# The Contribution of Metaevaluation to Program Evaluation: Proposition of a Model

Helga C. Hedler and Namara Gibram

*Faculdade Alvorada, Curso de Administração, Brasília/DF*

**Background:** This theoretical article points to the fundamental difference between meta-analysis and metaevaluation. A model of metaevaluation for social programs is presented based on prior practical research.

**Purpose:** The purpose is to present a model for metaevaluation as a tool that can be used in other studies. Theory points to the need of a qualitative framework to go beyond the understanding of meta-analysis for program evaluation.

**Setting:** This theoretical article is based on an empirical research conducted at a Brazilian Governmental audit agency.

**Subjects:** The Government agency where the practical research was conducted is responsible for the effectiveness and accountability of social programs through audits that occurred from 2003 to 2006.

**Intervention:** Meetings and interviews were held with auditors that participated in the evaluation process going from planning to final reports as the model proposes.

**Research Design:** The model for metaevaluation has a qualitative approach used to evaluate prior evaluations for social programs.

**Data Collection and Analysis:** Data collection included structured interview with the chief manager of the agency in charge of evaluating governmental programs. Documents and reports were analyzed using qualitative method for content analysis. Synthesis of categories was applied to compare different analysis and summarize findings.

**Findings:** Metaevaluation and meta-analysis are different research methods with a different approach. Meta evaluation is a qualitative method useful when evaluating prior evaluations. Yet the quantitative approach of meta-analysis applies better for first evaluations. Meta evaluation may include other methods to help strengthen the evaluation results.

**Conclusions:** Metaevaluation aligns theory and practice for program evaluation. The proposed model for metaevaluation may hold value for future theoretical and empirical work.

*Keywords:* metaevaluation; program evaluation; evaluation use

This theoretical article underwent the challenge of increasing knowledge on program evaluation: metaevaluation, a theme with few studies conducted in Brazil. The model proposed herein was based on data obtained by an auditing study carried through by the Brazilian Federal Audit Court (TCU). The result of this metaevaluation research composes the scope of another article; therefore the present work focuses on the theoretical and explicative traces of the premises which sustain the metaevaluation model and its applications.

## Evaluation of Programs and Their Concepts

The term evaluation can take several *lato sensu* meanings; among them, evaluations which are generally made in daily relation to things, people or situations (Cano, 2004). In such evaluations, value judgments are made. Therefore, in this sense, evaluating consists in issuing a value judgment or attributing value to something. This generic definition may be applied to several deliberations performed regularly and it refers to evaluation in the informal sense. Formal and systematic evaluation is used to evaluate services or professional activities; it utilizes the same methods and techniques present in social research (Aguilar & Ander-Egg, 1995).

Evaluating means to determine merit, cost and value (Fernández-Ballesteros, Vedung, & Seyfried, 1998; Posavac & Carey, 2003; Stufflebeam & Shinkfield, 1987). Evaluation is a necessary task that constitutes part of programs, public policies, private projects, public regulations, public and private interventions.

The evaluation of programs, referred in this article as *evaluative research,* goes beyond these concepts and presents the discussion of evaluation as method, subject and establishment of scientific patterns. "...The development of the evaluative research presents at its core not only the importance of the evaluation as a judgment tool for procedures and actions, but also the concept that the evaluation represents

production of knowledge" (Barreira, 2002, p. 17).

In the case of public policies that bring forth plans and goals by program action, evaluation is a tool that propitiates information of the results reached by these programs (Ala-Harja & Sigurdur, 2000). Rossi and Freeman (1993) understand that the evaluative research must use the scientific method as a means to investigate social problems.

Oskamp (1984) characterizes evaluation of programs as an attempt to evaluate the operation, the impact, and the effectiveness of programs in public and private organizations. Program evaluation was developed by applying a scientific method to the knowledge of reality based on the stages and demands for such methods. Moreover, the collection and systematization of data for the conduct of program evaluation requires the adoption of valid and trustful procedures, in order to have considerable and useful results (Aguilar & Ander-Egg, 1995).

Aguilar and Ander-Egg (1995), revised several definitions for program evaluation and proposed one that summarizes what other authors such as Stufflebeam and Shinkfield, (1987), Fernández-Ballesteros et al., Vedung and Seyfried (1998), Cano (2002), Posavac and Carey (2003) have declared. The definition states that program evaluation is "a kind of social research applied in a systematic, planned and directive way in order to identify, obtain and provide valid and trustful data…to support judgment of merit and value of different components in a program…" This definition expresses the sense of utility that program evaluation bears as a practice connected to reality and to the needs of users, stakeholders, and those involved with the program, aiming for the enhancement of service rendering.

