

<http://dx.doi.org/10.18222/eeae.v30i73.5879>

“HOW ‘GOOD’ IS GOOD ENOUGH?”¹ – DEFINING EVALUATIVE WORTH AND MERIT CRITERIA²

ADRIANA BAUER^I

TRANSLATED BY: Fernando Effori de Mello^{II}

ABSTRACT

This text discusses concepts associated with the notion of value and the action of valuing in program evaluation, considering the need to make judgments about the quality of a policy, program, project or social action that is inherent to it. It seeks to contribute to foment the discussion about valuation in the evaluation of social programs. To that end, it reviews the US and Brazilian literature in order to identify how valuation questions have been theoretically proposed and made explicit in evaluative research. Reports and articles on program evaluation with a focus on education are examined. It concludes, based on this bibliographical analysis, that the discussion about the attribution of value has seldom been conducted in the national literature in an explicit way, which reinforces the need for improving it.

KEYWORDS PROGRAM EVALUATION • EVALUATION CRITERIA • MERIT • WORTH.

¹ This suggestive question is inspired by the questioning made by Cousins and Shulha (2008) in their study about the complexities inherent in choosing criteria and valuation standards.

² I dedicate this text to Chris Coryn for introducing me, during his classes at the Evaluation Center, to discussions about valuation, as well as evaluation theories, thus providing me with intensive learning about the evaluation of programs; to Nelson Antonio Simão Gimenes, my colleague at the Carlos Chagas Foundation who, after listening for years to my discussions about merit and value in a course, suggested that I systematize them in the form of a text; and to Paulo de Martino Jannuzzi, whose prolific intellectual output in the field of evaluation, monitoring and social indicators has shed light on my studies and inspired my thoughts about important problems in the field.

^I The Carlos Chagas Foundation (FCC) and the University of São Paulo (USP), São Paulo-SP, Brazil; <http://orcid.org/0000-0002-5942-918>; adbauer@fcc.org.br

^{II} Freelance translator, São Paulo-SP, Brasil; feffori@gmail.com

“QUÃO ‘BOM’ É SUFICIENTE?” DEFINIÇÃO DE CRITÉRIOS AVALIATIVOS DE VALOR E MÉRITO

RESUMO

Este texto procura discutir conceitos associados à noção de valor e à ação de valorar na avaliação de programas, tendo em vista a necessidade de fazer julgamentos acerca da qualidade de uma política, programa, projeto ou ação social que lhe é inerente. Busca-se contribuir para fomentar a discussão a respeito da prática de valoração nas avaliações de programas sociais. Para isso, é feita uma revisão das literaturas estadunidense e brasileira, a fim de identificar como as questões de valoração têm sido propostas teoricamente e explicitadas em pesquisas avaliativas. São examinados, de forma exploratória, relatórios de avaliação e artigos sobre avaliações de programas, com foco na área educacional. Conclui-se, com base na análise bibliográfica realizada, que a discussão sobre a atribuição de valor raramente tem sido feita na literatura nacional de forma explícita, ressaltando-se a necessidade de seu aprimoramento.

PALAVRAS-CHAVE AVALIAÇÃO DE PROGRAMAS • CRITÉRIOS DE AVALIAÇÃO • MÉRITO • VALOR.

“¿CUÁN ‘BUENO’ ES SUFICIENTE?” DEFINICIÓN DE CRITERIOS EVALUATIVOS DE VALOR Y MÉRITO

RESUMEN

Este texto busca discutir conceptos asociados a la noción de valor y a la acción de valorar en la evaluación de programas, teniendo en cuenta la necesidad de juzgar acerca de la calidad de una política, programa, proyecto o acción social que le es inherente. Se busca contribuir a fomentar la discusión acerca de la práctica de valoración en las evaluaciones de programas sociales. Para ello, se hace una revisión de la literatura estadunidense y brasileña, a fin de identificar cómo las cuestiones de valoración han sido propuestas teóricamente y explicitadas en investigaciones de evaluación. Se examinan, de forma exploratoria, informes de evaluación y artículos sobre evaluaciones de programas, con enfoque en el área educativa. Se concluye, en base al análisis bibliográfico realizado, que la discusión sobre la atribución de valor se realiza raras veces en la literatura nacional e internacional de forma explícita, lo que resalta la necesidad de su perfeccionamiento.

PALABRAS CLAVE EVALUACIÓN DE PROGRAMAS • CRITERIOS DE EVALUACIÓN • MÉRITO • VALOR.

INTRODUCTION

Bad is bad and good is good and it is the job of evaluators to decide which is which. (SCRIVEN, 1986, p. 19)

The controversial citation that starts this paper, taken from a text by Michael Scriven (1986), refers to one of the integral aspects of the evaluation of policies, programs, projects and social actions: the attribution of value to the evaluated entity,³ based on which decisions are (or should be) made.

The valuation aspect seems inseparable from the evaluation of programs, as can be seen even in the definition of what this type of evaluation is, from the perspective of several authors. In general, *evaluation* is said to be:

[...] determining the worth or merit of an evaluation object (whatever is valued). More broadly, evaluation

3 That entity can be an object, a program, a project, an action, or a person. In English, there is the term "evaluand", coined by Scriven, addressed to this entity. In Portuguese, there isn't a word equivalent in meaning with evaluand. Consequently, the word "entity" will be used in the sense of evaluand.

is identification, clarification and application of defensible criteria to determine an evaluation object's value (worth or merit) in relation to those criteria. (WORTHEN; SANDERS; FITZPATRICK, 2004, p. 35)

The process of determining the merit, worth or value of something, or the product of that process. (SCRIVEN, 1991, p. 139)

[...] making a value judgment about an intervention or any of its components in order to assist in decision-making. This judgment can arise from the application of criteria and norms (normative evaluation) or from a scientific procedure (evaluative research). (CONTANDRIOPOULOS *et al.*, 1997, p. 31)

While there is some agreement on the idea that merit and value are basic features of program evaluation, it is necessary to define what each of these concepts means and what relationships are woven between them, as well as their differences.

Firstly, it is worth noting that according to Guba and Lincoln (1980, p. 61), the distinction between merit and worth is arbitrary, since both are constituents of the notion of value.⁴ To Davidson (2005, p. 247), in turn, “merit is the absolute or relative quality of something, either globally or in relation to a particular criterion”, and she makes no distinction between merit and value.⁵

In any case, merit has been understood as the *intrinsic value* of something, and its judgment may vary according to the different judges participating in the evaluation process, since different actors tend to consider different merit indicators (GUBA; LINCOLN, 1980).

⁴ The authors explain their theoretical choices and analyze their proposal in relation to the postulates of Scriven, another important American theorist of evaluation programs, who also recognizes that the notion of value contains more than one dimension. It is not the purpose of this study to delve further into these authors' semantic-conceptual discussion, though the reader can find it in Guba and Lincoln (1980, p. 61).

⁵ This text does not address methodological aspects for determining merit or value, although I recognize that the methodological designs used in evaluations depend on the public values underlying them (JANNUZZI, 2016), as well as on the personal values and interests of those involved (VIANNA, 2005). For the purpose of this paper, it will suffice to say that, according to Davidson (2005), a judgment of merit requires: collecting relevant data that allow the evaluator to judge what is evaluated in a given dimension of quality; and the synthesis of the performances achieved by the object, in case of a multicriteria dimension.

In turn, worth has been understood as the value of an evaluated entity for a given context, i.e., its *extrinsic value*. According to Mathison (2005), determining value requires a complete understanding of the context in which the program, project, or action is deployed, as well as knowledge of the qualities and attributes of what is being evaluated.

