with utility. Research results document this position. Weiss and Bucavalas, in Rossi, Freeman, and Leviton, studying reactions of decision-makers to actual research reports, found that decision makers apply both a *trust test* and a *utility test* [1]. Truth was judged on the basis of research quality and on conformity to prior knowledge and expectations. Utility was judged on the basis of feasibility potential and degree of challenge to current policy. These results evidence the complexity of the utilization process.

Diverse strategies have been developed to foment the use of results; they involve tasks that should be carried out before, during, and after the evaluation. In the planning stage of the evaluation, it is important to identify potential users and on what aspect of the program they would like the evaluation to focus. It is also important to identify the type of information that the identified users need and involve them in the planning of the evaluation.

During the implementation of the evaluation, it is important to frequently interact with potential users in order to stay alert about the need for useful evaluation information. Based on these needs, the evaluation design should be adjusted. It is also recommended that potential users be invited to participate in this phase of the evaluation and then provided timely partial results as the evaluation progresses. During the evaluation, as well as after its conclusion, evaluators should prepare brief executive summaries, present findings and recommendations in diverse forums, and present them in forms that are suitable to the specific audience. In our project, we frequently present results in different formats and forums. For example, to present findings from the review of portfolios several exemplars illustrating the observed results were developed and presented in a coordinators meeting. These were analyzed by the coordinators to identify strengths and weaknesses. Using a nominal group technique, some strategies for dealing with the problem were explored and prioritized.

Evaluators can promote diverse types of use and they should adapt the promotion strategies to the type of use and target audience. Several types of uses have been identified [1]. Instrumental or direct use involves the documented and specific use of evaluation findings by decision makers and other stakeholders. Conceptual utilization refers to the use of results to influence thinking about issues in a general way. In the case of educational programs, the issue may center on specific aspects of education; for example, the teaching-learning process. Persuasive utilization refers to use of evaluation findings to support or to refute political positions, that is, to defend or attack the status quo.

The type of use that has been mainly promoted in PR-CETP has been instrumental for direct utilization, since the project is in the fourth year of a five-year term. The use of results for improvement of the project as it is developing has been promoted. During the planning stages, project administrators and institutional coordinators were identified as potential users of evaluation results. These personnel participated in diverse ways in the refinement of evaluation plans and development of initial data gathering instruments. During the implementation stage, frequent interaction with these potential users have occurred in meetings, as well as in informal conversations in order to stay alert about the needs for evaluation information. Based on the identified needs, the evaluation design has been adjusted. Partial results from the evaluation have been provided in different ways. Institutional evaluators provided information to institutional coordinators about results coming from yearly student and professor focus groups, interviews with deans or directors, and the professors' collaborative survey. In fact, some coordinators have participated in some of these data gathering processes. At the central level, results are presented in both informal and formal ways. An example of the former is the presentation of exemplars illustrating the results obtained from the review of portfolios that was previously described. Examples of the latter are formal oral presentations to coordinators, administrators, and other project participants, as well as written reports for the NSF. Conceptual use is starting to be promoted through publication of articles in this special edition.

### **Final Comments**

Upon facing an evaluation, it is necessary to make many decisions. One of them is the epistemological position that will sustain it. Many times, this position is taken based on the training the evaluators have received. However, this should not be the main element that prevails in this decision. Recent developments in contemporary science indicate that there is no one single paradigm or method that allows us to better understand a phenomenon. Each of them has strengths and limitations. From this position, it follows that the selection of methods should be based on their utility for specific purposes. Moreover, the use of mixed methods could provide a better understanding of the studied program than can be achieved with any one method alone. Upon facing the task of making evaluations, it is thus no longer pertinent to ask which of the two approaches, qualitative or quantitative, should be used. Rather, it is pertinent to ask what method or what combination of methods would be more appropriate to achieve the established purposes and answer the posed evaluation questions in the context of the specific program.