Regarding service rendering, according to Gray, Jenkins, and Segsworth (1993), quoted by Fernández-Ballesteros et al. (1998), the control of public expenses and management of assistance programs or policies have been the

main focus of program evaluations in the past three decades. Therefore, there would be two perspectives within the evaluation of programs: the first directed to contribution, planning and improvement of the program and the second considering the verification of its effectiveness and impact. Other than legal principles, regulation and financial management of public finances aiding actions, program evaluations are instruments for controlling government actions within the public scope.

## Evaluation Ex-ante, Intermediate, and Ex-post

A definition of evaluation ex-ante is provided by the Evaluation Research Society (ERS) (1988) which defines it as analysis of start-end, pre-installation, viability analysis or contextual analysis. This definition includes evaluative activities that come before the implantation of a program. *Ex-ante* evaluation aims to ratify, to research or to emit a precise estimative of conception sufficiency, operational viability, sources of financial resources, and availability of organizational support. The results provide a useful direction to refine the program planning, determining the appropriate implantation level, and the decision regarding the installation or not of the program.

The *intermediate evaluation* is one of the ways of obtaining knowledge about the program. It aims to subsidize the program management procedure as feedback for its implantation and development. In this case, the evaluators and clients are generally internal, most likely program managers. Evaluation issues assessed are those related to event management, which are connected to program impact (Ala-Harja & Sigurdur, 2000). Its main contribution lies in the program formulation (Posavac & Carey, 2003).

According to ERS (1998), an intermediate kind of evaluation is formative evaluation, also known as the evaluation process in a continuous program, aiming for modifications and improvements. Its activities may include the management of strategy analysis, evaluation of human resources, and attitude research regarding the program. In some cases formative evaluation involves the development of field research on a small scale before a more comprehensive implementation. The informative evaluator works in a team along with the formulators and program administrators, and they participate directly in the decision making to perform all the necessary changes.

*Ex-post* evaluation deals with the evaluation of a working program. This kind of evaluation is conducted when the program has been implanted, in order to reach stated objectives (Ala-Harja & Sigurdur, 2000). For this reason, it is also called *additive evaluation*. Additive evaluation influence programs, projects, and plans.

## Program, Project, and Plan

The program, project, and plan modalities are social interventions which differ in scope and duration. Hence, the project is a "minimal unit for the destination of resources and by means of an integrated set of activities, a way to transform part of reality, provisioning for a scarcity or altering a problematic situation" (Cotta, 1998, p. 104). A set of projects aiming for the same objective form a program. Finally, the plan aggregates similar programs, thus defining the directives for social interventions.

The plan conception demands a wider comprehension when dealing with social intervention. For instance, in Brazilian public policies, plans are developed to establish directives for a policy. Multiyear plans created by the government have a wide scope: they predict directives, costs, budgets for the areas in which the government will work on, and enable programs to be unfolded in several areas.

## Similarities and Differences between Auditing and Program Evaluation

The plan conception demands a wider comprehension when dealing with social intervention. For instance, in Brazilian public policies, plans are developed to establish the directives for a policy. The multiyear plans created by the government have a wide scope: they predict directives, costs, budgets for the areas in which the government will work on, and enable programs to be unfolded in several areas.

In the early fifties, the auditing searched for the rationalization of management and distribution of resources for the Defense Programs and Missions of the Government. This effort increased in the Department of Defense, Planning, Programming and Budget Systems. However, such effort was peripheral and related to the accounting-perspective verification, whose main goals were: (1) planning the cost-effectiveness of the program and then evaluating such cost-effectiveness, and; (2) checking if the cost-effectiveness was the result of the planning procedures. Despite this restrict approach, analyses such as: technical-political analyses, cost-benefit and cost-effectiveness, were also conducted by the economic area as an attempt to comprehend program activities. However, the focus of the auditing was on planning so that the techniques could outline the probable future results of the programs. They were not aimed at identifying the current effects of program implantation of existing policies (Chelimsky, 1985).

## Metaevaluation: Characterization, Background and Differences Regarding Meta-analysis

### Meta-analysis

Meta-analysis can be described as "…a statistical technique utilized in the development of syntheses with general conclusions, regarding several studies investigating similar areas of research…" (Smith & Bond, 1999, p. 15). The meta-analysis calculates the effect size regardless of the complexion of the real standard used by a certain researchers. The size effect of a study is the result of the difference between the scores obtained by the experimental subjects and the control group, divided by the standard deviation of the scores of subjects in the control group. The size effect provides the average of different studies determining whether or not the experimental effect that was being investigated in a consistent way, could be found. If the sampling of studies is large enough, the influence of variation in the experimental delineation, geographic localization, study data and size effect can be predicted.

In areas where there is a great quantity of studies about a certain object, it is possible to have quantitative literature reviews and these studies become known as meta-analysis (Hunter & Schmidt, 1996, 1999; Rossi & Freeman, 1993).