In general terms:

- merit: the quality intrinsic in the evaluated program or object;
- worth: extrinsic quality of the program or object evaluated in relation to the context for which it was proposed, how it meets the real needs of its target audience, among other elements external to the proposal that should be considered in its evaluation;
- significance: it can be understood as the importance of the evaluated program/project/object, i.e., the general conclusion on its merit and value after all relevant considerations have been processed.

One can conclude from the above that there is variation in the bases of judgment both by merit and by value and significance, although the variability is greater in the latter, given the contextual character of value and of attribution of significance. In the words of Guba and Lincoln (1980, p. 64):

[...] that variation arises from the fact that judgments of merit are tied to intrinsic and therefore relatively stable characteristics of the entity itself, while judgments of value depend upon the interaction of the entity with some context and thus vary as contexts vary. [...] If, on the other hand, one is talking about the different dimensions along which merit and worth may vary, one needs to consider time, degree of consensus, and boundary factors. Assessments of merit are made in terms of criteria which are relatively invariant over time, while worth assessments are made in terms of criteria that may alter rapidly with changing social, economic, or other short-term conditions. Moreover, there is likely to be a relatively high degree of consensus about merit criteria, while worth criteria depend, among other things, on value sets likely to be very different both between and within social and organizational groupings.

It should be noted, however, that not all three dimensions occur concomitantly. A program can have merit but no worth or significance in a given context. If it fails to meet the demands of its stakeholders,⁶ then it is said to have no worth. In addition, any program (or policy) without merit is also worthless, because if it fails to do well what it has set out to do, then it cannot be used to meet the needs of its stakeholders. Thus, a poorly designed program has neither merit nor value.

It would therefore be the evaluator's job to clearly determine what he intends to evaluate (object), as well as the quality dimensions related to that object, in order to define possible merit and value criteria for the program. However, some authors differ from this idea. According to Schwandt (2005, p. 443-444), many evaluators disagree that assigning merit, value or significance is inherent in the evaluation of social programs, as they argue that the evaluator's role is rather to describe and explain reality without judgment.

The present text argues that the debate about the establishment of merit, value and significance criteria for a program is fundamental to ensure public transparency and ethics in the evaluation processes, although the various authors who have dedicated to developing theorization about program evaluation have different positions on the issue.

Shadish, Cook and Leviton (1991) dedicated to analyzing how the valuation issue is understood and dealt with in the thought of theorists at different times in the recent history of program evaluation. In their book *Foundations of Program Evaluation: Theories of Practice*, they analyze the studies of seven authors⁷ which they consider to be representatives of three different stages of evaluative theories: stage 1, represented by Campbell and Scriven, whose main concern would be to shed light on the solution of social problems; stage 2, with thinkers such as Weiss, Wholey and Stake, who are said to be committed to generating alternatives to program evaluation, thus emphasizing the use of results and pragmatism; and stage 3, with Cronbach and Rossi, centered on theories that try to integrate "the past", considering the main concerns explored in the previous stages.

In their analysis of the selected works and theorists, Shadish, Cook, and Leviton (1991) considered five different dimensions: knowledge construction (how to construct valid knowledge about the evaluated object); valuation (how

⁶ I.e., a program's target audience, those who propose a program/policy, those who execute it, and the society in general.

⁷ These are: Michael Scriven, Donald Campbell, Carol Weiss, Joseph Wholey, Robert Stake, Lee Cronbach e Peter Rossi.

to assign value to the results of evaluations); social program (understanding theorists' positions on how programs change); use of results (their indications on how to use the results in the policy process); and finally, their positions on how to organize the evaluative practice.

This paper is interested in exploring these authors' considerations on the dimension of valuation, which, according to them, essentially discusses the role played by values and valuing processes from each theoretical perspective. Its central discussion is the ways in which values can be associated with the description of the programs. The authors argue that values are ubiquitous in program evaluation and, in general, theorists who are concerned with the issue believe it is important to answer whether the object or entity evaluated is good/adequate, making evident the notion of "good" that is being considered, as well as the reasons for that conclusion. As a component of an evaluation theory, the valuation dimension is concerned with discussing "how evaluators can make value problems explicit, deal with them openly, and produce an analysis that is sensitive to the implications of the values of programs" (SHADISH; COOK; LEVITON, 1991, p. 47). The authors rely on Beauchamp to argue that valuation theories can be metatheories, prescriptive theories or descriptive theories, which they define as follows:

Metatheory: the study of the nature of and justification for valuing;

Prescriptive theory: theory that advocates the primacy of particular values; Descriptive theory: theory that describes values without advocating one as best (SHADISH; COOK; LEVITON, 1991, p. 48)

To exemplify their positions, the authors argue that Michael Scriven is the only theorist who has a metatheory of valuation, while Ernest House proposed a prescriptive theory, advocating the need for a theory of justice to underlie the analysis of evaluation results. They also argue that Joseph Wholey proposes a descriptive theory as he argues that the evaluator's job is just to try to understand and describe the values of those who propose each policy/program or social action. The authors emphasize that theorists who take on a prescriptive theory have a critical perspective and an "intellectual authority" that are not consistent with descriptive theories, since they are confronted with the need to reflect on what is "good" for the human condition, for the public good in broader terms, rather than just within the scope of the program (SHADISH; COOK; LEVITON, 1991, p. 49).

To the authors of *Foundations of Program Evaluation*, evaluation theorists of stage 1, such as Scriven, tend to define evaluation criteria based on social justice values, so it is fair to assume that they have prescriptive valuation theories as their horizon. In turn, second-generation evaluation authors such as Carol Weiss and Robert Stake would recognize the plural nature of values, admitting that the intentional users of evaluation results may contribute in constructing the evaluation criteria from a descriptive perspective of valuation; finally, authors representing the third generation of evaluation, such as Cronbach, Rossi and Wholey, would argue that evaluation criteria should be established by the program beneficiaries, also from a descriptive perspective.

Although they have built a consistent valuation theory and argued that concerns with how to make judgments about evaluated programs should be part of valuation theorization, Shadish, Cook and Leviton recognize that this aspect has not received the necessary attention, neither in theoretical nor in practical terms, at least not explicitly or systematically in the works they selected. In their words:

Yet evaluators acknowledge that values deserve more attention (House, 1980). Nearly all the theorists in this book agree that evaluation is about determining value, merit, or worth, not just about describing programs. (SHADISH; COOK; LEVITON, 1991, p. 49)

This scenario unveiled by the authors indicates a necessity to rediscuss the meanings and purposes of program evaluation itself, a field that has expanded in recent years but needs to be continually under discussion, with a view to its theoretical and practical improvement.

PROGRAM EVALUATION VS. EVALUATIVE RESEARCH VS. POLICY ANALYSIS

As explained in the beginning of this paper, the evaluation process of a program or policy determines, or should determine, criteria to judge the value or merit of an evaluated object based on data collected through empirical investigation carried out with the application of techniques of Social Sciences.

However, not every appreciation of a policy or program contains a value judgment. As seen earlier, researchers disagree whether all evaluation should lead to a judgment. But there are also studies that are not dedicated

to discussing the merit or value of its object since they conduct their analysis from a different theoretical perspective. Thus, before we proceed with the discussion on valuation theories, it is necessary to distinguish between program evaluation and policy analysis, since both research tools for public programs differ in several aspects, including with regard to the issue of value judgment.

In the present study, the determination of merit or value is assumed to be an essential task that differentiates evaluation from other types of research or analysis to which policies are submitted, regardless of the evaluation purposes – whether summative or formative –, or from more traditional or collaborative approaches (DAVIDSON, 2005; COUSINS; SHULHA, 2008).