Other decisions faced when carrying out an evaluation are interrelated to the epistemological one. The fact that most evaluators no longer believe that their methods yield

absolute and objective truth, but rather all data collection, qualitative as well as quantitative, operates within a socio-cultural context and is affected to some extent by the perceptions and beliefs of investigators and data collectors, has influenced how evaluation is conceived and practiced today. For example, it is no longer considered necessary for evaluators to maintain a distance from program personnel in order to avoid biases, but rather that they have a clear understanding of their potential influence and deal with it professionally. Moreover, the need to balance technical rigor with utility has also influenced how evaluation is practiced. Promotion of use is now considered one of the main responsibilities of evaluators. Internal evaluations can enhance instrumental use since evaluators can more intimately understand the programs and be more attuned to stakeholders' information needs. Collaborative evaluations also enhance utilization since close collaboration with individuals who will use evaluation findings will ensure that the evaluation is responsive to their needs and produces information that they can and will actually use [6]. The PR-CETP evaluation is an example of an evaluation with these characteristics. Evaluators can collaborate with science, math, and education faculty to promote changes that enhance teacher preparation. ■

## Bio

Milagros Bravo is Professor of Educational Research Methods in the College of Education, at the University of Puerto Rico, Río Piedras, and Coordinator of the Evaluation Component of the Puerto Rico Collaborative for Excellence in Teacher Preparation (PR-CETP).

## References

- [1] P.H. Rossi, H.E. Freeman, and M.W. Lipsey, *Evaluation: A systematic approach*, Sage Publications, Thousand Oaks, CA, 1999.
- [2] W.R. Shadish, T.D. Cook, and L.C. Leviton, *Foundations of Program Evaluation: Theories of Practice*, Sage Publications, Newbury Park, CA, 1991.
- [3] T.D. Cook and C.S. Reichardt, *Qualitative and Quantitative Methods in Evaluation Research*, Sage Publications, Beverly Hills, CA, 1979.
- [4] Y.S. Lincoln and E.G. Guba, *Naturalistic Enquiry*, Sage Publications, Newbury Park, CA, 1985.
- [5] E.G. Guba and Y.S. Lincoln, *Fourth generation evaluation*, Sage Publications, Newbury Park, CA, 1989.
- [6] M.Q. Patton, *Qualitative Research and Evaluation Methods*, Sage Publications, Thousand Oaks, CA, 2001.

- [7] J. Frechtling and L. Sharp (eds.), *User-Friendly Handbook for Mixed Method Evaluations*, National Science Foundation, Arlington, VA, 1997.
- [8] F. Capra, *The Turning Point: Science, society and the rising culture*, Bantam Books, New York, 1982.
- [9] R. Descartes, "Discourse on the method," in E. Anscombe and P.T. Geuch (eds.), *Descartes Philosophical Writings*, Bobbs-Merrill, Indianapolis, IN, 1971.
- [10] D. Polkinghorne, *Methodology for the human sciences: Systems of inquiry*, State University of New York Press, Albany, NY, 1983.
- [11] J.G. Greene and C. McClintock, "Triangulation in Evaluation: Design and Analysis Issues," *Evaluation Research*, **9** (1985) 523-547.
- [12] A. Fink, *Evaluation for Education and Psychology*, Sage Publications, Thousand Oaks, CA, 1995.
- [13] C.E. Depner, E. Wethington, and B. Ingersoll-Dayton, "Social Support: methodological issues in design and measurement," *Journal of Social Issues*, **40(4)** (1984) 37-54.
- [14] E. Chelimsky, "The Politics of Program Evaluation," in D.S. Condray, H.S. Bloom, and R.J. Light (eds.), *Evaluation Practice in Review*, Jossey-Bass, San Francisco, CA, 1987.
- [15] A.J. Love, *Internal Evaluation: Building Organizations from Within*, Sage Publications, Newbury Park, CA, 1991.
- [16] G.E. Mills, *Action research: A guide for the teacher researcher*, Merrill, Upper Saddle River, NJ, 2000.
- [17] M.Q. Patton, *Utilization Focused Evaluation*, Sage Publications, Thousand Oaks, CA, 1997.