Although in the exact sense, meta-analysis is not a delineation of research, it is an alternative to evaluation projects that can be useful in some situations, for instance, when more time is necessary for collection of original data. The findings of the meta-analysis are particularly useful in the delineation stage of a program, because they summarize the existing knowledge regarding similar programs which have been implanted, therefore providing knowledge for the new program.

Some authors confuse the meta-analysis procedure with the metaevaluation method. For instance, Ashworth, Cebulla, Greenberg, and Walker (2004) conducted a metaevaluation utilizing the meta-analysis procedure. They justify the metaevaluation in the meta-analysis procedure because they believe it favors the explicative power of replication, rigorous accumulation of evidence, revision and summarizing. Besides that, they disagree with authors such as Patton (2001) and Günther

(2006), who define metaevaluation as a qualitative method. The previous researchers believe that the qualitative approach is insufficient to support the metaevaluation and have not yet realized that this kind of perspective is now surpassed in scientific literature. The research method must be chosen by taking into consideration, among other factors, the characteristics of the phenomenon to be studied and whenever possible, both approaches should be applied for a wider and deeper comprehension of data and the reality in discussion.

The objective of the meta-analysis is to provide a description of real correlation distributions between independent and dependant variables. Therefore, if all the studies have been correctly conducted, the distribution of correlations can be directly used to estimate the distribution of the real correlation and if not, be submitted to corrections (Hunter & Schmidt, 1996). The meta-analysis approach will significantly contribute to the delineation of programs that have considerable gain, taking into account the existing social science research and professional reports of established programs.

*Metaevaluation*

Metaevaluations bear three main characteristics (Woodside & Sakay, 2001):

1. They are syntheses of findings and inferences of evaluative research about the program performance. They report the effectiveness of managing the goals achieved by the programs and provide information about two characteristics: *Well* managed programs and *poorly* managed programs.
2. They inform about the validity and utility of evaluation methods, offering guidance regarding useful evaluation methods.

3. They provide strong evidence regarding the program impact, subsiding the decision making process regarding it. Hence, the results of the metaevaluation assist and justify the increase of trust by the interested parts and managers of programs in the evaluation results.

Historically speaking, metaevaluation started in 1960 when evaluators such as Scriven, Stake, and Stufflebeam began discussing procedures and formal criteria (Worthen, Sanders, & Fitzpatrick, 2004). The term "evaluation of the evaluation" was created by Orata in 1940 and metaevaluation by Scriven in 1969 (Cook & Gruder, 1978). In accordance to Patton (2001), a metaevaluation is a re-analysis of an evaluative study, which has been already concluded; taking into consideration several aspects of the previous study such as methodology, subject selection, adopted criteria, results and analysis.

Guba and Lincoln (quoted by Schwandt, 1989), stated that the concept of metaevaluation was conceptually modeled as an inspection audit, introduced to establish the validity of naturalistic research (qualitative). The definition of Schwandt corroborates the definition of Patton (2001), because it comprehends the metaevaluation as a method of checking the quality of an evaluation. For such, it requires the examination of the evaluation method and its procedures to reach results and conclusions. Yet, for Woodside and Sakay (2001), metaevaluation includes the evaluation of utility and validation of two or more studies, which comprise the same issue.

Sometimes, metaevaluation is confused with meta-analysis mainly by those that do not use this method to evaluate programs. In order to clarify this matter, there is a need to compare them pointing to their similarities and differences. Table 1 shows the characteristics of the study object, application procedures, data analysis and distinction between meta-analysis and metaevaluation.

Table 1
Comparing Meta-analysis and Metaevaluation

| Characteristics | Meta-analysis | Metaevaluation |
|---|---|---|
| Study object | Any kind of study | Concluded evaluation (s) |
| Data source | Secondary | Secondary |
| Aplication procedures | Different studies are organized, following a criterion or variable, utilizing a temporal or thematic approach | Selection of the concluded evaluation (s) regarding the evaluative study or different studies with the same thematic |
| Data analysis | Quantitative (statistics). A synthesis of similar findings, calculating the size effect among the studies. | Qualitative (content analysis, criteria analysis). A new evaluation is done. The procedures and methods are compared to prior studies applying pre-established criteria. Improvements are suggested or a new model is presented. |
| Usage | Generally academic, but can also subsidize professional practices. | Either academic or professional. Serves as a reference for programs of the specific field studied. |

As highlighted in Table 1, there are similarities between meta-analysis and metaevaluation. Similarities occur when secondary sources of data and their applications are present. The specifications appear in relation to the study object and the data analysis procedure. In the meta-analysis, the object can be any type of study. In the metaevaluation, the study object is exclusively composed of evaluations which have been already concluded.

In meta-analysis, the procedure for data analysis is quantitative and statistical, with the calculation of the size effect providing the average between different studies, determining if there was or there was not the investigated experimental effect. It also indicates the consistency among the findings. Yet, the metaevaluation uses qualitative analysis procedures such as: Content analysis or the criteria checking by international organizations of evaluation as the Joint Committee or ERS (1998).