As explored in a previous study (BAUER, 2011), the distinctive element between evaluation and research is precisely the value judgment on a given object or program, which is inherent in the former. Although research can produce evidence on the value or merit of a given object of study, this production is understood to be indirect, secondary, since the focus of research is the generation of new knowledge for the field. In turn, program evaluation focuses on the search for evidence that allows a judgment to be made. That very concern is also central in evaluative research.⁸

Authors seem to agree that the product of evaluation and evaluative research is necessarily a value judgment to be used in decision making.

This issue appears differently when the analyst resorts to policy analysis.

It is worth to distinguish between evaluation and policy analysis. While policy analysis would be primarily concerned with identifying trends and seeking generalizable principles, program evaluation is aimed at providing information to decision-making about the quality of programs, with value judgments being inherent in it. To Figueiredo and Figueiredo (1986, p. 107), policy analysis is concerned with studying how decisions are made and

8 Here, it is worth opening a parenthesis to deal with the existing distinction in the literature between evaluation and evaluative research, a discussion that dates to the 1970's and today seems to have been overcome. Modern program evaluation was disseminated in the 1960's in the United States, focusing on educational programs. The expansion of and the efforts for this type of evaluation were not necessarily accompanied by theoretical and methodological rigor. Concerned with the quality of the evaluations that were being produced, and aware of the influence of their results on American social policy, Suchman (1971) proposed the distinction between evaluation and evaluative research: "first, we distinguish between evaluation as the general process of making judgments of worth regardless of the basis for such judgments, and evaluative research as referring to the use of the scientific method for collecting data concerning the degree to which some specified activity achieves some desired effect." (SUCHMAN, 1971, p. 45). This distinction was incorporated by other authors. Worthen, Sanders, and Fitzpatrick (2004, p. 36), for example, define evaluative research as "any evaluation that employs rigorous methodology from the field of social sciences." In the Brazilian literature, there is reference to *cânones científicos mínimos* (minimum scientific canons) also in the writings of Draibe (2001, p. 18). Currently, the distinction between evaluation and evaluative research has been less emphasized in the literature, with a tendency to incorporate the idea of methodological rigor into the concept of evaluation.

identifying “what factors can influence the decision-making process, as well as the characteristics of that process”. Muller and Surel (2002) point out that, if policy analysis initially has had any intent to make value assertions on policies and programs, that field of study was redirected to more complex analyzes focused on understanding governmental action (or the lack thereof):

Originated in the United States as a set of research devices whose ambition was to provide recipes of “good” government, policy analysis has progressively moved away from its operational orientation to become an actual discipline of Political Science, progressively autonomous in its teaching and research structures. Gradually fostering another view of the state, the analysis of public policies has since contributed to show that the state is not (is no longer?) that absolute social form in history.

[...]

The analyst’s work must therefore simultaneously consider the intentions of decision-makers, even if they are confused, and the meaning-building processes in practice throughout government action’s development phase. However, in every case, the researcher must take care not to take over the place of policy actors in determining the meaning of the policy. (MULLER; SUREL, 2002, p. 8, 20-21)

In Brazil, one can find a possible definition of public policy analysis in the work of Marta Arretche (2009, p. 30):

Any public policy can be designed and implemented in a number of ways [...] Public policy analysis aims to reconstruct these various characteristics so as to apprehend them in a coherent, comprehensible whole. Or rather to give meaning and understanding to the erratic character of governmental action.⁹

⁹ In the original: “Qualquer política pública pode ser formulada e implementada de diversos modos [...] A análise de políticas públicas busca reconstruir estas diversas características, de forma a apreendê-las em um todo coerente e compreensível. Ou, melhor dizendo, para dar sentido e entendimento ao caráter errático da ação pública”.

The author's statements, which allow us to differentiate the scope of policy analysis from that of public policy evaluation, allow reaffirming considerations made elsewhere:

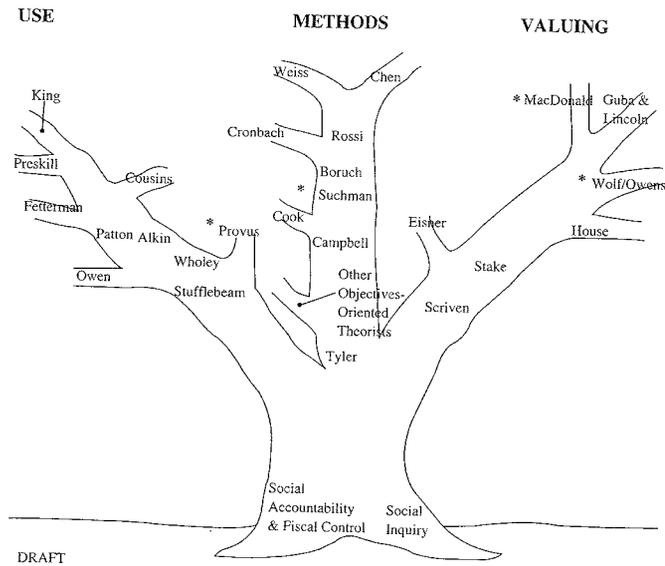
Whether under the aegis of evaluation or evaluative research, the field has developed extremely fast, and the search for improvement and methodological rigor has not yet come to an end. There are several theoretical approaches and movements in evaluation, and many authors are seeking to categorize trends, currents, or generations in evaluation.¹⁰ (BAUER, 2011, p. 10)

To illustrate this diversity, it is worth mentioning the work of Madaus, Scriven and Stufflebeam (1983), which describes the evolution of evaluation movements in five main periods: the pre-Tylerian period (before 1930); the Tylerian era (1930-1945); the age of innocence (1946-1957); the age of realism (1958-1973); and the age of professionalism, from 1973 onwards. With regard to evaluative approaches, Stufflebeam and Shinkfield (2007, p.145) classify and analyze 26 evaluations, also denominating some of them as pseudo-assessments when they “fail to produce and report on valid evaluations of merit and value” to the audiences who would be interested in those outcomes.

The authors provide the graphic representation of a genealogical tree of evaluation (Figure 1), emphasizing that the consolidation and universalization of the field of evaluation was based on a two-fold foundation of accountability and social inquiry, both situated in the trunk of the tree. From this common trunk, the authors organize the contributions of theorists in three branches which indicate emphases or concerns of different natures: the theorists who dealt with the discussion of evaluation methods; those who focused on the uses of the outcomes of the information obtained through evaluation; and those who discuss the question of value and merit in evaluation. The thoughts of some of these authors regarding the valuing issue are explored in the next section.

10 In the original: “Quer seja sob a égide da avaliação ou da pesquisa avaliativa, o campo se desenvolveu vertiginosamente, e a busca pelo aprimoramento e rigor metodológico ainda não chegou ao fim. Várias são as abordagens teóricas e os movimentos em avaliação, e muitos são os autores que buscam categorizar as tendências, correntes ou gerações em avaliação”.

FIGURE 1 - Theoretical Tree of Evaluation Revisited



Source: Alkin and Christie (2004, p. 389).

INTERNATIONAL THEORETICAL CONTRIBUTIONS TO THE VALUING ISSUE

Michael Scriven is undoubtedly the first theorist to express his concern about the issue of valuing and making value judgments, thus assuming an important role in theorizing about these aspects. To him, every evaluation process should follow a logic defined in four steps: selecting merit criteria; setting performance standards; measuring the performance; and summarizing the outcomes in a statement of value. In the first stage, the author argues that the needs of society should be considered.