## Quality Standards for Metaevaluation

In the area of government program evaluation, there is not a unique set of standards for the auditor's procedures as a meta-evaluator (Schwandt, 1989). The quality standards for metaevaluation originate together with it.

Posavac and Carey (2003) wrote a chapter in their work *Evaluation* to clarify the establishment of criteria and standards in the evaluation of projects. Therefore, evaluating would demand the issuing of a judgment based in values and also the establishment of criteria and standards. According to Posavac and Carey, the established criteria and standards need to be clear and explicit so they can subsidize useful evaluations. Therefore, they must represent the objectives of the program, the institutional efforts, measurable and trustful characteristics, including those selected with the *stakeholders*.

The evaluators of the period from 1960 to 1970 created lists to check what would constitute a *"good"* or *"poor"* program evaluation. Hence, at the end of the seventies, a project was launched aiming to develop a set of directives applied to the educational evaluations to be established as a general consensus of evaluation quality. The formulation of these directives started in 1975, and was coordinated by Stufflebeam, with authorization granted by the Joint Committee on Standards for Educational Evaluation, since then known as the Joint Committee (Worthen et al., 2004).

The directives of the Joint Committee consist of 30 standards, including definitions, fundamental logic, directives, common errors, illustrative cases, and descriptions of evaluation practices. In accordance to Stufflebeam and

Shinkfield (1987), the norms for program evaluation are: utility, viability, propriety, and precision.

The utility norms are those directed to people and groups that have the task of evaluating in other words, those directly responsible for the evaluation process. Such norms should help identify the "good" and "bad" functioning of the evaluated object, providing clear information regarding the virtues and defects in the evaluation, besides providing suggestions for improvement.

The viability norms refer to the use of evaluative procedures, which can be utilized, considering and applying the possible control over political forces, which may somehow interfere in the evaluation.

The norms related to ethics in evaluation relate to the explicit commitments, which assure the cooperation, protection, the rights of those involved in the evaluation and the accuracy of results.

Finally, the precision norms are those that clearly describe the evaluated object in its evolution, context, revealing virtues, defects in evaluation planning, proceedings and conclusions.

## Methodology for Metaevaluation

The metaevaluation data can be worked by different qualitative analysis techniques. The authors of this article suggest: Content analysis, summary of categories and conceptual model of the program and checking of criteria by the *Joint Committee.*

*Content analysis.* The content analysis is a technique for text analysis aiming to obtain through systematic and objective procedures, recurring themes grouped to compose an empirically defined category. These categories facilitate the interpretation of data related to the research object. Among the several types of category analysis, the *theme analysis* is widely utilized (Bardin, 1977).

The procedure for content analysis starts by reading the common parts of the text. After that, counting rules are established for the recurrence of words or sentences based on the theme. The procedure continues with the theme analysis to identify the nucleus sense for the text sentences, considering the frequency in which they appear and their relevance to the research interest.

The previously selected themes are then grouped in categories, considering how often they appear, homogeneity among them, pertinence and exclusivity (Bardin, 1977). It is advisable that the categories be submitted to *judges* for semantic analysis. The categories can be previously established or they can freely emerge from the analyzed text.

*Synthesis of categories.* The synthesis of categories was developed by Gibram (2004) as a way to broaden the scope for content analysis. The procedure consists in regrouping the categories in thematic axes. These axes can be previously constructed according to theoretical parameters being studied or defined from the analyzed contents. The axes are important in research dealing with great quantity of information or complex themes. The procedure of creating thematic axis can also be useful when there are many categories, hindering the conclusion of the results.

The thematic axes formed by grouping categories provide a broader and more realistic view of the problem being studied. If all themes were analyzed, as Bardin (1977) suggests, grouped in simple categories, the study of more complex problematic issues would lose the representation of relational and textual significance. Hence, the thematic axis, formed by several related categories preserve the significance while organizing multiple themes for analysis and result reports.

*Program conceptual models.* The social programs are delineated for action under different problematic situations, and as such, they bear a

conceptual and technical structure to support them in accordance to the area of program execution. For example, the programs in the area of social assistance are based on individual rights and social policies that follow the Operational Norms (NOB) established by the Unique System of Social Assistance (SUAS). These are guidelines for the implantation and management of program procedures. Therefore, during the execution of the metaevaluation, it is necessary to search in the specific literature related to the program, for sources that contextualize the program thematic as well as other program evaluations similar to the meta-evaluated ones.

In the specific literature about metaevaluation and in the search of a proper model for execution, different procedures were found for data analysis as means to subside results and conclusions. For instance, Woodside and Sakai (2001) utilized as procedure for metaevaluation analysis the theoretical model of Kotler (mentioned in Woodside & Sakai, 2001), designed specifically for the area of Federal Marketing and Tourism. In order to evaluate the program planning, they compared the procedures applied in the previous evaluation with the premises of a SWOT analysis. To evaluate the program implementation they traced comparisons of the previously adopted procedure with the concepts of Mintzberg (mentioned by Woodside & Sakai, 2001), about planned and deliberated strategy. The results of the previous study were analyzed under the perspective of the use of impact indicators for the Federal Marketing and Tourism Program. Moreover, at the end of this article, Woodside and Sakai proposed a model for future evaluations of similar programs.

*Checking the Criteria According to the Joint Committee.* According to the orientation of the Joint Committee, in order to execute metaevaluation, a checklist must be constructed based on the criteria to be contemplated in an evaluation. In the example showed in Picture 2, questions

were chosen based on the criteria referring to the methodology of evaluative research for which consulting of experts in specific program theme domain was not necessary.

Worthen et al. (2004), believe that the verification of the Joint Committee checklist should go beyond the indication of the use or not of the criteria (check yes or no). They suggest the adoption of a scale with scoring points to measure the criteria. For example, a scale from zero to three as indicated in Table 2.

In order to correctly score the questions in the Checklist presented in Table 2, the evaluator must comprehend that:

- ✓ In Question 15, the measures to guarantee the minimum quantity of errors refer to: Application of a pilot test; control group; data collection before and after; random sampling or another procedure of control for internal and external validity in evaluative research.
- ✓ In Question 18, in order to evaluate *the adequate training of the team for the execution of the auditing*, the scoring of all questions must be considered because they refer to the adequate use and how pertinent the methodological processes are, what kind of analysis, results, conclusions and recommendations are produced. Hence:
  - ▪ *Without training:* the team who gets scores greater or equal to 49% of the total (less than 9 questions);
  - ▪ *Partial training:* scores from 50 to 69% (from 9 to 12 questions);
  - ▪ *Adequate training:* the team who gets scores "yes" in more than 70% of the total (13 questions).
- ✓ In Question 19, *there was no participation* refers to when there was no hiring of specialists or any consultation asked of them. P*artial* was considered when specialists participated in only one stage (planning or evaluation execution). *There*

was participation when consultants were hired or consulted.

Table 2 shows a sample for verifying the questions referring to the Joint Committee criteria. Such questions are related to the political context, program characteristics, approach, methods, techniques and difficulties for the execution of evaluation.

Table 2
Checklist of Questions Based on the Criteria of the Joint Committee

| Questions | Does not apply | no | partially | yes |
|---|---|---|---|---|
| **Politic context** | 0 | 1 | 2 | 3 |
| 1. Were the program audience and participants of the audit identified? | | | | |
| 2. Did the report clearly describe the program context? | | | | |
| 3. Did the audit consider how the different groups of interest acted in the program? | | | | |
| **Characteristics of the program** | | | | |
| 4. Was the collected data broad enough to understand the functioning of the program? | | | | |
| 5. Did the report clearly describe the program? | | | | |
| 6. Did the report clearly describe the objectives of the program? | | | | |
| **Auditing approach** | | | | |
| 7. Were the in formations collected sufficient to reflect the objectives of the audit? | | | | |
| 8. Did the report clearly describe the results of the audit? | | | | |
| 9. Did the report clearly describe the conclusions of the audit? | | | | |
| 10. Did the report clearly justify the recommendations made by the audit? | | | | |
| **Methods and techniques** | | | | |
| 11. Were the techniques of data analysis explicit? | | | | |
| 12. Did the report clearly describe the methodological procedures of the audit? | | | | |
| 13. Were the procedures of information collection clearly described? | | | | |
| 14. Were the instruments for information collection valid? | | | | |
| 15. Were all the necessary measures taken in order to assure the minimum amount of errors during the data collection? | | | | |
| 16. Was the quantitative information adequately analyzed? | | | | |
| 17. Was the qualitative information adequately analyzed? | | | | |
| **Auditing accomplishment difficulties** | | | | |
| 18. Did the auditing team have adequate training to undergo the audit? | | | | |
| 19. Did external consultants for specific areas participate in the audit? | | | | |
| 20. Were the audit's resources (time, money and employees) adequate for accomplishing the foreseen activities? | | | | |

## Proposition of a Model for Metaevaluation

Metaevaluation itself can be the object of another metaevaluation; in this case several requirements must be considered. The metaevaluation demands a set of procedures, standards and criteria for judging the evaluation quality (Schwandt, 1989).

According to Patton (2001), Schwandt (1989), and Woodside and Sakai (2001), metaevaluation can be defined as a method of research where one or more stages of the evaluative studies concluded are re-analyzed; there is a comparison of the previous evaluations with quality standards and validity accepted in the scientific community and at the end there is a new evaluation issued regarding the analyzed evaluative study.