Evaluation is not merely the application of social science methods to solve social problems; rather, evaluators use social science methods to examine the merit, worth and significance of a program or policy for the purposes of describing values associated with different stakeholder groups, as well as reaching evaluative conclusions about good and bad solutions to social problems. (SCRIVEN, 2003, p. 21)

Scriven strongly criticizes what he perceives as the evaluators' tendency, especially in some approaches, to not assume the valuing perspective of their work. His annoyance is understandable when we observe that evaluation is an instrument of guarantee of social rights, and it's the evaluator's job to protect them. The evaluator is therefore a "public protector" in his area (SCRIVEN, 1971, p. 53). Consequently, those who perform evaluation or evaluative research cannot escape the task of judging whether a program works or not, whether it is good or bad, whether it achieves its goals or not.

In addition, Scriven suggests that the evaluator should seek a synthesis, a final judgment, even when the evaluation of a program, policy or action presents multiple results. According to Alkin and Christie (2004), he is the only author who advocates the need for this synthesis in terms of quality judgment of the assessed entity.

However, even though the evaluator plays a very important role in Scriven's theorizing as he must determine criteria of merit and value and quality standards in the analysis of the results obtained (SCRIVEN, 1976), the author is careful to point out that the decision on which criteria and standards to choose is not for the evaluator to make exclusively.

It is his job to discuss values extensively with specialists, relying on empirical facts, in order to determine the values that are fundamental. To the author, such caution would validate the valuing assertions and considerations (SHADISH; COOK; LEVITON, 1991; ALKIN; CHRISTIE, 2004). It is worth noting that Scriven gives more weight to the opinion of experts on the values that should be considered in evaluation than to those involved in proposing the program.¹¹

Elliot Eisner, however, assumes a position quite different from Scriven's, although both attach central importance to the valuing issue in their theorizations.

To Eisner (1998, p. 80 *apud* ALKIN; CHRISTIE, 2004, p. 34-36), evaluation "is concerned with making value judgments about the quality of some object, situation, or process" (EISNER, 1998, p. 80). The role of the evaluator, according to him, is very similar to that of the art critic. Based on his knowledge (connoisseurship), the evaluator selects the central aspects that determine the quality of what is evaluated. Criticism and connoisseurship are the central aspects of the author's theorizing, would be the attempt to

¹¹ Within the limits of this text, our focus is not to discuss positive and negative aspects of each approach. The reader who wants to delve into the subject can refer to Shadish, Cook and Leviton (1991) and Alkin and Christie (2004).

answer three essential questions: what is happening, what is the meaning of what is happening, and what is the value of what is happening. His proposal is therefore based on qualitative approaches, with observation having a critical role in making value judgments (ALKIN; CHRISTIE, 2004; MATHISON, 2005).

Ernest House (2004) also dedicates attention to the value component in his considerations about evaluation. To him, evaluators should focus on making, based on facts and values, the evaluation conclusions:

Evaluators should focus on drawing conclusions, a process that involves employing appropriate methods as a component, but also finding the right comparison standards and criteria and approaching them in data collection, analysis, and inference of conclusions. (HOUSE, 2004, p. 219)

The author argues that the principles of Rawls' theory of justice¹² should be considered in the evaluation process and should find judgments that economically and politically prioritize the interests of the underprivileged.

Assuming that there is no value neutrality in evaluation, nor should there be such neutrality, House (2004) accepts that the evaluator takes on a favorable position on the reduction of social injustices. That does not mean that those interested in the evaluation should be disregarded. For him, both the nature of the evaluated entity and the intuitions of those involved with what is being evaluated should be respected in defining the evaluation criteria and the data to be collected.

With a different approach from the authors mentioned above, Robert Stake (2004) believes that there is not one single, true or best value based on which a program can be judged, since the distinct programs can have different values for different people. According to Shadish, Cook and Leviton (1991), Stake considers that the values of those interested in the evaluation should be considered by the evaluator when making judging the results, and should be reflected in them. Thus, his theorization admits a diversity of perspectives among those involved in the program. In his words:

¹² John Rawls, in his book *A Theory of Justice*, originally published in 1971, defends the theory of justice as fairness and establishes two principles of justice for institutions, distinguishing them from the principles applied to individuals and their actions, which he considers fundamental: the principles of freedom and equal opportunities. It is not the purpose of this work to look further into the discussion about Rawlsian theory, only contextualize House's discussion.

[...] directed toward discovering merit and gaps in the program, the responsive evaluation recognizes multiple sources of valuing as well as multiple foundations. It is respectful of multiple and sometimes contradictory standards kept by different individuals and groups, with a reluctance to press for consensus. (STAKE, 2004, p. 210)

In the 1980's, Wolf and Owens developed an evaluative approach that clearly followed a valuing orientation. Their concern was to minimize the possibility of involuntary bias by the evaluators, whether in defining the objectives to be evaluated or choosing the basis for the evaluation of results. As in a trial, two or three groups of evaluators are called upon to take opposite positions on the program (the contraposition method). The attention the authors dedicated to how the trial would be conducted enabled their classification by Alkin and Christie in this branch of the evaluation. Considering that the goal of a trial is the public interest, one can assume that the discussions between the opposing groups about the results obtained – the listing of aspects that are favorable and contrary to the program – enable adjusting the valuing bases without involving, however, a decision by the evaluator about which values should be considered within the program.

In turn, to Barry MacDonald (1995), the diverse stakeholders in the program have different visions about its objectives, as well as its quality dimensions and levels. According to them, the evaluator's role is that of a negotiator. The evaluator's main activity, as far as valuation is concerned, is thus described by the author: "Rather he will collect and communicate the alternative definitions, perspectives and judgement held by people in and around the program". (MACDONALD *et al.*, 1979, p. 127-128 *apud* ALKIN; CHRISTIE, 2004, p. 39). Although he argues that specialists should conduct democratic evaluations, the author recognizes that the contexts for evaluating are diverse and the evaluator may be confronted with the need to conduct bureaucratic and autocratic evaluation proposals. Even in these contexts, the evaluator's role would be to represent someone else's value and accept it. That does not mean that the author does not recognize that the specialist has his own values; to the contrary, he considers that the evaluator has a political role and that his values are always present in the evaluation process, from the choice of the method (MACDONALD, 1995).

Table 1 presents a classification of the evaluative proposals of the aforementioned authors in relation to the type of value theory that is believed to underlie their thinking.

TABLE 1 - Classification of authors according to valuing theories presented by Shadish, Cook and Leviton (1991)

AUTHORS	EVALUATIVE PROPOSAL	TYPE OF VALUE THEORY
Scriven	<i>Goal-free evaluation</i>	Meta-theory and prescriptive theory
Guba e Lincoln	<i>Constructivist evaluation</i>	Descriptive
Stake	<i>Responsive evaluation</i>	Descriptive
House	<i>Deliberative democratic evaluation</i>	Prescriptive
Eisner	<i>Connoisseurship</i>	Descriptive
Wolf e Owens	<i>Adversary Evaluation</i>	Descriptive
MacDonald	-	Descriptive

Source: Alkin e Christie (2004); Shadish, Cook e Leviton (1991); Mathison (2005).

In addition to the authors highlighted, we can find initiatives of evaluation practitioners and of multilateral organizations which indicate desirable aspects, values and criteria to be considered in evaluation practice.

An example of the above can be found in the document titled *Manual de planificación, seguimiento y evaluación de los resultados de desarrollo*, of the United Nations Development Programme (UNDP) (PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO, 2009). Among the various elements discussed, there is a clear position on criteria commonly included in UNDP assessments: relevance, effectiveness, efficiency, sustainability and impact of development efforts, the importance of which is justified in the document. A summarized explanation about these criteria can be found in Jannuzzi (2018, p. 16):

In a simplified way, the Relevance of a program or project is associated with the degree of its pertinence to the public demands' priority, that is, its adherence to the agenda of political priorities in a given society. The Effectiveness of a program or project is an attribute related to the fulfillment degree of the program objectives or the degree to which it meets the demand that motivated it. Efficiency, on the other hand, is associated with the quality of and concern with how resources are used in producing program outputs.

Sustainability refers to the ability of the program or project to generate permanent changes in the reality on which it acted.

Finally, Impact refers to the medium and long term effects on beneficiaries and society – whether positive or not – which are directly or indirectly attributable to the program or project. In general, when considering the positive impacts, we alternatively use the term Effectiveness.¹³

In addition to clearly suggesting a pragmatic valuing basis for the programs, the document seems to be consistent with the prescriptive perspective of valuation as it argues that other principles, such as respect for human rights and equity, reflected in the document through its concern about gender issues, should be considered in UNDP evaluations, as illustrated in the following excerpt:

En línea con los esfuerzos de desarrollo del PNUD, las evaluaciones de la organización están guiadas por los principios de igualdad de género, un enfoque basado en los derechos y el desarrollo humano. Por tanto, y según corresponda, las evaluaciones del PNUD valorarán el grado en el que las iniciativas de la organización han tratado los temas de inclusión social y de género, de igualdad y de empoderamiento; han contribuido a fortalecer la aplicación de estos principios a los diferentes esfuerzos de desarrollo en un país dado; y han incorporado en el diseño de la iniciativa el compromiso del PNUD con los enfoques basados en los derechos y la problemática de género. (PNUD, 2009, p. 169-170, emphasis added)

13 In the original: "De forma simplificada, Relevância de um programa ou projeto está associado ao grau de pertinência do mesmo às demandas públicas prioritárias, isto é, sua aderência à agenda de prioridades políticas de uma determinada sociedade. Eficácia de um programa ou projeto é um atributo relacionado ao grau de cumprimento dos objetivos do mesmo ou de atendimento da demanda motivadora do programa. Eficiência, por outro lado, está associado à qualidade e preocupação de como os recursos são utilizados na produção dos resultados do programa. A Sustentabilidade refere-se à capacidade do programa ou projeto em gerar mudanças permanentes na realidade em que atuaram. Por fim, o Impacto diz respeito aos efeitos de médio e longo prazo sobre os beneficiários e sociedade – positivos ou não- direta ou indiretamente atribuíveis ao programa ou projeto. Em geral, quando considerados os impactos positivos, emprega-se alternativamente o termo Efetividade".

These are positions of some field theorists and institutions that have studied the subject. From the above, one can see that the valuing issue gains attention at different points in the program evaluation history and, in most cases, it emerges from authors' evaluation practice, which confronts them with ethical dilemmas and the need to take a position in facing the challenge of laying the bases for judging the programs. The next section is dedicated to exploring how such issues are handled by national authors and evaluators. To that end, a non-exhaustive bibliographical review was conducted in theses, dissertations, articles and evaluation reports dealing with educational programs and policies, which are presented in the next section.

DISCUSSING VALUING IN PROGRAM EVALUATIONS IN THE NATIONAL TERRITORY

In Brazil, there are few studies focusing on the valuing issue inherent in the evaluation of educational programs, policies and projects. In general, researchers and evaluators do not clearly adopt or propose bases for judgments about the evaluated proposals, except for the thoughts of Heraldo Vianna (2005) and Paulo Jannuzzi (2016).

In discussing the collective nature of program evaluations, Vianna (2005) recognizes the diversity of interests, views and values that are inherent to it, and he argues, based on House and Howe (2003), that these aspects should be debated democratically in order to reach sensible and appropriate judgments. To the author, the values to be considered are the ones that matter to the program, although the search for these values, from which judgments about the program and its quality will be made, is a complex task that does not always result in consensus. In referring specifically to the school programs evaluation, Vianna (2005, p. 51) points out that “values, beliefs and multiple social demands must be taken into account in relation to the various types of learning and the different situations at the school”. One can therefore infer that his source for establishing the bases for judging value and merit are the various persons involved in its achievement.

Paulo Jannuzzi (2016) assumes that three contending political and ideological conceptions underlie the proposed program evaluations in Brazil: economic efficiency, procedural efficacy and social effectiveness, all of which influence the theoretical and methodological choices and public values that are dealt with by those who engage in public policy evaluation. These values, which the author identifies as “quality of public spending”, “procedural

compliance” and “improvement in search of the greatest social impact”, are based on different conceptions of justice and offer different bases for proposing value criteria and merit on what is being evaluated. The author advocates that social effectiveness and, therefore, greater social impact should be the guiding elements of program evaluators, since they are closely related to the principles of justice, equality and social well-being that are at the core of the Federal Constitution of 1988. Jannuzzi explores, with a clarity that is rare in the Brazilian literature on program evaluation, how republican public values influence (or should influence) merit and value criteria used in judging programs and projects, which are consistent with the prescriptive theories of value explored earlier.

The interest in understanding how Brazilian authors and researchers of evaluation have dealt with problems related to the valuing issue is old and dates to the time when I was writing my doctoral thesis (BAUER, 2011). At that time, the bibliographic survey conducted showed that Brazil was in the infancy with regard to program evaluations and their impacts, which was the main interest of that doctoral research. Thus, due to the theoretical-methodological paths I had chosen, no discussion was conducted about how the valuing issue was laid out in the national literature.

Within the limits of the present text, an attempt is made here to fill this gap. Although the author resumes and reports on her experience and decisions on valuation, other works are considered. In order to identify and select them, a decision was made to perform a broad search using Google Scholar. The decision is justified by the intention to identify, in a short time, works of different natures, which are usually indicated by this tool.

The selection was made through the use of descriptors combining different terms: evaluation of educational programs AND merit; evaluation of educational programs AND value; and evaluation of educational programs AND judgement. Several references were found which dealt with other areas (evaluation of health programs, computer science programs, etc.), or with other dimensions of evaluation (learning, institutional, higher education evaluation, etc.). Given the huge number of references, only those that, by reading the titles, were directly related the educational area and to evaluation of programs were selected. The second step was to consolidate the references in a single file, eliminating repetitions, since the same works appeared more than once, although different descriptors had been used. This first phase of selection resulted in a database with 85 references, including articles, evaluation reports, theses and dissertations.

Then, each of the references was located on the internet. When it was not possible to access the work electronically, it was discarded. The reading of the abstracts allowed a new selection, leaving a database with 32 studies. In these, an effort was made to determine whether the authors conducted their own discussion about the valuing issue (more explicitly or implicitly), or if it appeared reproduced by the review of the literature, without necessarily representing the author's thought, or, still, whether that discussion was absent.

A classification of the studies in prescriptive or descriptive was also tried. The authors included were those who sought evaluation parameters outside the program or action evaluated, whether in the relevant literature or through indication of the evaluation standards agreed in the framework of expert associations, such as the Joint Committee. Those who based their analysis of results on the opinion of stakeholders or on the indications found in documentation of the program itself were considered as assuming a descriptive theory. Finally, when the author's basis for judging was not clear, or the work did not involve value judgments, it was classified as "absent" or "not applicable".