In the revised literature (Ashworth et al., 2004; Chelimsky, 1985; Cook & Gruder, 1978; Patton, 2001; Schwandt, 1989; Stufflebeam & Shinkfield, 1987; Woodside & Sakai, 2001;), studies pointed to different procedures for conducting metaevaluations and also different conceptions regarding the process. In this sense, no direct response to the question of the necessary stages and techniques for the conduction of a meta-evaluative study were found.

The procedure of meta-analysis as means to obtain a metaevaluation was not considered because it would only answer to the main purpose of metaevaluation if, at the end, a new evaluation regarding the analyzed evaluation procedure had been drawn. The meta-analysis can be utilized in metaevaluation only if it is used together with other qualitative procedures validated by program evaluation associations, to finally generate a new evaluation.

The authors of the present article disagree with Ashworth et al. (2004) that consider meta-analysis the *best way* or a self-sufficient procedure for conducting a metaevaluation.

## Conceptual Premise: The Programs in the Social Reality

The Brazilian social reality has structural problems, which produce hunger, poverty and social disaggregation. In this article, the social program is understood as a "systematic intervention planned with the purpose of achieving change in the social reality" (Cano, 2004, p. 9). The social programs are developed by public policies and emerge to supply the needs detected in the environment of a certain population (Posavac & Carey, 2003).

The social programs are created to intervene in these situations; however, due to their originating complexity, the possibility of action is limited and may cause both advances and regressions. An advance is considered when these programs transcend governments and become continuous services. This way their execution becomes independent of the government policy, being assured by the Laws and Policies of the State.

The TCU bears a social function associated with the control and supervision of the public affairs, as well as it's patrimonial and economic aspects regarding public administration (Mendes et al., 1999). Therefore, the court is responsible for conducting the audit of social programs. The evaluation modalities used are auditing of operational performance and program evaluation (Brasil, 2000).

In Brazil, the discontinuity of social programs is very common. This is more evident in large-scale programs that produce little documented and systematized results. Notwithstanding governmental planning, focus is generally on the development of plans, programs and projects, neglecting the stages of inspection, procedure evaluation, results and impacts (Silva, 2002).

The social reality in which the social programs are inserted present challenges for the management of programs, both in the effectiveness of actions and in program evaluation. Hence, the social reality being fluid and mutable supports the drawn programs that need constant evaluation and monitoring. These evaluations also need to be meta-evaluated.

## The Conceptual Model of Metaevaluation

The graphic model presented in Figure 1 is supported by the supposition that the metaevaluations are applied to evaluative studies within the context of social reality, and that this context may influence its realization.

The meta-evaluative studies contemplate previous studies, in any program phase, whether they are *ex-ante*, intermediate and *ex-post* and they influence any study bearing an evaluative drawing (e.g., evaluation, policy, plan, project program, auditing).

Besides that, the metaevaluation depends on a set of quality criteria to make it valid. Such value judgment criteria are shared by the

international community of evaluators as a necessary guide for the evaluation of another evaluation. In the current model, the quality standards and validity of metaevaluation adopted were the ones recommended by the Joint Committee to attest the global quality of the evaluation and issue a new evaluation in relation to the previously conducted one. In the example described in Table 2, the checklist of questions was based on the criteria of the Joint Committee.

The metaevaluation may have a set of analysis procedures in order to reach a final result – the emission of a new value judgment, a new evaluation. This article has presented some quality standards and the methodology used in metaevaluation involving techniques of data analysis including case studies, content analysis, syntheses of categories and conceptual models in the specific program area.