TABLE 2 – Analysis of works by Brazilian authors based on valuation theories presented by Shadish, Cook and Leviton (1991)

PAPERS, THESIS, DISSERTATIONS AND EVALUATION REPORTS	DISCUSSION ON VALUATION	TYPE OF THEORY
ABRAMOWICZ, Mere. Avaliação, tomada de decisões e políticas: subsídios para um repensar. <i>Estudos em Avaliação Educacional</i> , São Paulo, n. 10, p. 81-101, jul./dez. 1994.	Explicit	Prescriptive
ANDRIOLA, Wagner B. Utilização do modelo CIPP na avaliação de programas sociais: o caso do Projeto Educando para a Liberdade da Secad/MEC. <i>REICE: Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación</i> , Madrid, v. 8, n. 4, p. 65-82, 2010.	Absent	Not Applicable
BAUER, Adriana. <i>Avaliação de impacto de formação docente e serviço: o programa Letra e Vida</i> . Tese (Doutorado) - Universidade de São Paulo, São Paulo, 2011.	Explicit	Prescriptive
BITTENCOURT, Jaqueline M. V. <i>Uma avaliação da efetividade do Programa de Alimentação Escolar no Município de Guaíba</i> . Dissertação (Mestrado) - UFRGS, Porto Alegre, 2007.	Implicit	Descriptive
BRANDÃO, Maria de F. R. Um modelo de avaliação de programa de inclusão digital e social. Tese (Doutorado) - UnB: Brasília, 2009.	Explicit	Descriptive
CAMPOS, Maria de F. H. <i>et al.</i> Avaliação de políticas e programas governamentais: experiências no mestrado profissional. <i>Avaliação: Revista Avaliação de Políticas Públicas</i> , v. 1, n. 1, p. 49-58, jan./jun. 2008.	Literature review	Not Applicable
CARNEIRO, Liliane B. <i>Características e avaliação de programas brasileiros de atendimento educacional ao superdotado</i> . 2015. 178 f. Tese (doutorado) - UnB, Brasília, 2015.	Implicit	Descriptive
CORREA, Vera L. de <i>Avaliação de programas educacionais: a experiência das escolas cooperativas em Maringá</i> . Tese (Doutorado) - FGV, Rio de Janeiro, 1993.	Literature review	Descriptive

(to be continued)

(continuation)

PAPERS, THESIS, DISSERTATIONS AND EVALUATION REPORTS	DISCUSSION ON VALUATION	TYPE OF THEORY
DAVOK, Delsi F. <i>et al. Modelo de meta-avaliação de processos de avaliação da qualidade de cursos de graduação. Tese (Doutorado)</i> – UFSC, Florianópolis, 2006.	Explicit	Prescriptive
DAVOK, Delsi F. Quality in education. <i>Avaliação: Revista da avaliação da Educação Superior</i> , Campinas, v. 12, n. 3, p. 505-513, set. 2007.	Explicit	Not Applicable
DUTRA, Marina L. da S. <i>Avaliação de treinamento: em busca de um modelo efetivo. Tese (Doutorado)</i> – FGV, Rio de Janeiro, 1979.	Implicit	Prescriptive
FERREIRA, Helder; CASSIOLATO, Martha; GONZALEZ, Roberto. <i>Uma experiência de desenvolvimento metodológico para avaliação de programas: o modelo lógico do programa segundo tempo. Brasília: Ipea</i> , 2009.	Implicit	Descriptive
FIRME, Thereza P. Os avanços da avaliação no século XXI. <i>Revista Educação Geográfica em Foco</i> , v. 1, n. 1, p. 1-4, jan./jul. 2017.	Explicit	Descriptive
GIMENES, Nelson A. S. Estudo metavalativo do processo de auto-avaliação em uma Instituição de Educação Superior no Brasil. <i>Estudos em Avaliação Educacional</i> , São Paulo, v. 18, n. 37, p. 217-243, maio/ago. 2007.	Implicit	Prescriptive
GOLDBERG, Maria A. A.; BARRETTO, Elba S. S.; MENEZES, Sônia M. C. Avaliação educacional e educação de adultos. <i>Cadernos de Pesquisa</i> , São Paulo, n. 8, p. 7-110, 1973.	Explicit	Prescriptive
MEDEIROS NO, Benedito. <i>Avaliação dos impactos dos processos de inclusão digital e informacional nos usuários de programas e projetos no Brasil. Tese (Doutorado)</i> – UnB, Brasília, 2012.	Absent	Descriptive
MIORANZA, Claudio. <i>Desenvolvimento e aplicação de modelo multidimensional para a avaliação da qualidade educacional no programa de Pós-Graduação Stricto Sensu do IPEN. Tese (Doutorado)</i> – Ipen/ USP, São Paulo, 2009.	Explicit	Prescriptive
ORLANDO FO, Ovidio; SÁ, Virgínio, I. M. Avaliação externa da gestão escolar do Programa Nova Escola do Estado do Rio de Janeiro: um estudo reflexivo sobre o seu primeiro ciclo de realização (2000-2003), passados quinze anos de sua implementação. <i>Ensaio: Avaliação e Políticas Públicas em Educação</i> , v. 24, n. 91, p. 275-307, abr./jun. 2016.	Explicit	Descriptive
PINTO, Edvan W. F. <i>Programa Nacional de Integração da Educação Profissional com a Educação Básica na Modalidade de Educação de Jovens e Adultos (PROEJA): uma avaliação de impactos nas condições de trabalho e renda dos egressos no município de Açailândia-MA. Dissertação (Mestrado)</i> – UFMA, São Luís, 2016.	Absent	Absent
RIBEIRO, Claudia. <i>Programa Alfabetizar com Sucesso – Programa de acompanhamento dos anos iniciais da rede pública de Pernambuco: a avaliação do município de Condado. Dissertação (Mestrado)</i> – UFJF, Juiz de Fora, 2015.	Absent	Absent
ROCHA, Ana R.; CAMPOS, Gilda H. B. Avaliação da qualidade de software educacional. <i>Em Aberto</i> , Brasília, v. 12, n. 57, p. 32-44, 1993.	Absent	Prescriptive
RODRIGUES, Rosângela S. <i>et al. Modelo de avaliação para cursos no ensino a distância: estrutura, aplicação e avaliação. Dissertação (Mestrado)</i> – UFSC, Florianópolis, 1998.	Literature review	Prescriptive
ROSEMBERG, Fúlvia. Políticas de educação infantil e avaliação. <i>Cadernos de Pesquisa</i> , São Paulo, v. 43, n. 148, p. 44-75, jan./abr. 2013.	Explicit	Prescriptive
SERPA, Selma M. H. C. <i>Para que avaliar? Identificando a tipologia, os propósitos e a utilização das avaliações de programas governamentais no Brasil. Dissertação (Mestrado)</i> – UnB, Brasília, 2010.	Literature review	Not Applicable

(to be continued)

(continuation)

PAPERS, THESIS, DISSERTATIONS AND EVALUATION REPORTS	DISCUSSION ON VALUATION	TYPE OF THEORY
SILVA, Danilma de M. <i>Desvelando o PRONATEC: uma avaliação política do programa</i> . Dissertação (Mestrado) – Universidade Federal do Rio Grande do Norte, Natal, 2015.	Literature review	Absent
SOMERA, Elizabeth A. S. Reflexões sobre vertentes da avaliação educacional. <i>Acesso do Averso</i> , Araçatuba, SP, v. 6, n. 6, p. 56-68, ago. 2008.	Literature review	Not Applicable
SOUSA, Clarilza P. Descrição de uma trajetória na/da avaliação educacional. <i>Ideias</i> , São Paulo, v. 30, p. 161-174, 1998.	Literature review	Not Applicable
VIANNA, Heraldo M. A prática da avaliação educacional: algumas colocações metodológicas. <i>Cadernos de Pesquisa</i> , São Paulo, n. 69, p. 40-47, maio 1989.	Literature review	Not Applicable
VIANNA, Heraldo M. Avaliação de Programas Educacionais: duas questões. <i>Estudos em Avaliação Educacional</i> , São Paulo, v. 16, n. 32, p. 43-56, jul./dez. 2005.	Explicit	Descriptive
VIANNA, Heraldo M. Avaliação educacional: problemas gerais e formação do avaliador. <i>Estudos em Avaliação Educacional</i> , São Paulo, v. 25, n. 60 (especial), p. 74-84, dez. 2014.	Bibliographic Review	Not Applicable
VIANNA, Heraldo M. Fundamentos de um programa de avaliação educacional. <i>Estudos em Avaliação Educacional</i> , São Paulo, n. 28, p. 23-38, jul./dez. 2003.	Bibliographic Review	Not Applicable
VIANNA, Heraldo M. Novos estudos em avaliação educacional. <i>Estudos em Avaliação Educacional</i> , São Paulo, n. 19, p. 77-169, jan./jun. 1999.	Bibliographic Review	Not Applicable

Source: Google Scholar and SciELO (Scientific Electronic Library Online). Prepared by the author.