The results of these qualitative data analyses subsidize at the end of the process, the judgment of the previous evaluative study comparing it to the criteria established in the conducted analysis, therefore enabling the emission of a new evaluation. Hence, the new evaluation will present the strong aspects to be valued and the weak aspects to be corrected.
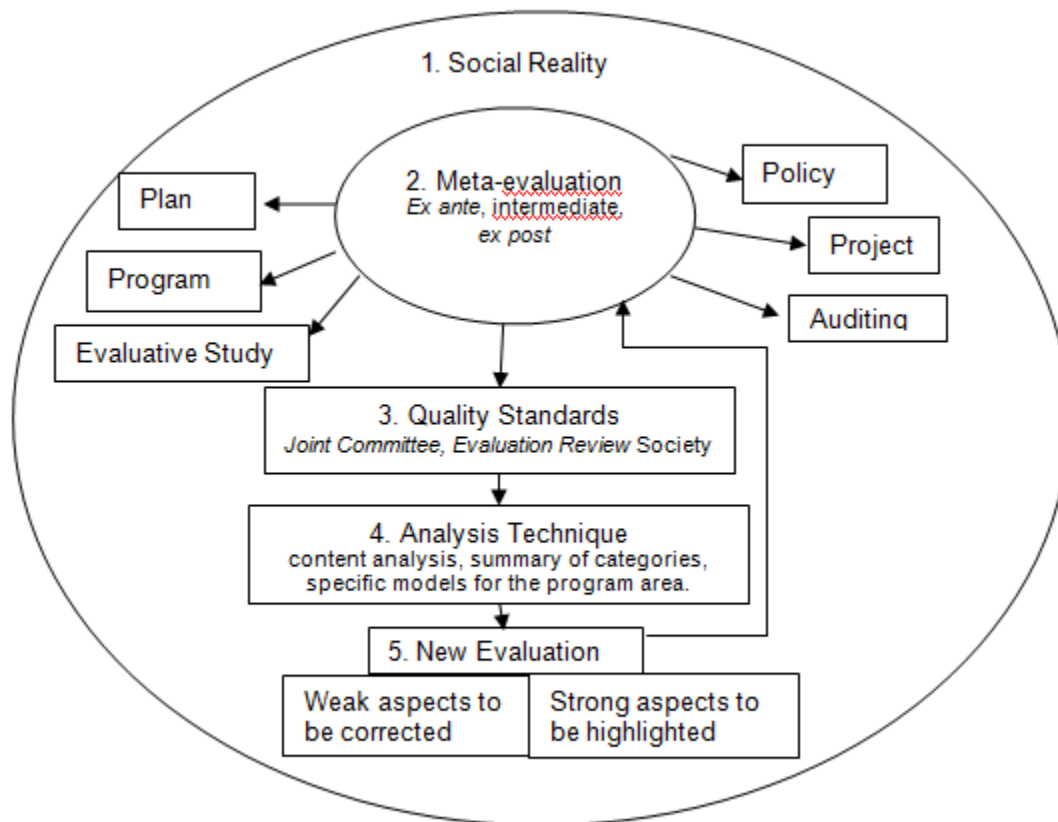


Figure 1. Conceptual Model of Metaevaluaiton

## Conclusion

Some considerations will be made in relation to the applicability of the metaevaluation. The metaevaluation can take several shapes and vary from professional critiques to evaluation reports or be used in procedures of re-analysis of original data. It can be formative or summative in kind. The summative metaevaluation is a flashback activity developed by an independent external agent over the process and the product of the evaluation comparing it to a group of patterns for evaluation. In order to do this kind of metaevaluation, the evaluator must have an extremely coherent political and prudent attitude and behavior towards the correct actions and procedures.

Metaevaluation demands hard work from those conducting it, as well as the collaboration of other consultants or researchers to judge the criteria utilized in the analysis of previous evaluations. Sometimes the meta-evaluator may develop different hypotheses and/or collect new information about the program studied in the previous evaluation. In case of programs generating wide public interest, the metaevaluations analyze the results of different evaluations of these programs (including evaluations of units or program components) in order to verify their global impact. Thus, it is important to remember the need to preserve the ethics, precision, and fidelity to the new metaevaluation results. When divulging the metaevaluation results, a great deal of caution must be taken in regard to questioning the previous evaluation. An ethical and cautious positioning is required from those conducting it in order to avoid bias and the inadequate use of results. A negative evaluation can harm the credibility and merit of a given institution or group of evaluators. It is important to have in mind that the meta-evaluators do not analyze the collected data, but the inferences other evaluators previously made about them, so they emit an evaluation based on personal inferences over the results obtained by other people.

The metaevaluation can be stimulated by several interests such as academic research or demands of the agencies that coordinate and supervise the program. It should be made clear that the evaluator does not have to accept the original results obtained by previous studies. This method is a qualitative instrument which provides the means for analyzing and implanting improvements in the existing evaluations.

This article has presented the procedures for data analyzes, such as synthesis of categories for content analysis which bears the potential to enhance the understanding of wide range categories originated from the reading of extensive printed material about the meta-evaluated auditing.

The applicability of this model can be observed in the propositions of improvements presented in the audit conducted by the Federal Audit Court. The method of data analysis provided the knowledge of the procedure for realization of ANOP, enhancing its strengths and weaknesses. Above all, it was verified that the auditing model adopted by the Federal Audit Court was positively evaluated by the Joint Committee in most of the established criteria. The suggestions for improvement referred to the methodological aspects regarding the sample that was used, the instruments for data collection and the need to improve the qualitative analyses performed.

The metaevaluation carried through by Hedler (2007) can also contribute to the improvement of social programs since the same suggestions for improvement presented in the previous evaluation can be applied by the programs themselves. They can implant them in their monitoring and internal evaluations.

Therefore, this article proposes to enhance the discussion about the utility of metaevaluation, as well as the discussion about the model presented and its future applicability.

# References

Aguilar, M. J., & Ander-Egg, E. (1995). *Avaliação de serviços e programas sociais.* Petrópolis: Vozes.

Ala-Harja, M., & Sigurdur, H. (2000, Outubro-Dezembro). *Em direção às melhores práticas de avaliação.* Revista do Serviço Público. Fundação Escola Nacional de Administração Pública, v.1, nº 1. 51. Brasília: ENAP.

Ashworth, K., Cebulla. A., & Greenberg, D. W, Robert (2004). Metaevaluation: discovering what works beast in welfare provision. *Evaluation.* Vol. 10. [On-line] Obtido em 18 de fevereiro de 2007, de http://evi.sagepub.com/cgi/content/abstract.

Bardin, L. (1977). *Análise de Conteúdo.* Lisboa: Edições 70

Barreira, M. C. R. N. (2002). *Avaliação Participativa de Programas Sociais.* São Paulo: Editoria Veras CPIHTS.

Brasil. (2000). Tribunal de Contas da União. *Manual de auditoria de natureza operacional.* Brasília: TCU. Coordenadoria de Fiscalização e Controle. 114p.

Cano, I. (2004). *Introdução à avaliação de programas sociais.* (2ª ed.) Rio de Janeiro: Editora. FGV.

Chelimsky, E. (1985). Comparing and contrasting auditing and evaluation: some notes on their relationship. *Evaluation Review.* vol. 9, nº 4, p. 483-503 [On-line], obtido em 18 de fevereiro de 2007, de http://erx.sagepub.com/cgi/content/abstract/9/4/483.

Cook, T. D., & Gruder, C. L. (1978). Metaevaluation research. *Evaluation Review.* [On-line], February 18th 2007, de http://erx.sagepub.com/cgi/content/abstract/2/1/5.

Cotta, T. C. (1998, Abril - Junho). *Metodologias de avaliação de programas e projetos sociais*: análise do resultado de impacto. Revista do Serviço Público. Fundação Escola Nacional de Administração Pública. 49, nº.2. Brasília: ENAP.

Evaluation Research Society [ERS] (1998). *Standards for program evaluation.* San Francisco, CA: Jossey-Bass.

Fernández-Ballesteros, R., Vedung, E., & Seyfried, E. (1998). Psychology in program evaluation. *European Psychologist, 3,*143-154.

Gibram, N. F. R. (2004). *Trabalho e familia: um estudo da congruência dinâmica de demandas multiplas* (*Work and family: a study on the dynamic congruency of multiple demands).* Thesis presented for Doctor's degree in Psychology by the University of Brasília.

Günther, H. (2006). Pesquisa qualitativa *versus* pesquisa quantitativa: Esta é a questão? *Série: Textos de Psicologia Ambiental,* Nº 07. Brasília, DF: UnB, Laboratório de Psicologia Ambiental.

Hedler, H. C. (2007). *Meta-avaliação em de Auditorias de Natureza Operacional do Tribunal de Contas da União*: Um estudo sobre auditorias de Programas Sociais. Thesis presented for Doctor's degree in Psychology by the University of Brasilia.

Hunter, J. E., & Schmidt, F. L. (1996). Measurement error in psychological research: lessons from 26 research scenarios. *Psychological Methods, 1*(2), 199-223

Hunter, J. E., & Schmidt, F. L. (1999). Comparison of three meta-analysis revisited: An analysis of Johnson, Lullen, and Salas *(1995). Journal of Applied Psychology, 84*(1), 144-148.

Mendes, A. M. B., Tamayo, A., Paz, M. G. T., Neiva, E. R., Tamayo, N., Silva, P. T., Souza, A. C., Martins, A. J., & David, R. G. (1999). *Análise da cultura organizacional do Tribunal de Contas da União – TCU.* Brasília: Relatório Final. O&T Consultoria/FINATEC/Unb.

Oskamp, S. (1984). *Applied social psychology.* New Jersey: Prentice Hall.

Patton, M. Q. (2001). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Posavac, E. J., & Carey, R. G. (2003). *Program evaluation*. Methods and case Studies (6th ed.). New Jersey: Prentice Hall.

Rossi, P. H., & Freeman, H., E. (1993). Evaluation: A systematic approach (5th ed.). Thousand Oaks, CA: Sage.

Schwandt, T. A. (1989). The politics of verifying trustworthiness in evaluation auditing. *American Journal of Evaluation*, *10*, 33-40.

Silva, P. L. B. (2002). *A avaliação de programas públicos:* Reflexões sobre a experiência brasileira**.** Relatório técnico. Brasília: IPEA.

Smith, P. B., & Bond, M. H. (1999). *Social psychology across cultures*. Allyn e Bacon: EUA.

Stufflebeam, D. L., & Shinkfield, A. J. (1987). *Evaluación sistemática*: guía teórica y práctica. Barcelona: Ediciones Paidós Ibérica.

Woodside, A. G., & Sakai, M. Y. (2001). Metaevaluation of performance audits of government tourism-marketing programs. *Journal of Travel Research*; *39*, 369. [On-line], Obtido em 18 de fevereiro de 2007, de http://jtr.sagepub.com/cgi/content/abstract/39/4/369

Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (2004). *Avaliação de programas*: concepções e práticas. São Paulo: Editora Gente.