The analysis of the papers shown in table reveals some aspects of interest. There are several studies that see themselves as evaluative and do not explicitly carry out a theoretical discussion about the value issue. Most of them simply review the literature, pointing out concepts and problems related to the valuing issue, yet without the author's explicit position on it.

About a third of the works implicitly or explicitly present a descriptive theory of the evaluation, representing a very similar scenario to that which US theorists point out in their analyzes. In general, these are evaluations that take qualitative approaches to evaluation, with an emphasis on case studies and opinion surveys. This result should be understood in relation to the theoretical tradition of Brazilian evaluation, which, especially in the 1980s, has highlighted the political dimension of evaluation and the existence of several values among those interested in evaluation processes. In any case, even this theoretical production, which influenced the evaluative proposals at the end of the 20th century, does not address directly (or in depth) the valuing issues. Thus, it was not possible to identify a Brazilian meta-theory about the valuing issue. The author who dedicates most attention to this aspect seems to be Heraldo Vianna, although his considerations are scattered in several of the articles that he wrote throughout his career.

Finally, and as an illustration of the decisions to be taken by those interested in conducting a program evaluation, below are explicitly presented,

based on the author's previous experience (BAUER, 2011), the decisions made on the occasion of an evaluation of the impact of a program for training literacy teachers, which certainly influenced the analysis and results achieved. It is not the purpose of this article to delve into the proposed methodological model,¹⁴ only discuss the choices that were made at the time of the study.

In order to analyze the impacts of this program, it was necessary to identify the variables that would be considered the desired results of the program. At the time it was carried out, it was argued that literacy should be a process for understanding and interacting with the world, making pupils able to access and use different language skills to solve real problems. This conception brought the need of a broader, more holistic process of literacy, and thus the criteria of merit for evaluating the Program were defined taking into account students' ability to demonstrate such language skills, even though that goal was not made explicit. In addition, since the program focused on teacher learning, it was considered necessary to evaluate what they had learned after participating in the program, considering the new theories of literacy and teaching methodology peculiar to the design of the program. In the production of the instruments of research and observation of the classroom environment, the evaluation addressed the expected practices of teachers participating in the course, according not only to the program materials, but also indicators of quality in education, with an emphasis on literacy, produced by Educative Action in partnership with MEC and the United Nations Children's Fund (Unicef).

These decisions were not based on methodological assumptions but referred to the definition of the program goals to be evaluated and the criteria with which its achievements would be judged. Propositions of Daniel Stufflebeam (2002, p. 1) were considered:

Since evaluation systems are context dependent, they must take into account constituents' needs, wants, and expectations plus other variables such as pertinent societal values, customs, and mores; relevant laws and statutes; economic dynamics; political forces; media interests; pertinent substantive criteria; organizational mission, goals, and priorities;

14 For more information on this aspect, see Bauer (2010, 2011).

organizational governance, management, protocols, and operating routines; and the organization's history and current challenges.¹⁵

Thus, one can say that the criteria used in the analysis, understood as “standards on which judgments are based” (STUFFLEBEAM, 2001, p. 1), were formulated based on values and objectives intrinsic in the program, but they also incorporated “pertinent societal values” or “pertinent substantive criteria”. These two were incorporated, for example, with the decision to use quality indicators in education as an indication of the effectiveness of the program, and with the investigation of its influence on improving pupils' educational outcomes and literacy levels. Thus, the focus on the criteria to be evaluated was expanded to include socially oriented objectives (literacy for all pupils), regardless of how clearly this objective was pursued by the program. The establishment of criteria and indicators for data analysis was based on the researcher's understanding of the program's theoretical and methodological assumptions. These decisions were discussed in the thesis. If the basis for judging the results had been the values of the stakeholders – whether those who propose the program or the teachers who participated in it – the conclusions of the work would have been different.

FINAL CONSIDERATIONS

Every program, project or action is part of a broader public policy and constitutes the translation into reality of the strategies designed to achieve the intended objectives and goals. In that sense, public policy translates as an intentional action of the State on society, hopefully with a view to the common good.

Political action presents, within itself, the conceptions of state officials on principles that guide the fundamental aspects of the area on which the policy is being designed. And political practice presupposes setting goals for the area in question. In determining principles and goals, a policy assumes a projective capacity, directing for a limited period the paths taken by society. Such paths reflect the conception of man, world, society, state, social justice, human rights (among many others) of the authorities who institute policies.

¹⁵ Available at: <http://www.wmich.edu/evalctr/checklists/institutionalizingeval.pdf>. Accessed on May 1st 2011

These conceptions are not always explicit, however, in the documents that organize the policy, which often has to be inferred through the acts outlined for its execution, expressed in the process of implementing the programs and projects that realize it, as between the lines of laws, regulations and the documents designed for its implementation, as well as other elements providing indications of the policy (DRAIBE, 2001). Or, still, by the behavior and conceptions of those responsible for designing and implementing the policy. Therefore, it is not possible to affirm that a political direction is neutral or free of human subjectivity. As the concretization of a given policy into actions (at a theoretical level), educational programs carry assumptions and intentionalities that need to be analyzed critically. Jannuzzi (2018) even advances in this analysis to postulate the need for republican public values to be at the basis of both program design and program evaluation.

Understanding these aspects is fundamental, especially when the researcher's objective is to contribute to the evaluation of governmental action, the conceptions that underlie it and its applicability. The evaluator can assess the objectives, actions and strategies outlined in the proposal, discussing their adequacy to the challenges and needs of society, or he can look at it from the perspective of the various stakeholders in the policy. His choice requires reflection and careful investigation, since it will often determine the basis for analyzing the quality of the program, a task that no public policy evaluator should escape.

It is the discussion about these options and their explication in the evaluation, whether in the evaluation of programs or in evaluative research, which allows the development of good theory about program evaluation. Thus, it is the job of the evaluator (or evaluative researcher) to look into the valuing aspects in order to discuss their assumptions, and also specify his own system of values and beliefs about evaluative practice, thus contributing to expand the debate about this important aspect, which is inextricable to evaluation practice.

REFERENCES

ALKIN, Marvin; CHRISTIE, Cristina A. An evaluation theory tree. In: ALKIN, Marvin C. (ed.). *Evaluation roots: tracing theorists' views and influences*. Thousand Oaks: Sage Publications, 2004. p. 12-65.

ARRETCHE, Marta. Tendências no estudo sobre avaliação. In: RICO, E. M. (org.). *Avaliação de políticas sociais: uma questão em debate*. São Paulo: Cortez/IEE, 2009. p. 29-40.

BAUER, Adriana. Avaliação de impacto no Brasil: é possível mensurar impactos de programas de formação docente? *Estudos em Avaliação Educacional*, São Paulo, v. 21, n. 46, p. 229-252, maio/ago. 2010.

BAUER, Adriana. *Avaliação de impacto de formação docente e serviço: o programa Letra e Vida*. 2011. 232 f. Tese (Doutorado em Educação) – Faculdade de Educação, Universidade de São Paulo, São Paulo, 2011. doi:10.11606/T.48.2011.tde-15072011-105940. Acesso em: 12 out. 2018.

CONTANDRIOPOULOS, André-Pierre; CHAMPAGNE, François; DENIS, Jean-Louis; PINEAULT, Raynald. A avaliação na área da saúde: conceitos e métodos. In: HARTZ, Zulmira M. de Araújo (org.). *Avaliação em saúde: dos modelos conceituais à prática na análise da implantação de programas*. Rio de Janeiro: Fiocruz, 1997. p. 29-47.

COUSINS, Bradley; SHULHA, Lyn M. Complexities in setting program standards in collaborative evaluation. In: SMITH, Nick L.; BRANDON, Paulo R. *Fundamental issues in evaluation*. New York: The Guilford Press, 2008. p. 139-158.

DAVIDSON, Jane. Merit. In: MATHISON, Sandra. *Encyclopedia of evaluation*. Thousand Oaks, CA: Sage Publications, 2005. p. 245.

DRAIBE, Sônia M. Avaliação de implementação: esboço de uma metodologia de trabalho em políticas públicas. In: BARREIRA, Maria Cecília R. Nobre; CARVALHO, Maria do Carmo B. (org.). *Tendências e perspectivas na avaliação de políticas e programas sociais*. São Paulo: IEE/PUC-SP, 2001. p. 13-42.

FIGUEIREDO, Marcus; FIGUEIREDO, Argelina. Avaliação política e avaliação de políticas: um quadro de referência teórica. *Análise e Conjunturas*, Belo Horizonte, v. 1, n. 3, p. 107-127, set./dez. 1986.

GUBA, Egon G.; LINCOLN, Yvonna S. The distinction between merit and worth in evaluation. *Educational Evaluation and Policy Analysis*, v. 2, n. 4, p. 61-71, 1980.

HOUSE, Ernest R.; HOWE, Kenneth R. Deliberative democratic evaluation. In: KELLAGHAN, T.; STUFFLEBEAM, D. L. (ed.). *International handbook of educational evaluation*. Boston: Kluwer Academic, 2003. v. 2, p. 79.

HOUSE, Ernest R. Intellectual history in evaluation. In: ALKIN, Marvin C. (ed.). *Evaluation roots: tracing theorists' views and influences*. Thousand Oaks: Sage Publications, 2004. p. 218-224.

JANNUZZI, Paulo M. Eficiência econômica, eficácia procedural ou efetividade social: três valores em disputa na avaliação de políticas e programas sociais. *Desenvolvimento em Debate*, Rio de Janeiro, v. 4, n. 1, p. 117-142, 2016.

JANNUZZI, Paulo M. A implementação como centro da avaliação de políticas públicas: reflexões à luz da experiência da SAGI/MDS 2004-maio 2016. In: SEMINÁRIO BRASILEIRO SOBRE IMPLEMENTAÇÃO DE POLÍTICAS PÚBLICAS, 1., Brasília, 20-21 de março de 2018. *Anais [...]*. Brasília: Enap, 2018.

MACDONALD, Barry. La evaluación como servicio público: perspectivas de futuro. In: SÁEZ BREZMES, María José (ed.). *Conceptualizando la evaluación en España*. Madrid: Universidad de Alcalá de Henares, 1995. (Colección Pautas, 3). Disponível em: <https://ueaeprints.uea.ac.uk/31854/>. Acesso em: 10 jul. 2018.

MACDONALD, Barry; JENKINS, David; KEMMIS, Stephen; VERMA, Gajendra. Racism and education evaluation. In: VERMA, Gajendra; BUGLEY, Christopher (ed.). *Racism and educational evaluation education and identity*. London: Macmillian, 1979.

MADAUS, George; SCRIVEN, Michael; STUFFLEBEAM, Daniel L. Program evaluation: an historical overview. In: MADAUS, George; SCRIVEN, Michael; STUFFLEBEAM, Daniel L. (ed.). *Evaluation models: viewpoints on Educational and Human Services Evaluation*. Boston: Kluwer-Nijhoff, 1983. p. 3-22.

MATHISON, Sandra. *Encyclopedia of evaluation*. Thousand Oaks, CA: Sage Publications, 2005.

MULLER, Pierre; SUREL, Yves. *A análise das políticas públicas*. Tradução: Agemir Bavaresco, Alceu R. Ferraro. Pelotas: Educat, 2002.

PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO. Garantía de calidad del diseño y la metodología de la evaluación. *Manual de planificación, seguimiento y evaluación de los resultados de desarrollo*. New York: PNUD, 2009. p. 161-178.

RAWLS, John. *Uma teoria da justiça*. São Paulo: Martins Fontes, 1997.

SCHWANDT, Thomas. Values. In: MATHISON, Sandra. *Encyclopedia of evaluation*. Thousand Oaks, CA: Sage Publications, 2005. p. 443-444.

SCRIVEN, Michael. Evaluating educational programs. In: CARO, Francis G. (ed.). *Readings in evaluation research*. New York: Russell Sage Foundation, 1971. p. 49-53.

SCRIVEN, Michael. Payoffs from evaluation, In: ABT, Clark C. *The evaluation of social programs*. Beverly Hill: Sage Publications, 1976. p. 217-224.

SCRIVEN, Michael. New frontiers of evaluation. *Evaluation Practice*. Science Direct, v. 7, n. 1, p. 7-44, Feb. 1986.

SCRIVEN, Michael. *Evaluation thesaurus*. Newbury Park, CA: Sage Publications, 1991.

SCRIVEN, Michael. Hard-won lessons in program evaluation. *New directions for program evaluation*. San Francisco, CA: AEA/Sage Publications, 1993.

SCRIVEN, Michael. Evaluation in the new millennium: the transdisciplinary vision. In: DONALDSON, Stewart I.; SCRIVEN, Michael (ed.). *Evaluating social programs and problems*. Mahwah NY: Erlbaum, 2003. p. 19-41.

SHADISH, Willian; COOK, Thomas; LEVITON, Laura. *Foundations of program evaluation: theories of practice*. Newbury Park, CA: Sage Publications, 1991.

STAKE, Robert. Stake and responsive evaluation. In: ALKIN, Marvin C. (ed.). *Evaluation roots: tracing theorists' views and influences*. Thousand Oaks: Sage Publications, 2004. p. 203-217.

STUFFLEBEAM, Daniel L. Evaluation values and criteria checklist. *Evaluation Checklists Project*. Kalamazoo, MI: The Evaluation Center, March 2001.

STUFFLEBEAM, Daniel L. Institutionalizing evaluation checklist. *Evaluation Checklists Project*. Kalamazoo, MI: The Evaluation Center, 2002. Disponível em: http://www.wmich.edu/evalctr/archive_checklists/institutionalizingeval.pdf. Acesso em: 23 fev. 2011.

STUFFLEBEAM, Daniel L.; SHINKFIELD, Antony J. *Evaluation theory, models, and applications*. San Francisco, CA: Jossey-Bass, 2007.

SUCHMAN, Edward A. Evaluating educational programs. In: CARO, Francis (ed.). *Readings in evaluation research*. New York: Russell Sage Foundation, 1971. p. 43-48.

VIANNA, Heraldo M. Avaliação de programas educacionais: duas questões. *Estudos em Avaliação Educacional*, São Paulo, v. 16, n. 32, p. 43-56, jul./dez. 2005.

WORTHEN, Blane; SANDERS, James; FITZPATRICK, Jody. *Avaliação de programas: concepções e práticas*. São Paulo: Edusp/Gente, 2004.

Received on: 15 JULHO 2018 July 15, 2018

Approved for publication on: October 22, 2018



This content is licensed under a Creative Commons attribution-type BY-NC.